BIG DATA

CHO NHÀ QUẢN LÝ

ATAL MALVIYA
MIKE MALMGREN

Khánh Trang *dịch*

NHÀ XUẤT BẢN CÔNG THƯƠNG
THAIHABOOKS

BIG DATA

CHO NHÀ QUẢN LÝ

Big Data for Managers: Creating Value

©2019 Atal Malviya & Mike Malmgren

All rights reserved.

Authorised translation from the English language edition published by Routledge, a member of the Taylor & Francis Group.

BIG DATA CHO NHÀ QUẢN LÝ

Bản quyền tiếng Việt © 2020, Công ty Cổ phần Sách Thái Hà

Không phần nào trong cuốn sách này được sao chép hoặc chuyển sang bất cứ dạng thức hoặc phương tiện nào, dù là điện tử, in ấn, ghi âm hay bất cứ hệ thống phục hồi và lưu trữ thông tin nào nếu không có sự cho phép bằng văn bản của Công ty Cổ phần Sách Thái Hà.

Biên tập viên Thaihabooks: **Hiền Nhip**

Thiết kế bìa & Trình bày: **Vương Mai – Đàm Oanh**

Công ty Cổ phần Sách Thái Hà
Trụ sở chính: 119 C5 Tô Hiệu -
Nghĩa Tân - Cầu Giấy - Hà Nội
Tel: (024) 3793 0480; Fax: (024) 6287 3238
Chi nhánh TPHCM: 88/28 Đào Duy Anh, Phường 9, Q. Phú Nhuận
Tel: (028) 22532641
Website: www.thaihabooks.com /nhasachthaiha.vn

Thái Hà Books trân trọng cảm ơn mọi ý kiến đóng góp của độc giả.
Mọi thông tin xin gửi về:

- Góp ý về biên tập: publication@thaihabooks.com
- Tư vấn dịch vụ xuất bản: dichvuxuatban@thaihabooks.com

BIG DATA

CHO NHÀ QUẢN LÝ

ATAL MALVIYA
MIKE MALMGREN

Khánh Trang *dịch*

Mục lục

Lời nói đầu *7*

1. GIỚI THIỆU **13**

Đối với nhà quản lý 15

Một cuốn sách không chuyên sâu về kỹ thuật 16

Bố cục sách 17

2. CUỘC CÁCH MẠNG DỮ LIỆU LỚN **23**

Các quyết định và quá trình tạo giá trị dựa trên dữ liệu 23

Lịch sử dữ liệu và dữ liệu lớn 37

Dữ liệu và phân tích 47

Phân tích dữ liệu và thống kê 48

Phân tích dữ liệu và điện toán 49

Tìm kiếm trên mạng với Google 53

Phân tích dữ liệu lớn trên đám mây 53

Dữ liệu có cấu trúc 54

Dữ liệu phi cấu trúc 60

Dữ liệu lớn 66

Tóm tắt 81

3. KIẾN TẠO GIÁ TRỊ TỪ DỮ LIỆU LỚN **85**

Các yếu tố thúc đẩy giá trị trong tổ chức thương mại 86

Giá trị thị trường và các giá trị phi tài chính 99

Các khoản đầu tư vào nơi dữ liệu lớn có thể tạo ra giá trị 105

Tóm tắt 143

4. CÁC KỸ THUẬT VÀ GIẢI PHÁP DỮ LIỆU LỚN **145**

Phân tích dữ liệu lớn 146

Các kỹ thuật phân tích dữ liệu 158

Tóm tắt 212

5. GIỚI THIỆU MÔ HÌNH C-ADAPT 215
Mô hình C-ADAPT trong quá trình kiến tạo giá trị 221
Bảng tính C-ADAPT 259
Tóm tắt 297

6. CÁC TRƯỜNG HỢP VÍ DỤ ĐIỂN HÌNH VỀ DỮ LIỆU LỚN 299
Ooredoo (tên cũ là Qtel) 301
Domino's Pizza 306
Công ty kinh doanh phần mềm chống virus hàng đầu 309
Gate Gourmet 313
Tesco 317
Hãng hàng không Delta Airlines 320
Intel 324
TXU Energy 327
OmedaRx 334
John Deere 339
Airbnb 342
Walmart 346
Huffington Post 350
Tóm tắt 354

7. NHỮNG NGƯỜI THỰC HÀNH NÓI GÌ 357
Dữ liệu lớn là quan trọng – rất quan trọng! 360
Giá trị chính từ dữ liệu lớn 365
Các thách thức khi triển khai dự án dữ liệu lớn 376
Tóm tắt 379

8. KẾT LUẬN VÀ BÀN LUẬN 381

LỜI NÓI ĐẦU

Hơn mười năm trước, Jeanne Harris và tôi cùng nhau viết một cuốn sách mang tên *Competing on Analytics* (tạm dịch: Cạnh tranh trong lĩnh vực phân tích). Một chủ đề mà chúng tôi thường xuyên đem ra thảo luận trong quá trình nghiên cứu và viết cuốn sách, đó là liệu tất cả các ngành nghề đều có thể hoặc mong muốn phát triển những khả năng liên quan đến phân tích không. Tôi nhớ lúc ấy mình nói rằng khả năng ngành thời trang tích cực vận dụng kỹ thuật phân tích là không cao, vì dường như đây là một hình mẫu tiêu biểu đại diện cho các ngành hoạt động dựa trên trực giác; hơn nữa, khi ấy tôi cũng không dám chắc liệu có bao giờ đủ dữ liệu tốt để thực hiện phân tích hay không.

Dĩ nhiên, tôi đã nhận định sai. Ngành thời trang giờ đây đã dịch chuyển thiên hẳn về hướng áp dụng các kỹ thuật phân tích và dữ liệu lớn (big data).

Chỉ cần tìm kiếm nhanh trên Google cụm từ khóa "fashion analytics" (phân tích thời trang), chúng ta sẽ gặp vô số nội dung như "bảy cách mới mẻ để sử dụng dữ liệu lớn trong ngành thời trang" và "năm xu hướng phân tích hàng đầu trong thời trang bán lẻ". Vậy đấy, một ngành từng phụ thuộc vào trực giác giờ đây đã mở rộng cửa tiếp nhận dữ liệu và phân tích, coi đó là phương tiện chủ đạo để ra quyết định.

Và thời trang không phải là ngành duy nhất trải qua cuộc lột xác đó. Hai tác giả của cuốn sách này nói rằng "dữ liệu lớn" là một loại dữ liệu đặc thù, khối lượng đặc biệt lớn và tương đối phi cấu trúc. Tuy đồng tình với định nghĩa này, nhưng tôi cho rằng dữ liệu lớn thực sự đã trở thành một khái niệm rộng hơn thế rất nhiều. Nó đã trở thành cách chúng ta sống hiện nay. Như các tác giả đã chứng minh thông qua nhiều ví dụ được nêu trong cuốn sách này, dữ liệu đã thâm nhập vào gần như mọi lĩnh vực ngành nghề và mọi khía cạnh hoạt động của con người. Nó thống trị địa hạt thể thao, chính trị, mạng xã hội, chăm sóc sức khỏe, và hầu hết các doanh nghiệp.

Như vậy, ngày nay các tổ chức có cơ hội tiếp cận nguồn dữ liệu lớn hơn bao giờ hết. Để hiểu được kho dữ liệu đó, chúng ta phải phân tích nó, tìm ra những mô típ và xu hướng. Người ta có thể gọi tập hợp các hoạt động đó là "dữ liệu lớn" (như cách dùng từ của

các tác giả cuốn sách này), "tình báo doanh nghiệp", "khoa học phân tích", "máy học", hay "chuối xanh", nhưng chung quy lại, chúng là một yếu tố thiết yếu của cuộc sống trong các nền kinh tế và các xã hội hiện nay. Thuật ngữ có thể thay đổi, nhưng bản thân các hoạt động hầu như vẫn giữ nguyên, ít nhất là trong vài thập niên trở lại đây.

Quá trình dịch chuyển từ một khái niệm kỹ thuật tương đối hẹp trở thành một phong trào xã hội và kinh tế như vậy có nhiều ý nghĩa khác nhau. Chẳng hạn, điều đó có nghĩa dữ liệu lớn không còn là một xu hướng nhất thời nữa. Trong giai đoạn ban đầu khi ý tưởng này mới phôi thai, các công ty triển khai những dự án dữ liệu lớn chỉ bởi họ cảm thấy mình nên thử. Có nhiều "bằng chứng khả thi"[1] và nguyên mẫu sản phẩm, nhưng rất ít dự án mang lại những ứng dụng trong sản xuất hoặc dẫn đến những quy trình nghiệp vụ mới.

Tuy nhiên, cốt lõi của cuốn sách này là quan điểm cho rằng các dự án dữ liệu lớn đang từng bước trở thành một hoạt động nghiệp vụ chính thống. Và cũng giống như mọi sáng kiến khác trong kinh doanh,

---

[1] Bằng chứng khả thi (proof of concept): Là bằng chứng cho thấy một ý tưởng được đề xuất nào đó là khả thi. Thông thường, các bằng chứng loại này được thu thập thông qua việc triển khai thí nghiệm hoặc thực hiện dự án thí điểm. (Chú thích của người dịch – ND)

các dự án này chỉ đáng thực hiện nếu chúng mang lại giá trị nhiều hơn so với chi phí bỏ ra. Các phần nói về cách kiến tạo và đo lường giá trị trong cuốn sách này sẽ giúp ích cho những ai muốn thu được lợi nhuận kinh tế từ các hoạt động đầu tư vào dữ liệu lớn.

Ban đầu, cách lựa chọn các dự án dữ liệu lớn thường mang đậm tính cá nhân – ai đó trong tổ chức có mối quan tâm đặc biệt đến một vấn đề, hoặc một vị quản lý có ảnh hưởng cấp ngân sách cho một dự án mang lại lợi ích cho đơn vị hoặc bộ phận của mình. Nhưng đó không phải cơ sở dài hạn đáng tin cậy cho quá trình ra quyết định liên quan đến các dự án dữ liệu lớn. Vì thế, trong cuốn sách này có một phần rất hữu ích bàn về mô hình C-ADAPT; mô hình này sử dụng một quy trình có hệ thống nhằm xác định, tìm hiểu, phát triển, và thử nghiệm các phân tích định lượng trong nội bộ doanh nghiệp hoặc tổ chức. Mô hình này được giải thích và minh họa thông qua một ví dụ rất chi tiết. Khi áp dụng một quy trình như thế, bạn có thể yên tâm rằng các dự án dữ liệu lớn sẽ được lên kế hoạch và triển khai để mang lại giá trị cho tổ chức.

Dữ liệu lớn còn phải chịu một cái tiếng là khó hiểu và mơ hồ. Nhưng cuốn sách này đã xóa tan ngộ nhận đó. Chương 4 trong cuốn sách bàn về "Các kỹ thuật và giải pháp về dữ liệu lớn" – chỉ nội thực tế

rằng một khối lượng kiến thức hữu ích khổng lồ về chủ đề này lại có thể gói gọn trong khuôn khổ một chương sách cũng đủ để thấy việc tiếp cận các phân tích dữ liệu lớn hoàn toàn không quá khó.

Tuy nhiên, điều gây ấn tượng nhất với tôi ở cuốn sách này không phải nội dung hay chủ đề cụ thể nào, mà là tính thực tế và sự hữu dụng. Tôi rất ủng hộ sự hợp tác giữa giới hàn lâm và giới thực hành, và cuốn sách này là một sự hợp tác như thế. Ashridge và Hult vốn vẫn luôn tập trung vào các hoạt động nghiên cứu và viết lách mà cả các nhà quản lý thiên về thực hành lẫn các chuyên gia đều có thể sử dụng được, và *Big data cho nhà quản lý* hoàn toàn phù hợp và nhất quán với truyền thống đó. Chương 6 với rất nhiều ví dụ là một minh chứng rõ ràng cho thấy cuốn sách này hướng đến đối tượng độc giả là những người thích sự cụ thể hơn các khái niệm trừu tượng. Một minh chứng khác là các kết quả khảo sát và các cuộc phỏng vấn mà hai tác giả thực hiện với những người thực hành dữ liệu lớn, được trình bày ở Chương 7.

Như vậy, nếu mục tiêu của bạn là tìm kiếm những kiến thức thực tế về cách phân tích dữ liệu lớn (hoặc nhỏ) để mang lại kết quả trong hoạt động kinh doanh hoặc giá trị về mặt tổ chức, cuốn sách này dành cho bạn. Bạn sẽ được tiếp cận các khung, phương pháp, hướng dẫn, và ví dụ hỗ trợ cho quá trình này. Như lời

thoại trong phim *Star Trek*, "kháng cự là vô ích". Thế giới đang dịch chuyển về hướng dữ liệu lớn, và bạn cũng nên sớm tìm cách tham gia vào phong trào này, nếu không bạn sẽ bị bỏ lại đằng sau đấy.

Thomas H. Davenport

Giáo sư, Đại học Babson, và nghiên cứu viên thuộc chương trình Sáng kiến về Nền Kinh tế số của MIT; đồng tác giả cuốn *Competing on Analytics* (Cạnh tranh bằng phân tích), *Big Data @Work* (tạm dịch: Dữ liệu lớn ở nơi làm việc), và *Only Humans Need Apply* (tạm dịch: Chỉ con người mới cần vận dụng.)

1

GIỚI THIỆU

Kiến tạo giá trị là mục tiêu cơ bản nhất của bất kỳ doanh nghiệp nào. Mặc dù giá trị này có thể liên quan đến tài chính hay phi tài chính, nhưng tùy thuộc vào hoạt động kinh doanh và các dự án mà bạn đang theo đuổi, mục tiêu duy nhất của hầu hết các doanh nghiệp là kiến tạo nên giá trị tài chính cho các cổ đông. Các doanh nghiệp vận dụng nhiều phương thức và công nghệ sáng tạo khác nhau để tạo ra giá trị này, và trong danh sách các công nghệ hay công cụ họ sử dụng, dữ liệu lớn là sự bổ sung gần đây nhất.

Mục tiêu của cuốn sách này là tránh những ồn ào xung quanh dữ liệu lớn để tìm hiểu khía cạnh ít mang tính kỹ thuật mà thiên về ý nghĩa kinh doanh nhiều hơn. Đồng thời, cuốn sách cũng cung cấp thông tin để hỗ trợ hàng triệu nhà quản lý đang phải đối mặt

với áp lực đầu tư và ra các quyết định liên quan đến dữ liệu lớn cho tổ chức của họ. Trọng tâm của dữ liệu lớn là công nghệ và việc sử dụng công nghệ để thu thập thông tin ở quy mô lớn chưa từng có. Tin vui là nguồn thông tin này có khả năng trang bị cho doanh nghiệp một lợi thế trước đối thủ cạnh tranh hoặc giúp họ phục vụ khách hàng tốt hơn. Tuy nhiên, để tận dụng được những tiến bộ kỹ thuật trong xử lý dữ liệu và thông tin, doanh nghiệp phải chấp nhận đầu tư tiền bạc, nguồn lực, và thời gian. Đối với nhiều nhà quản lý doanh nghiệp, môi trường kỹ thuật thay đổi không ngừng này đặt ra một thách thức: Làm sao để biết nên đầu tư vào dữ liệu lớn ở đâu thì hợp lý và thu được lợi ích. Để đưa ra một quyết định đầu tư trong kinh doanh, bạn cần nắm rõ những yếu tố tác động liên quan đến cách tạo ra giá trị, và với dữ liệu lớn là kết quả khả thi của các quyết định đó.

Có thể bạn và đội ngũ của mình khởi động và triển khai dự án dữ liệu lớn thành công nhất để đạt được một mục tiêu cụ thể hoặc để giải quyết một thách thức, nhưng không phải lúc nào các bạn cũng tạo ra được giá trị ngắn hoặc dài hạn cho công ty. Vì thế, với cuốn sách này, chúng tôi sẽ vạch ra một cách tiếp cận và bối cảnh rất thực tế cho các nhà quản lý đã hoặc đang có ý định đặt chân vào hành trình thực hiện một dự án dữ liệu lớn.

Một mặt, chúng tôi cung cấp các công cụ, kỹ thuật, và quy trình để triển khai thực hiện các dự án dữ liệu lớn; nặt khác, chúng tôi vạch ra quy trình tạo giá trị và cách đo lường giá trị được tạo ra – và chúng tôi mong muốn rằng các nhà quản lý sẽ đọc, học hỏi, và thực hàm cả hai khía cạnh này để đưa ra được những quyế định kinh doanh hợp lý, từ đó tạo ra giá trị bền vững dài hạn cho công ty.

ĐỐI VỚI NHÀ QUẢN LÝ

Từ kinh nghệm làm việc trực tiếp với các nhà quản lý ở nhiều tổ chức có quy mô khác nhau, chúng tôi nhận thấy rằng mặc dù công nghệ và phát minh mang lại giá trị lớn clo doanh nghiệp, nhưng các nhà quản lý không phụ tách mảng kỹ thuật thường gặp khó khăn trong việc hểu tiềm năng đích thực của các công nghệ đó và sử dụng chúng cho tổ chức của mình vào đúng thời điểm. Cùng lúc, các trang web, thiết bị di động, mạng xã hội và các nền tảng khác đang không ngừng tạo ra nhữn lưu lượng dữ liệu phi cấu trúc khổng lồ, hàm chứa trong đó cả một kho thông tin giá trị về thị trường. Khi nở khóa được kho "Dữ liệu lớn" này, bạn sẽ giành đực sức mạnh để xây dựng giá trị bền vững theo nhiều hnh thức khác nhau.

Mục tiê của cuốn sách này là nhằm giúp các nhà quản lý hiểu được rằng họ có thể sử dụng dữ liệu lớn

như thế nào và theo cách nào để tạo ra nhiều doanh thu, tiết kiệm chi phí, và đưa ra nhiều sản phẩm sáng tạo – nhưng trên hết, nó có thể mang lại giá trị bền vững cho doanh nghiệp. Cuốn sách có thể giúp nhà quản lý hiểu dữ liệu lớn là gì và không là gì. Nó cũng có thể giúp nhà quản lý có được những cuộc trao đổi hiệu quả và ý nghĩa với những người bên trong và bên ngoài tổ chức đồng tình rằng dữ liệu lớn là câu trả lời cho các vấn đề trong kinh doanh nhưng vẫn còn khá mơ hồ, chưa hiểu rõ việc đầu tư vào đó có thể tạo ra giá trị như thế nào.

Mục tiêu của cuốn sách này là nhằm giúp độc giả nói chung và các nhà quản lý nói riêng hiểu được tiềm năng của dữ liệu lớn, lên kế hoạch triển khai các dự án cần thiết, nhưng quan trọng hơn, giúp họ đánh giá được kết quả đầu tư để bảo đảm rằng các dự án đó mang lại giá trị cho tổ chức.

Một lợi ích lớn mà cuốn sách mang lại là mô hình kinh doanh và khung C-ADAPT độc quyền mà nhà quản lý có thể sử dụng để xác định xem nên triển khai các dự án dữ liệu lớn như thế nào, ở đâu, và có thể tạo ra giá trị bằng cách nào với dữ liệu lớn.

MỘT CUỐN SÁCH KHÔNG CHUYÊN SÂU VỀ KỸ THUẬT

Đây là cuốn sách viết cho các nhà quản lý, không phải sách kỹ thuật dành cho các chuyên gia hay sinh viên

công nghệ và IT. Nhà quản lý cần nắm được cái nhìn tổng quan, nhất là để có thể hiểu và bàn bạc với các chuyên gia IT và các nhà tư vấn. Nói như thế không có nghĩa rằng chúng tôi hoàn toàn không sử dụng từ ngữ chuyên môn hay không đề cập đến quá trình vận động phát triển của hoạt động tạo và sử dụng dữ liệu, tuy vậy, chúng tôi đã cố gắng trình bày ở mức độ tối thiểu.

Cuốn sách này giới thiệu một số lượng vừa đủ các khía cạnh kỹ thuật của dữ liệu lớn mà các nhà quản lý nhiều tham vọng muốn nhảy vào biển cơ hội mênh mông này cần nắm rõ. Chúng tôi không bàn chi tiết đến khía cạnh phần cứng hay phương tiện lưu trữ dữ liệu lớn, mà tập trung hơn vào mặt phân tích và trí tuệ của dữ liệu lớn, cũng là nguồn tạo ra giá trị cho hầu hết các doanh nghiệp.

BỐ CỤC SÁCH

Mặc dù đây không phải sách kỹ thuật, nhưng Chương 2 có hai phần. Phần thứ nhất bàn về cách áp dụng các công nghệ dữ liệu lớn của một số tổ chức, với kết quả thường là tạo ra những cú đột phá trong các thị trường truyền thống. Phần thứ hai khái quát sơ lược về lịch sử phát triển của ngành khoa học dữ liệu và một số công nghệ. Đây là phần quan trọng vì nó cung cấp một số kiến thức cơ bản về các thuật ngữ và công

nghệ làm nền tảng cho dữ liệu lớn. Cụ thể, nó giải thích về sự khác biệt giữa có cấu trúc và phi cấu trúc. Nguồn dữ liệu phi cấu trúc phát triển bùng nổ từ các cuộc trao đổi, blog, và website trên mạng internet và có thể hàm chứa nhiều kiến giải quan trọng về khách hàng, các xu hướng và quan điểm mà hàng triệu người dùng internet chia sẻ. Trọng tâm của dữ liệu lớn là khai thác mỏ dữ liệu quý giá này.

Chương 3 là phần dẫn nhập cung cấp những kiến thức nền tảng để tìm hiểu về việc kiến tạo giá trị, theo đó giải thích cách tính toán giá trị tài chính và đưa ra một số ví dụ về cách sử dụng dữ liệu lớn để tạo giá trị. Vì đối tượng độc giả mục tiêu của cuốn sách này là nhà quản lý trong các tổ chức, nên chúng tôi thiết nghĩ họ cần nắm rõ về các yếu tố tác động trong quá trình kiến tạo giá trị, vốn là một khía cạnh then chốt đối với bất kỳ quyết định đầu tư nào, trong đó có dữ liệu lớn. Chương này lấy bối cảnh là đời sống tài chính của một doanh nghiệp quy mô vừa tiêu biểu, vì thế chúng tôi sử dụng các mẫu báo cáo kết quả kinh doanh và bản cân đối kế toán của loại hình doanh nghiệp này để chỉ ra tác động tiềm năng của việc đầu tư vào dữ liệu lớn. Mục tiêu của chương này là trang bị cho độc giả kiến thức vững chắc về các yếu tố tác động trong quá trình kiến tạo giá trị. Chúng tôi hy vọng các nhà quản lý sẽ vận dụng những gì học được từ chương này

để đo lường mức độ thành công của các dự án mà họ triển khai trong tương lai. Chúng tôi cũng giới thiệu một mô hình và khung ở Chương 5; trong giai đoạn cuối cùng của mô hình này, ý tưởng kiến tạo giá trị sẽ được kiểm định và đưa vào triển khai.

Chương 4 là phần mô tả các công nghệ dữ liệu lớn. Chương này đưa chúng ta vào giai đoạn tiếp theo: hiểu rằng dữ liệu và tác động của nó đối với quá trình kiến tạo giá trị có liên quan đến việc phân tích và rút ra các liến giải từ nguồn dữ liệu thu thập được. Ở chương này, chúng ta sẽ cùng nhau bàn về nhiều kỹ thuật phân tích dữ liệu khác nhau, từ các mô hình thống kê truyền thống đến các lĩnh vực mới mẻ như khoa học phân tích dự đoán và các kỹ thuật trực quan hóa dữ liệu. Chương này có một vài chỗ bàn sâu về khía cạnh kỹ thuật; tuy nhiên, điều này khó tránh khỏi bởi nền tảng của dữ liệu lớn là công nghệ. Mục tiêu của chúng tôi là đưa ra một phần trình bày tổng quan cũng như lập danh sách các điểm có lợi và bất lợi để mà quản lý có thể tự tin trao đổi với các nhà khoa học dữ liệu hoặc các đội kỹ thuật. Một phần quan trọng của dữ liệu lớn là dữ liệu phi cấu trúc, chẳng hạn dữ liệu từ email, blog, và các dữ liệu dựa trên văn bản, nên chương này cũng sẽ trình bày một phần tổng quan về các kỹ thuật phân tích khác nhau có thể vận dụng cho loại hình phân tích này.

Chương 5 là trọng tâm của cuốn sách. Dựa trên công việc và những quan sát rút ra được từ nhiều nghiên cứu về dữ liệu lớn, chúng tôi xin trân trọng giới thiệu với bạn đọc một mô hình có thể giúp nhà quản lý triển khai thành công các dự án dữ liệu lớn. **Mô hình C-ADAPT** về kiến tạo giá trị bằng dữ liệu lớn ở Chương 5 là một mô hình có hệ thống và là một biểu mẫu thiết thực hỗ trợ đắc lực cho các nhà quản lý được giao trọng trách xây dựng chiến lược cho các dự án dữ liệu lớn. Chúng tôi sẽ giới thiệu về mô hình và khung này, đồng thời giải thích về các thành phần khác nhau trong đó cũng như hướng dẫn độc giả cách sử dụng chúng. Giai đoạn cuối cùng của mô hình này sẽ giúp bạn đo lường được giá trị tạo ra, qua đó đảm bảo rằng dự án vừa thực hiện là sự đầu tư thỏa đáng về thời gian và công sức, hoặc gợi ý cho bạn về những thay đổi có thể thực hiện trong dự án tiếp theo. Chúng tôi cũng giới thiệu công cụ chính để sử dụng kèm với mô hình trên – **bảng tính C-ADAPT** – nhằm hỗ trợ bạn trong quá trình triển khai dự án, đồng thời giúp bạn tập trung để đưa dự án đi đúng hướng.

Chương 6 bao gồm một loạt các trường hợp ví dụ xoay quanh việc phân tích và tạo giá trị liên quan đến dữ liệu lớn với sự tham gia góp mặt của nhiều doanh nghiệp, từ những công ty hoạt động trong lĩnh vực công nghệ như Intel đến những công ty hoạt động

trong lĩnh vực nông nghiệp như John Deere. Chúng tôi đưa vào đây nhiều ví dụ đa dạng nhằm chứng minh cho độc giả thấy phân tích dữ liệu lớn có thể được sử dụng ở bất kỳ lĩnh vực ngành nghề nào, với bất kỳ loại hình doanh nghiệp nào. Nhưng điều quan trọng nhất là thông qua tất cả các trường hợp này, bạn sẽ thấy rằng mô hình C-ADAPT có thể được thiết lập hoặc vận dụng khi cần thiết. Các tác giả của cuốn sách đã có kinh nghiệm làm việc trực tiếp trong một số trường hợp ví dụ được nêu, ngoài ra còn có một số trường hợp là đóng góp của những người khác trong nhiều lĩnh vực đa dạng. Ý định của chúng tôi là giới thiệu tất cả các trường hợp này để bạn có thể dễ dàng xác định vấn đề cụ thể mà mình đang phải đối mặt. Chúng tôi đã cố gắng trình bày theo hướng cô đọng súc tích và trực tiếp nhất có thể – nếu bạn cần thêm thông tin chi tiết hoặc muốn trao đổi thêm, hãy liên hệ thẳng với các tác giả.

Chương cuối cùng tóm lược lại những bài học chính trong cuốn sách, đồng thời đề cập đến một số vấn đề mà chúng tôi cho rằng xảy ra trong bối cảnh công nghệ thay đổi nhanh chóng với nhiều yếu tố đột phá như hiện nay – đây là những vấn đề mà nhà quản lý ở các doanh nghiệp sẽ cần tìm hiểu và đưa ra quyết định.

Nếu muốn đọc nhanh, bạn có thể đọc lướt qua Chương 4 – Các kỹ thuật và giải pháp dữ liệu lớn.

Nếu cảm thấy mình đã nắm rõ về khái niệm và quy trình kiến tạo giá trị, bạn có thể đọc lướt qua Chương 3.

Sau khi bạn đọc xong cuốn sách này, chúng tôi hy vọng bạn sẽ tự tin hơn về chủ đề này và về quy trình kiến tạo giá trị.

2

CUỘC CÁCH MẠNG DỮ LIỆU LỚN

CÁC QUYẾT ĐỊNH VÀ QUÁ TRÌNH TẠO GIÁ TRỊ DỰA TRÊN DỮ LIỆU

Nếu doanh nghiệp của bạn sản xuất các sản phẩm hoặc dịch vụ mà không doanh nghiệp nào khác cung cấp, không có sản phẩm/dịch vụ thay thế cho sản phẩm/dịch vụ của bạn trên thị trường, nhu cầu đối với sản phẩm/dịch vụ của bạn ngày càng cao và khách hàng biết cách tiếp cận với bạn, khi đó doanh nghiệp của bạn đang sở hữu một vị thế rất đắc lợi. Thông thường, có thể xếp loại hình kinh doanh này vào nhóm kinh doanh độc quyền, và điều duy nhất bạn lỳ vọng là không có doanh nghiệp nào khác bắt tay vào sản xuất một sản phẩm tương tự hoặc sản phẩm thay thế trong tương lai. Trong thế giới ngày nay, doanh nghiệp độc quyền chỉ có thể hình thành trong những tình huống hết sức đặc biệt – chẳng hạn,

khi chính phủ trao vị thế độc quyền cho một doanh nghiệp như công ty bưu chính. Nhưng đối với phần lớn các doanh nghiệp đang hoạt động trên thị trường, viễn cảnh này là phi thực tế và rất khó khăn. Thực ra ngay cả trong trường hợp một doanh nghiệp được trao vị thế kinh doanh độc quyền, cũng không ai dám khẳng định rằng họ có thể duy trì hoạt động trong những điều kiện bất biến.

Sự cạnh tranh là yếu tố chủ đạo trong bất kỳ thị trường nào – trên thế giới hiện nay, rất khó tìm được lĩnh vực kinh doanh nào vắng bóng sự cạnh tranh. Các doanh nghiệp nỗ lực dựng lên những rào cản gia nhập bằng cách sử dụng những phương thức và chiến lược độc nhất vô nhị nhằm ngăn chặn đối thủ bước chân vào thị trường mà họ đang hoạt động. Đôi khi những chiến lược này cũng xuất phát từ chính sách định vị hoặc phân khúc thị trường của doanh nghiệp. Trong một môi trường cạnh tranh lành mạnh, sự cạnh tranh đóng vai trò chất xúc tác cho sự đổi mới sáng tạo và phát triển trên thị trường.

Ở Anh, hai chuỗi siêu thị Aldi và Lidl nổi tiếng với các sản phẩm giá rẻ, và họ thu hút những khách hàng nhạy cảm về giá tìm đến hệ thống cửa hàng của mình. Trong quý IV năm 2015, hai chuỗi siêu thị này đạt mức tăng trưởng doanh số bán hàng lần lượt là 13,3% và 18,5%. Cùng lúc đó, Waitrose, chuỗi siêu

thị chuyên cung cấp các sản phẩm giá cao, lại chứng kiến đợt sụt giảm doanh số tồi tệ nhất kể từ năm 2006. Sau đó, tới năm 2016, Waitrose tập trung vào đổi mới sáng tạo và ra mắt chiến dịch giảm giá mới mang tên "Hãy lựa chọn sản phẩm riêng cho bạn", trong đó họ quyết định giảm giá 20% cho 10 loại sản phẩm được khách hàng mua nhiều nhất trong số 100 sản phẩm. Chiến lược này đạt được thành công lớn, và chiến dịch mới giúp doanh số bán hàng của họ gia tăng 3,7% trong khi các chuỗi siêu thị khác như Tesco, Asda, và Morrisons suy giảm. Chiến lược giảm giá của Waitrose lấy cảm hứng từ thành công của các siêu thị giá rẻ và bầu không khí thị trường sau giai đoạn suy thoái ảnh hưởng đến sức mua của khách hàng. Cơ sở để đề ra chiến lược giảm giá hết sức đúng đắn này là những hoạt động nghiên cứu thị trường sâu rộng, bao gồm các dữ liệu về thị trường và đối thủ cạnh tranh, các phân tích thực hiện trên dữ liệu về hành vi người dùng, và tình báo doanh nghiệp. Đối với Waitrose, "Hãy lựa chọn sản phẩm riêng cho bạn" là một trong những chiến lược thành công nhất giúp họ thu hút được 700.000 khách hàng mới chỉ trong vòng ba tháng đầu tiên sau khi triển khai chiến lược. Đây chỉ là một trong vô số các ví dụ thực tế trong đó các doanh nghiệp bán lẻ thu thập dữ liệu bán hàng và sử dụng các kỹ thuật phân tích để lập ra hành trình khách hàng và hành vi mua sắm một cách chính xác.

Tesco, một chuỗi bán lẻ lớn ở Anh, nổi tiếng với chiếc thẻ trung thành (Thẻ Câu lạc bộ Tesco) – họ dùng các dữ liệu chi tiêu và sử dụng để thiết lập các xu hướng mua sắm và hành vi tiêu dùng của khách hàng.

Năm 2008, Brian Chesky và Joe Gebbia thành lập Airbnb. Thực ra Airbnb là một thị trường mở, người dùng nền tảng này sẽ trả phí để có thể đăng tải thông tin quảng cáo, tìm kiếm, và thuê nhà nghỉ/ phòng trống. Một mình Airbnb đã đứng ra thách thức cả ngành khách sạn và lưu trú vốn đã tồn tại hàng trăm năm. Đối với các công ty lớn trong ngành như Marriott, giá trị chủ đạo của họ nằm ở các tài sản cố định, và hoạt động của họ diễn ra trên cơ sở là quy mô và vị trí của các bất động sản mà họ sở hữu. Theo truyền thống, lĩnh vực khách sạn được cho là ngành đòi hỏi các hoạt động đầu tư vốn mạnh mẽ vì có liên quan đến các chi phí về bất động sản. Như chúng ta đã biết, trong nhiều trường hợp, sự đổi mới sáng tạo xuất phát từ các lĩnh vực ngành nghề khác. Các công ty công nghệ đã đưa ra những tiên đoán về một kỷ nguyên mới trong lĩnh vực khách sạn. Airbnb và nhiều công ty tương tự khác như OYO Rooms và Zoho Rooms đã tiếp cận vấn đề này từ một góc độ khác hẳn. Họ nhận thấy có nhiều khách du lịch và nhân viên của các doanh nghiệp nhỏ không đủ ngân sách hoặc kinh phí để thuê phòng khách sạn khi phát sinh nhu

cầu gấp, bên cạnh đó, hàng nghìn cư dân có phòng dư trong nhà có thể cho thuê. Airbnb bắt tay vào giải quyết vấn đề này từ gốc và tạo ra những cách thức khác nhau để chủ sở hữu phòng trống và du khách có thể kết nối và hợp tác với nhau. Tinh giản số tài sản sở hữu là một điểm mạnh chủ chốt của Airbnb, và cũng là một trong những lý do chính đằng sau sự phát triển thần kỳ của họ trong một thời gian ngắn. Năm 2014, khi Marriott tuyên bố dự định bổ sung thêm 30.000 phòng trong năm tiếp theo, Chesky đã đăng tải một tweet trên tài khoản Twitter của mình rằng Airbnb có thể bổ sung thêm 30.000 phòng trong vòng hai tuần.

Tới mùa xuân năm 2014, Airbnb có 550.000 bất động sản đăng tải trên website của họ với 10 triệu khách đăng ký. Điều mà Airbnb làm trong ngành khách sạn cũng chính là điều mà Uber làm trong ngành taxi.

Năm 2008, vào một ngày trời đổ tuyết, Travis Kalanick loay hoay mãi vẫn không gọi được taxi để đến tham dự một cuộc hội thảo. Từ tình cảnh oái oăm này, Kalanick chợt nhận ra rằng cần phải có dịch vụ gọi taxi qua điện thoại di động với chỉ hai cú chạm bàn phím. Sau đó, tại hội thảo công nghệ LeWeb, anh đã có một cuộc trao đổi định mệnh với Garrett Camp, nhà sáng lập của StumbleUpon. Camp chia sẻ với anh về ý tưởng mở một dịch vụ thuê xe cao cấp rất thuận

tiện mà không tốn tới 800 đô-la cho một chuyến đi – mức giá mà bản thân Camp từng phải cắn răng bỏ ra. Nhận ra đây là vấn đề mà rất nhiều người ở khắp nơi trên thế giới đều gặp phải, Kalanick đã cùng Garrett Camp thành lập Uber Cabs, và bây giờ được biết tới với tên gọi đơn giản hơn: Uber. Uber không phải một hãng taxi, mà là giải pháp công nghệ đối với một vấn đề từng được coi là liên quan đến lĩnh vực vận tải. Hành khách có thể gọi xe Uber thông qua một ứng dụng di động trên Google Play Store và các cửa hàng ứng dụng. Khi người dùng bật ứng dụng định vị GPS, Uber sẽ thông báo cho các lái xe đang hoạt động gần đó, nhờ vậy bảo đảm rằng khách hàng bắt được xe trong thời gian nhanh nhất có thể vào đúng thời điểm cần đón. Hiện dịch vụ của Uber đã xuất hiện ở nhiều quốc gia trên thế giới, bao gồm Mỹ, Anh, Ấn Độ, Pháp, và Australia. Uber đã và đang phải đối mặt với một số trường hợp phản đối từ các hãng taxi hiện hành và các tổ chức khác, nhưng giải pháp đầy tính sáng tạo của họ vẫn luôn nhận được sự ủng hộ nhiệt tình của hàng triệu người dùng sẵn sàng sử dụng dịch vụ của Uber. Tại thời điểm tháng 9 năm 2016, Uber thực hiện trung bình 1 triệu chuyến xe mỗi ngày và mỗi tháng nhận thêm 50.000 tài xế mới làm đối tác.

Airbnb và Uber là hai ví dụ điển hình về các doanh nghiệp mang tinh thần đổi mới sáng tạo; những tân

binh này đã làm thay đổi cách thức kinh doanh của các công ty đang hoạt động trong ngành. Dĩ nhiên, chúng ta không thể giả định rằng Uber sẽ thay thế tất cả các hãng taxi hay Airbnb sẽ nuốt chửng cả ngành khách sạn, nhưng hai công ty này đã tìm ra được một cách thức rất sáng tạo nhằm giải quyết một vấn đề cho người dùng dịch vụ, và may mắn là có hàng triệu người sẵn sàng sử dụng các dịch vụ của họ.

Trên thực tế, tất cả những gì một doanh nghiệp phải biết là làm thế nào để cung cấp giải pháp đáp ứng các nhu cầu của khách hàng theo hướng **nhanh hơn, tốt hơn, hoặc rẻ hơn** so với đối thủ cạnh tranh. Để làm được điều này, doanh nghiệp cần đưa ra những lựa chọn chiến lược để giữ lấy vị thế dẫn đầu trong ngành. Bruce Henderson, nhà sáng lập hãng tư vấn Boston Consulting Group, và Michael Porter, giảng viên tại trường Kinh doanh Harvard, là hai nhà lãnh đạo tư tưởng uy tín từng giới thiệu khái niệm chiến lược kinh doanh cho giới kinh doanh trong giai đoạn từ năm 1965 đến 1980. Ở đây chúng tôi sẽ không bàn chi tiết về những đóng góp giá trị của họ, bởi bạn có thể dễ dàng tìm thấy các công trình của họ đăng tải trên internet. Đối với những người như Henderson và Porter, ý tưởng cơ bản về tư duy chiến lược trong kinh doanh xuất phát từ sự hiểu biết sâu sắc về các yếu tố và các bên liên quan khác nhau trong bất kỳ

doanh nghiệp nào (quy trình, sản phẩm, và nguồn lực) và thay đổi chúng để tạo ra một kết quả tốt hơn (giá trị), đồng thời tạo dựng một vị trí nhiều lợi thế. Những thay đổi xuất phát từ sự hiểu biết về kinh doanh này được xây dựng trên cơ sở là những thông tin giá trị về doanh nghiệp của bạn, bối cảnh cạnh tranh, và lĩnh vực ngành nghề mà bạn đang hoạt động – tóm lại đó là tình báo doanh nghiệp. Tình báo doanh nghiệp thường được coi là một khía cạnh kỹ thuật trong kinh doanh với những báo cáo và bảng biểu, đồ thị nhằm hỗ trợ nhà quản lý trong quá trình ra các quyết định mang tính chiến lược. Trên thực tế, khái niệm này không liên quan gì đến công nghệ, ngoại trừ việc bây giờ chúng ta đã có cách tạo ra nhiều thông tin phức tạp hơn, mang lại nhiều hiểu biết hơn cho doanh nghiệp.

Năm 1865, trong cuốn *Cyclopaedia of Commercial and Business Anecdotes* (tạm dịch: Bách khoa thư về các giai thoại trong thương mại và kinh doanh), Richard Millar Devens đã sử dụng cụm từ "tình báo doanh nghiệp" để nói về việc Henry Furnese, một lãnh đạo ngân hàng, đã tạo ra giá trị bằng cách nhận thông tin trước các đối thủ cạnh tranh và hành động kịp thời.

Thế giới kinh doanh đã thay đổi rất nhiều trong 150 năm qua. Đặc biệt là trong 20-30 năm qua, những đổi mới sáng tạo về công nghệ đã cho phép

các lãnh đạo doanh nghiệp đưa ra được những quyết định quan trọng một cách nhanh chóng và hiệu quả dựa trên những điểm bằng chứng (dữ liệu) vững chắc. Giờ đây các lãnh đạo doanh nghiệp có thể tiếp cận với nguồn thông tin và tin tức tình báo nhiều hơn bao giờ hết, đến độ hầu như không còn chỗ cho những giả định và phán đoán mò. Các doanh nghiệp và khách hàng hiện đang tạo ra nhiều dữ liệu hơn bao giờ hết, và doanh nghiệp có nguồn lực để lưu trữ, sử dụng, cũng như phân tích khối lượng dữ liệu khổng lồ này một cách nhanh chóng. Nhiều gã khổng lồ công nghệ như Oracle và SAP đang giúp các doanh nghiệp xây dựng các bộ công cụ tình báo doanh nghiệp (business intelligence – BI) để đưa vào các hệ thống cơ bản của họ, và các hệ thống tình báo doanh nghiệp của họ cũng đang hỗ trợ các doanh nghiệp trong việc đưa ra những lựa chọn thông minh và ra quyết định dựa trên cơ sở thông tin đầy đủ. Chẳng hạn, các nhà bán lẻ như Tesco ở Anh và Walmart ở Mỹ đã không còn phụ thuộc vào chiến lược "giảm giá thành, tăng doanh số" áp dụng chung cho tất cả khách hàng, mà hiện đã bắt đầu chuyển sang thu thập và phân tích hành vi chi tiêu của khách hàng để có thể gửi những thông tin mời chào mua hàng được đo ni đóng giày riêng cho từng nhóm khách hàng. Nếu bạn mua hàng trực tuyến trên Tesco hoặc sử dụng thẻ câu lạc bộ Tesco để nhận được các chương trình khuyến mãi, thì bạn sẽ

nhận được phiếu giảm giá cho các sản phẩm mà bạn sử dụng nhiều hơn, hoặc có thể muốn sử dụng trong tương lai. Tesco không chỉ biết gia đình bạn cần bao nhiêu chai sữa mỗi tuần, mà còn biết bạn có thể sẽ quan tâm đến loại sản phẩm mới nào. Sở dĩ họ làm được như vậy là vì sử dụng các thuật toán lập hồ sơ người dùng thông minh dựa trên bộ công cụ tình báo doanh nghiệp.

Nhiều website hiện đang sử dụng dữ liệu để phát triển công cụ đề xuất, nhờ đó có thể giới thiệu những sản phẩm khác nhau cho từng đối tượng khách hàng dựa trên thị hiếu và sở thích riêng của mỗi người. Amazon.com nổi tiếng với công cụ đề xuất của họ – nếu bạn mua một cuốn sách dạy nấu ăn trên Amazon.com, họ sẽ đề xuất các cuốn sách dạy nấu ăn tương tự ngay trên website và trong các thư tin tức gửi tới hộp thư của bạn. Phần cốt lõi ở đây là hệ thống đề xuất của gã khổng lồ trong lĩnh vực bán lẻ này, gọi là hệ thống "lọc hợp tác từ sản phẩm sang sản phẩm", được xây dựng dựa trên một số yếu tố thông tin đơn giản như "người dùng đã mua gì trong quá khứ", "họ đưa sản phẩm nào vào giỏ hàng trực tuyến", "những sản phẩm họ đã đánh giá và nhấn nút thích", và "các khách hàng khác đã xem và mua những gì"... Phương pháp đề xuất dựa trên hành vi cũng được áp dụng ở các cổng web khác như YouTube và LinkedIn.

Ra quyết định dựa trên dữ liệu cũng giúp doanh nghiệp thực hiện được những thay đổi mang tính đột phá trên thị trường. Ví dụ, Myntra.com, một công ty khởi nghiệp hàng đầu Ấn Độ trong lĩnh vực thương mại điện tử, thông báo rằng họ sẽ đóng cổng web vào ngày 1 tháng 5 năm 2015 và chỉ tiếp tục hoạt động trên các ứng dụng di động. Trong năm trước đó, 80% lưu lượng truy cập website Myntra.com và 70% doanh số bán hàng đến từ các thiết bị di động. Một trong những lý do chính của hiện tượng này là khách hàng có thể dễ dàng truy cập internet với chi phí thấp thông qua các thiết bị di động. Nhiều thành phố và làng mạc không có internet hoặc chỉ có đường truyền internet chất lượng thấp thông qua các kết nối có dây, và điện thoại di động đã giúp trám vào lỗ hổng này. Ấn Độ là một trong những quốc gia đã nhảy cóc qua kỷ nguyên của điện thoại bàn và kết nối internet có dây nhờ vào sự xuất hiện của các thiết bị di động giá rẻ và một thị trường dịch vụ viễn thông mang tính cạnh tranh cao, từ đó làm giảm đáng kể chi phí của các dịch vụ thoại và dữ liệu. Dựa trên những thực tế đó và nguồn tin tức tình báo doanh nghiệp do bộ phận sau bán hàng của Myntra.com cung cấp, ban lãnh đạo công ty đã quyết định thực hiện chiến lược chỉ tập trung cung cấp dịch vụ qua điện thoại di động để nhắm vào cơ sở khách hàng sử dụng thiết bị di động ngày càng phát triển nhanh chóng của Ấn Độ, với sức tăng trưởng theo cấp số nhân trong 2-3 năm qua.

Có nhiều ví dụ cho thấy những quyết định dựa trên dữ liệu đã và đang hỗ trợ các doanh nghiệp phát triển bứt phá. Các công ty như Uber hay Myntra.com đã tiến thêm một bước trong các sản phẩm/dịch vụ liên quan đến công nghệ. Nhưng bên cạnh đó cũng có không ít ví dụ cho thấy có những doanh nghiệp không "đọc" được thông điệp chính xác từ luồng dữ liệu trên thị trường, rốt cuộc đưa ra những quyết sách sai lầm dẫn đến thua lỗ. Morrisons-Kiddicare là một ví dụ điển hình về những thua lỗ lớn trong khía cạnh này.

Morrisons là chuỗi siêu thị lớn thứ ba ở Anh. Năm 2011, họ bỏ ra 70 triệu bảng để mua lại cửa hàng trực tuyến Kiddicare chuyên bán đồ cho mẹ và bé, với hy vọng rằng sẽ tạo được cơ hội bán hàng đa kênh bằng cách vận hành đồng thời cả hệ thống cửa hàng trực tuyến và phi trực tuyến. Nhưng Morrisons không nhận ra rằng xu hướng mua sắm trực tuyến và mua sắm qua thiết bị di động ở khu vực này khi đó đang phát triển nhanh chóng, tỉ lệ thâm nhập của điện thoại thông minh trên thị trường cũng cao hơn, do đó việc mở rộng sang không gian phi trực tuyến là một quyết định tốn kém. Morrisons mua lại chuỗi cửa hàng Best Buy thất bại ở Anh và chuyển đổi chúng thành hệ thống cửa hàng Kiddicare. Trong giai đoạn từ năm 2011 đến 2014, Morrisons liên tục thua lỗ nặng, và cuối cùng, vào năm 2014, họ buộc phải bán

lại Kiddicare cho quỹ đầu tư cổ phần riêng Endless với giá 2 triệu bảng; tổng thiệt hại mà Morrisons phải gánh trong toàn bộ thương vụ này là 163 triệu bảng. Hiện nay, Kiddicare đã quay trở lại làm cửa hàng trực tuyến.

Cuộc tranh luận xung quanh vấn đề dữ liệu và phân tích không chỉ giới hạn ở các tổ chức lớn và có lịch sử hoạt động lâu đời. Với sự hậu thuẫn của nhiều công cụ phân tích miễn phí hoặc có chi phí thấp, kết hợp với mức chi phí lưu trữ dữ liệu giảm đáng kể nhờ công nghệ điện toán đám mây, nhiều doanh nghiệp vừa và nhỏ đã bắt đầu tận dụng các kỹ thuật phân tích dữ liệu trong hoạt động hằng ngày. Các công cụ như Google Analytics đã trở thành công cụ bắt buộc phải có đối với hầu hết mọi doanh nghiệp hoạt động trong không gian trực tuyến. Nhiều công ty khởi nghiệp trong lĩnh vực công nghệ hiện cũng đang tích cực giới thiệu nhiều công cụ phân tích thông minh hơn với mức chi phí thấp hơn, và ngay cả thị trường kinh doanh công cụ phân tích bây giờ cũng càng lúc càng cạnh tranh hơn. Xin đơn cử một ví dụ, trong giai đoạn từ 5 đến 7 năm, nhiều công ty phân tích mạng xã hội đã ra mắt các giải pháp riêng của họ và thiết lập được vị trí rất vững chắc trong nhóm các doanh nghiệp vừa và nhỏ. Cơ hội và tính cạnh tranh trên thị trường đã bắt đầu củng cố vững chắc hơn cho lĩnh

vực này. Trong vòng sáu năm qua, Hootsuite, một công ty phân tích mạng xã hội được nhiều người ưa dùng đã huy động được hơn 246 triệu đô-la tiền vốn và tiến hành mua lại hơn mười công ty khác, trong đó có cả các công ty đối thủ là Seesmic năm 2012 và UberVU năm 2014.

Các doanh nghiệp không chỉ ra quyết định dựa trên dữ liệu nhằm mục đích bán thêm nhiều sản phẩm hoặc dịch vụ, mà còn sử dụng dữ liệu trong nhiều khía cạnh khác nhau với nhiều mục tiêu khác nhau, từ cải thiện hoạt động kinh doanh và lập kế hoạch nguồn lực đến phát triển sản phẩm và đổi mới sáng tạo. Trong mười năm qua, các khách hàng, doanh nghiệp, và thiết bị đã sản xuất ra lượng dữ liệu đồ sộ hơn bao giờ hết. Theo kết quả nghiên cứu của IBM năm 2012, 90% dữ liệu trên thế giới được sản xuất vào năm 2011 và 2012 – vì vậy, các doanh nghiệp có thể yên tâm ngày càng có nhiều dữ liệu để họ học hỏi, ra quyết định hành động, và cải thiện dần trong thế giới số này. Nguồn dữ liệu đó đang được tạo ra rất nhanh chóng với số lượng khổng lồ và có nhiều hình thức tồn tại khác nhau xuất phát từ nhiều nguồn đa dạng, bao gồm các thiết bị cảm biến, mạng xã hội, thiết bị di động, các hoạt động trực tuyến, tệp tin lưu trữ trong thiết bị, cùng vô vàn nguồn dữ liệu khác. Đây là một cơ hội lớn cho các nhà quản lý và lãnh

đạo doanh nghiệp có thể tận dụng và khai thác được những kho dữ liệu ngày càng gia tăng này.

"Dữ liệu lớn" là thuật ngữ dùng để chỉ nguồn dữ liệu ngày càng phát triển nhanh chóng, có quy mô rộng và hết sức đa dạng này; đặc điểm của nó là quá lớn đến nỗi không thể lưu trữ trong một máy chủ truyền thống duy nhất, quá phi cấu trúc đến nỗi không thể tiến hành phân tích dữ liệu trên các hệ thống tình báo doanh nghiệp kiểu cũ, và chuyển động quá nhanh chóng đến nỗi không thể đưa vào một nhà kho dữ liệu tĩnh như trước đây. Nguồn dữ liệu này có thể mang lại giá trị hạn chế cho các nhà quản lý doanh nghiệp, bởi có khả năng họ không lĩnh hội được giá trị thực sự của hiện tượng hết sức mới mẻ này. Từ lưu trữ đến phân tích, các doanh nghiệp có thể tiết kiệm được nhiều khoản chi phí lớn, tạo ra doanh thu và thu thập được những kiến giải giá trị, có tính khả thi cao. Dữ liệu lớn là một công cụ chính yếu để các nhà quản lý sử dụng trong quá trình xây dựng và triển khai thực thi các chiến lược thành công.

LỊCH SỬ DỮ LIỆU VÀ DỮ LIỆU LỚN

Bất kỳ vị CEO nào cũng cần có khả năng đặt ra một câu hỏi liên quan đến việc kết nối dữ liệu xuyên suốt tổ chức, có khả năng điều hành hiệu quả một công ty, và đặc biệt là có khả năng

phản ứng tốt trước những sự kiện bất ngờ. Phần lớn các tổ chức đều thiếu năng lực kết nối toàn bộ dữ liệu lại với nhau.

– Tim Berners-Lee (nhà phát minh mạng diện rộng toàn cầu World Wide Web)

Quản lý luồng dữ liệu đổ về với khối lượng lớn từ nhiều nguồn đa dạng khác nhau và chắt lọc giá trị từ đó không phải là thách thức mới phát sinh. Năm 1944, Fremont Rider, thủ thư của Đại học Wesleyan, nêu ý kiến rằng quy mô của các thư viện thuộc các trường đại học trên lãnh thổ nước Mỹ cứ sau 16 năm lại mở rộng gấp đôi. Ông cũng dự đoán rằng tới năm 2040, thư viện Đại học Yale sẽ có khoảng 200 triệu cuốn sách, chiếm trên 9.600 ki-lô-mét kệ sách. Rõ ràng, Rider đang nói về những thách thức của việc lưu trữ, quản lý, và sử dụng kho dữ liệu lớn hơn năng lực lưu trữ dữ liệu hiện hành. Dĩ nhiên, vào thời điểm đó, đây không được coi là một vấn đề liên quan đến dữ liệu lớn, nhưng trong bối cảnh này, Rider đang đề cập đến một hiện tượng tương tự mà ngày nay chúng ta liên hệ nó với dữ liệu lớn.

Từ thời Rider đến nay, chúng ta đã chứng kiến nhiều tiến bộ trong lĩnh vực công nghệ và đặc biệt trong lĩnh vực quản lý dữ liệu. Suốt một thời gian dài, dữ liệu được lưu trữ và quản lý một cách thông minh

trong khả năng xoay xở của các Hệ thống Quản lý Cơ sở Dữ liệu (Data Base Management Systems – DBMS) và sau đó là các Hệ thống Quản lý Cơ sở Dữ liệu Quan hệ (Relational Data Base Management Systems – RDBMS). DBMS được phát minh vào đầu thập niên 1960, còn RDBMS ra đời vào thập niên 1970, cả hai đều xuất phát từ các phòng thí nghiệm của tập đoàn IBM. Khi làm việc với các hệ thống tương đối nhỏ, chúng ta dễ dàng hài lòng với các cách tiếp cận của DBMS hay RDBMS cũng như các hệ thống thông minh được xây dựng trên đó, cho đến khi chúng ta gặp phải một hiện tượng gọi là internet.

Tim Berners-Lee sinh năm 1955 và từng theo học tại Đại học Oxford. Năm 1985, khi đang làm việc với Tổ chức Nghiên cứu Hạt nhân châu Âu (CERN), ông công bố một bài báo mang tiêu đề "Một đề xuất về quản lý thông tin". Trong bài báo này, Berners-Lee giới thiệu một hệ thống tạo, chia sẻ, và phân phối thông tin toàn cầu và đặt tên cho nó là Mạng Diện rộng Toàn cầu (World Wide Web). Khi đó, Berners-Lee không hề biết rằng hệ thống mới này sẽ mở ra cánh cửa cơ hội để toàn thế giới cùng thỏa sức tạo và chia sẻ những khối lượng thông tin lớn, và rằng sẽ khó lưu trữ cũng như quản lý luồng dữ liệu được tạo ra trong quá trình này với các hệ thống hiện hành. Lẽ ra các lãnh đạo trong ngành phải coi mốc này là sự

khởi đầu sớm của kỷ nguyên dữ liệu lớn, nhưng mãi tới đầu năm 2000, hai công ty hàng đầu trong thế giới công nghệ – Yahoo! và Google – mới xác định và dự đoán các vấn đề liên quan đến việc lưu trữ và quản lý lượng dữ liệu (lớn) khổng lồ, đồng thời bắt tay vào tìm kiếm giải pháp cho các vấn đề này. Công việc họ làm chủ yếu xuất phát từ nhu cầu của chính họ đối với việc quản lý các trang web và dữ liệu tìm kiếm vốn luôn luôn thách thức những giới hạn của các cơ sở hạ tầng và quy trình hiện hành.

Năm 2013, Roger Magoulas đưa ra định nghĩa về dữ liệu lớn – khi quy mô và hiệu suất hoạt động của dữ liệu trở thành một phần vấn đề. Dữ liệu lớn không chỉ giới hạn bởi định nghĩa mà Magoulas đề xuất. Nguồn dữ liệu khổng lồ, đa dạng, và phát triển với tốc độ nhanh chóng đã khiến các hệ thống lưu trữ hiện hành và hệ thống phân tích kiểu cũ trở nên lỗi thời – và đây là một vấn đề (đồng thời cũng là một cơ hội) thực sự lớn.

Việc lưu trữ và sử dụng lượng dữ liệu khổng lồ này chính là vấn đề đầu tiên cần xử lý trước khi chuyển sang xác định những cơ hội mà dữ liệu có thể mang đến cho các doanh nghiệp. Mặc dù quy mô ngày càng gia tăng của dữ liệu có thể là một vấn đề nghiêm trọng đối với nhiều doanh nghiệp, và một số công ty lớn có đủ nguồn lực để xử lý vấn đề lưu trữ dữ liệu này,

nhưng một số người vẫn tiếp tục nghiên cứu để giải quyết vấn đề theo những cách thức sáng tạo, đồng thời tập trung vào vấn đề lớn tiếp theo: truy xuất và quản lý dữ liệu. Các công ty quản lý hàng triệu trang web và cố gắng đánh chỉ mục (index) cho chúng trong các cơ sở dữ liệu của họ theo thời gian thực chính là lực lượng hiện hữu đằng sau sự ra đời của các công nghệ dữ liệu lớn. Google và Yahoo! đã đóng góp rất nhiều trong giai đoạn đầu của quá trình phát triển các công nghệ này.

Năm 2004, hai nhân viên của Google – Jeffery Dean và Sanjey Ghemawat – giới thiệu một bản đặc tả kỹ thuật xác định một mô hình lập trình gọi là MapReduce. Trong khoảng năm năm, nhiều nhân viên của Google, trong đó có Dean và Ghemawat, bận rộn làm việc với những lượng dữ liệu khổng lồ đến từ các hoạt động của người dùng internet như số lượng các trang web mà mỗi máy chủ tiếp cận, tập hợp các thắc mắc được hỏi nhiều nhất mỗi ngày... Họ có thể phân phối các tập dữ liệu lớn trên nhiều máy chủ khác nhau để xử lý song song và có thể sử dụng mô hình MapReduce để xử lý các tập dữ liệu lớn này trong một khoảng thời gian tương đối ngắn. Đây được coi là một thành công quan trọng trong giai đoạn đó. Nhưng đồng thời vào lúc này, các kỹ sư ở Yahoo! cũng đang miệt mài tìm cách giải quyết vấn

đề tương tự theo cách khác. Yahoo! đặt tên cho dự án này là Nutch, với người phụ trách dự án là Doug Cutting, giám đốc của thư viện số Internet Archive Search, và Mike Cafarella, một sinh viên cao học ở Đại học Washington.

Sau khi tham gia dự án Nutch được vài tháng và cố gắng tìm cách quản lý luồng dữ liệu của 1 tỉ trang web, Doug Cutting tình cờ bắt gặp bản đặc tả khái niệm của Google, trong đó trình bày rằng MapReduce có thể giúp giải quyết vấn đề tương tự nhưng theo cách thức hiệu quả hơn. Lấy cảm hứng từ khung MapReduce, Doug Cutting và Mike Cafarella lập ra khung dưới đây dựa trên một công nghệ nguồn mở và ra mắt phần mềm Hadoop vào năm 2005.

Apache Hadoop là một khung phần mềm mã nguồn mở dùng để lưu trữ phân tán và xử lý phân tán các tập dữ liệu cực lớn trên các cụm máy tính được xây dựng từ các phần cứng thông thường. Trong lĩnh vực lưu trữ dữ liệu, có một thực tế không thể tránh khỏi là thất bại về phần cứng khá phổ biến và nếu có thể, nên xử lý tự động các thất bại này. Hadoop đã chủ động giải quyết vấn đề tồn tại dai dẳng này, theo đó tất cả các module đều được thiết kế nhằm xử lý tự động các thất bại do khung này gây ra.

Phần cốt lõi của Apache Hadoop bao gồm một phần lưu trữ, gọi là Hệ thống File Phân tán Hadoop (Hadoop Distributed File System – HDFS), và một phần xử lý gọi là MapReduce. Hadoop chia các file dữ liệu thành những khối dữ liệu lớn rồi phân phối chúng trên một loạt các nút mạng trong một cụm máy tính. Để xử lý dữ liệu, Hadoop truyền mã đóng gói cho các nút mạng để chúng xử lý đồng thời dựa trên dữ liệu cần xử lý. Cách làm này tận dụng được tính cục bộ của dữ liệu – các nút mạng điều khiển những dữ liệu mà chúng có thể truy cập – để xử lý tập dữ liệu nhanh hơn và hiệu quả hơn so với kiến trúc siêu máy tính truyền thống vốn phụ thuộc vào hệ thống file song song, trong đó hoạt động tính toán và dữ liệu được phân phối thông qua việc kết nối mạng lưới tốc độ cao.

Cơ sở của khung Apache Hadoop bao gồm các module sau:

1. Hadoop Common – chứa các thư viện và tiện ích mà các module khác của Hadoop cần đến;
2. Hệ thống File Phân tán Hadoop (Hadoop Distributed File System – HDFS) – một hệ thống file phân tán nhằm lưu trữ dữ liệu trên các máy tính thông thường, cung cấp băng thông tổng có tốc độ rất cao trên toàn bộ cụm máy tính;

3. Hadoop YARN – một nền tảng quản lý nguồn lực chịu trách nhiệm quản lý các nguồn lực điện toán trong các cụm và sử dụng chúng để lên lịch trình cho các ứng dụng của người dùng;

4. Hadoop MapReduce – một phương án triển khai mô hình lập trình MapReduce để xử lý dữ liệu quy mô lớn.

Bản thân khung Hadoop chủ yếu được viết bằng ngôn ngữ lập trình Java, ngoài ra cũng có một số mã lệnh riêng cho từng chương trình viết bằng ngôn ngữ C và các tiện ích dòng lệnh viết theo câu lệnh shell[1]. Mặc dù MapReduce thường được viết bằng ngôn ngữ lập trình Java, nhưng bất kỳ ngôn ngữ lập trình nào cũng có thể được sử dụng với "Hadoop Streaming" để thực thi các phần "bản đồ" (map) và "giảm bớt" (reduce) trong chương trình của người dùng. Các dự án khác trong hệ sinh thái Hadoop góp phần làm phong phú thêm cho các giao diện người dùng.

Thuật ngữ Hadoop được mở rộng dần và giờ đây nó không những được dùng để chỉ các module cơ sở như mô tả ở trên, mà còn chỉ toàn bộ hệ sinh thái, hay tập hợp các gói phần mềm bổ sung, có thể được cài đặt bên trên hoặc bên cạnh Hadoop, chẳng hạn Apache Pig, Apache Hive, Apache HBase, Apache Phoenix, Apache Spark, Apache ZooKeeper, Cloudera Impala, Apache Flume, Apache Sqoop, Apache Oozie, Apache Storm.

Các hợp phần MapReduce và HDFS của Apache Hadoop được xây dựng dựa trên các tài liệu của Google về MapReduce và Hệ thống File Google.

---

[1] Câu lệnh shell (shell script): Là danh sách các lệnh được liệt kê theo thứ tự thi hành. (ND)

Nhiều công ty và công ty khởi nghiệp đã tạo ra các phiên bản Hadoop riêng cho mình dựa trên khung mã nguồn mở Apache. Về khía cạnh truy xuất và quản lý dữ liệu, nhiều công ty cũng đã tự tạo ngôn ngữ kịch bản riêng – chẳng hạn, ngôn ngữ Pig và Hive giúp các doanh nghiệp phân tích lượng dữ liệu khổng lồ được lưu trong các nút mạng của Hadoop.

Trong thế giới dư dả dữ liệu ngày nay, các doanh nghiệp thu thập lượng lớn dữ liệu từ rất nhiều nguồn đa dạng như mạng xã hội, lưu lượng truy cập web, file lưu trữ trong hệ thống, thông tin phản hồi của khách hàng, các hoạt động của nhân viên... Vì vậy, hiện nay chúng ta có thể truy cập nguồn dữ liệu dồi dào từ hầu như mọi khía cạnh trong đời sống kinh doanh, và đây là một cơ hội lớn cho các doanh nghiệp nếu họ có thể rút ra được những giá trị đúng đắn từ nguồn dữ liệu phù hợp – chúng ta sẽ bàn đến phần này ở các chương tiếp theo. Gần đây, chúng ta đã chứng kiến nhiều trường hợp trong đó sự tham gia của dữ liệu và hoạt động phân tích dữ liệu đã làm thay đổi tình hình hoàn toàn, bất kể là trong lĩnh vực kinh doanh hay chính trị. Cuộc bầu cử đầu tiên dựa trên dữ liệu lớn – tức cuộc bầu cử tổng thống Mỹ năm 2012 – là một ví dụ.

Năm 2012, đội ngũ thực hiện chiến dịch tranh cử của Barack Obama rõ ràng đã giành chiến thắng không chỉ trong cuộc bầu cử mà cả trong việc hiểu

hành vi liên tục thay đổi của cử tri nhờ phát triển các ứng dụng trên mạng xã hội và thiết bị di động. Đội ngũ của Obama đã nhận thức được tầm quan trọng của việc triển khai độ phức tạp của dữ liệu và những khó khăn đi kèm. Một trong những ưu tiên hàng đầu trước cuộc bầu cử là thực hiện một cuộc sát nhập cơ sở dữ liệu đại quy mô kéo dài 18 tháng để đưa tất cả dữ liệu về một kho chứa duy nhất. Sự tập trung về cơ sở dữ liệu này còn cho phép đội ngũ của Obama có được tầm nhìn mở rộng về cách sử dụng các chỉ số đo lường. "Chúng tôi sẽ đo lường từng yếu tố trong chiến dịch này," quản lý chiến dịch tranh cử Jim Messina chia sẻ với tạp chí *Time* trong một cuộc phỏng vấn. Messina cũng được giao cho những nguồn lực đáng tin cậy để làm việc, và ông được phép tuyển dụng một bộ phận phân tích đông đảo gấp năm lần so với đợt thực hiện chiến dịch tranh cử năm 2008. Cách tiếp cận nảy khác biệt, thực ra là đối lập hoàn toàn so với đội ngũ thực hiện tranh cử của phe đối thủ.

Dựa trên các thông tin mà toàn bộ đội ngũ Obama cùng nắm được, họ có thể chạy các cuộc bầu cử mô phỏng hơn 66.000 lần mỗi tối với cơ sở là các dữ liệu thu thập được trong ngày, nhằm bảo đảm rằng họ quan sát và lưu ý đến tác động của từng điểm dữ liệu gắn với cử tri cũng như đến tổng thể cuộc bầu cử. Ngoài ra, nhờ những thông tin về nhân khẩu học thu thập được

và so sánh với các yếu tố khác, họ có thể tìm ra những cách thức mục tiêu khác nhau để dựa vào đó mua các khung quảng cáo trên truyền hình nhằm tiếp cận với từng phân khúc cử tri nhỏ. Rõ ràng, chính chiến lược này đã mang lại thế thượng phong cho Obama trong cuộc bầu cử đầu tiên dựa vào dữ liệu lớn. Trước khi trao chiếc ghế tổng thống cho Donald Trump, Obama đã đắc cử hai lần và điều hành nước Mỹ trong tám năm. Trong cả hai cuộc bầu cử của Obama, các hoạt động phân tích dữ liệu lớn đều mang lại những giá trị lớn lao.

DỮ LIỆU VÀ PHÂN TÍCH

Bạn có thể có dữ liệu mà không có thông tin, nhưng bạn không thể có thông tin mà thiếu dữ liệu.

– Daniel Keys Moran (lập trình viên máy tính và tác giả dòng sách khoa học viễn tưởng)

Không thể phủ nhận rằng dữ liệu đang tồn tại khắp nơi xung quanh chúng ta. Chúng ta quan sát thấy chúng, sử dụng chúng, bỏ qua chúng, hoặc thậm chí không hề hay biết về chúng (trừ khi bạn đang sống ở một vùng xa xôi hẻo lánh không hề có bất kỳ sự liên hệ nào với thế giới bên ngoài), nhưng dữ liệu không hề là một "hiện tượng" mới mẻ. Dữ liệu đã tồn tại xung quanh chúng ta từ rất, rất lâu, khi loài người

bắt đầu biết giao tiếp, sử dụng ngôn ngữ ký hiệu hay các biến thể âm thanh đơn giản để chỉ các hoạt động hay đối tượng khác nhau.

Như Thomas Davenport và D. J. Patil đã viết trong một bài báo đăng tải năm 2012 mang tên "Chuyên gia khoa học dữ liệu: Công việc hấp dẫn nhất của thế kỷ 21", phân tích dữ liệu đang ngày càng thu hút được nhiều sự chú ý hơn và đây là một trong những công việc được săn đón nhiều nhất. Nó đang đóng một vai trò ngày càng quan trọng hơn trong các doanh nghiệp thuộc đủ mọi quy mô, kể cả các công ty khởi nghiệp quy mô nhỏ. Việc thực hành phân tích dữ liệu đã và đang phát triển dần dần theo thời gian, chủ yếu nhờ sự tiến bộ trong công nghệ điện toán. Trong suốt nhiều năm, những yếu tố nền tảng của phân tích dữ liệu vẫn giữ nguyên thông qua các quy trình xử lý dữ liệu và thực hiện các kỹ thuật phân tích trên dữ liệu đó.

PHÂN TÍCH DỮ LIỆU VÀ THỐNG KÊ

Phân tích dữ liệu bắt nguồn từ thống kê, một lĩnh vực có lịch sử tương đối lâu dài. Người ta nói rằng khởi thủy của ngành thống kê là Ai Cập cổ đại, khi nước này thực hiện điều tra dân số định kỳ để xây dựng kim tự tháp. Trong suốt chiều dài lịch sử, thống kê đã đóng một vai trò quan trọng đối với các chính

phủ ở khắp nơi trên thế giới, trong quá trình thiết lập các cuộc điều tra dân số tổng thể nhằm phục vụ các hoạt động hoạch định chính sách đa dạng của chính phủ (dĩ nhiên, trong đó bao gồm cả những vấn đề liên quan đến thuế). Tương tự, các doanh nghiệp đã và đang sử dụng dữ liệu và thống kê, bằng các biện pháp trực tiếp hoặc gián tiếp, trong một thời gian rất dài. Với những dữ liệu thu thập được, chúng ta có thể chuyển sang bước tiếp theo là phân tích dữ liệu. Phân tích dữ liệu là một quá trình bắt đầu với việc truy xuất dữ liệu từ nhiều nguồn khác nhau, sau đó tiến hành phân tích chúng nhằm phát hiện ra những thông tin hữu ích. Chẳng hạn, việc phân tích sự tăng trưởng dân số theo từng quận giúp chính phủ xác định được số lượng bệnh viện cần có ở từng khu vực. Cũng với thông tin tương tự về sự tăng trưởng dân số, các doanh nghiệp có thể ra quyết định mở rộng hay thu hẹp hoạt động ở các khu vực khác nhau.

PHÂN TÍCH DỮ LIỆU VÀ ĐIỆN TOÁN

Những tiến bộ trong cơ chế thu thập thông tin

Sự ra đời của máy vi tính và các tiến bộ tiếp theo trong công nghệ điện toán đã mở ra nhiều khả năng phân tích dữ liệu. Khi hoạt động thu thập dữ liệu còn được thực hiện bằng phương pháp thủ công (không có máy tính), đó là một quy trình dài, mệt mỏi, và dễ

mắc sai sót. Năm 1880, để hoàn tất một đợt tổng điều tra dân số ở Mỹ, người ta phải mất bảy năm thu thập dữ liệu và đưa ra bản báo cáo cuối cùng. Năm 1890, nhằm rút ngắn thời gian thực hiện, Herman Hollerith phát minh ra "máy lập bảng". Máy này có khả năng xử lý theo hệ thống dữ liệu được lưu trên các tấm thẻ bấm lỗ. Với sự hỗ trợ của chiếc máy này, vào năm 1890, người ta cũng thực hiện cuộc tổng điều tra dân số tương tự nhưng chỉ mất 18 tháng, với mức chi phí chỉ bằng một phần nhỏ so với trước kia.

Cơ sở dữ liệu quan hệ

Có thể coi bước ngoặt trong lĩnh vực thu thập và phân tích dữ liệu xuất hiện cùng với sự ra đời của RDB (relational database – cơ sở dữ liệu quan hệ) vào thập niên 1980, theo đó cho phép người dùng viết ngôn ngữ truy vấn (SQL) để truy xuất dữ liệu từ cơ sở dữ liệu cho trước. Sự kết hợp giữa RDB và SQL cho phép người dùng bắt đầu chạy và phân tích các truy vấn theo yêu cầu tùy vào nhu cầu kinh doanh, và do vậy, những thay đổi trong nguyên tắc kinh doanh có thể được đưa vào rất dễ dàng. Điều này khiến quá trình lấy dữ liệu dễ dàng hơn và giúp phổ quát hóa việc sử dụng cơ sở dữ liệu. Như bạn có thể thấy, sự kết hợp giữa một bên là quá trình thu thập dữ liệu dễ dàng hơn với chi phí thấp hơn và một bên là công nghệ lưu

trữ và truy xuất dữ liệu rẻ hơn, nhanh hơn đã và đang nới rộng ranh giới của khả năng sử dụng dữ liệu.

Kho dữ liệu và tình báo doanh nghiệp

Tới cuối thập niên 1980, khối lượng dữ liệu được thu thập gia tăng nhanh chóng, tỉ lệ thuận với chi phí mua ổ đĩa cứng để lưu trữ số dữ liệu này. Đây là một thách thức lớn đối với các doanh nghiệp tin vào giá trị của dữ liệu, nhưng khối lượng dữ liệu ngày càng lớn và chi phí lưu trữ ngày càng gia tăng khiến họ chật vật xoay xở. Sau đó, William H. Inmon đề xuất một khái niệm gọi là Kho Dữ liệu (Data Warehouse) – một hệ thống nhằm tối ưu hóa quá trình báo cáo và phân tích dữ liệu trên cơ sở thời gian dữ liệu, loại hình hoạt động, và nhãn thời gian gắn với dữ liệu. Đây là cách tiếp cận khác so với cách lưu trữ dữ liệu truyền thống, trong đó nó chỉ là một kho lưu trữ và một quy trình truy xuất phẳng tuyến tính, dữ liệu được lưu trữ với nhãn thời gian, và các thao tác như DELETE (xóa) và UPDATE (cập nhật) được sử dụng ít thường xuyên hơn nhiều. Như vậy, nếu bộ phận bán hàng của doanh nghiệp muốn so sánh xu hướng dữ liệu bán hàng cho từng quý, dữ liệu có thể được lưu trữ trên cơ sở hằng quý đi kèm với nhãn thời gian tương ứng và có thể được truy vấn trên cơ sở nhãn thời gian. Ngoài ra, vào năm 1989, Howard Dresner ở hãng nghiên cứu Gartner

cũng giới thiệu khái niệm BI (Business Intelligence – tình báo doanh nghiệp). BI hỗ trợ cho quá trình ra quyết định trong kinh doanh thông qua hoạt động tìm kiếm, thu thập, và phân tích các dữ liệu. Có thể coi sự ra đời của khái niệm này là một điều hết sức tự nhiên, vì vào thời điểm đó các công nghệ như cơ sở dữ liệu và kho dữ liệu đã xuất hiện để hỗ trợ cho nó. Các công ty lớn đặc biệt chào đón BI bằng việc phân tích dữ liệu khách hàng một cách có hệ thống khi đưa ra những quyết định kinh doanh. Các doanh nghiệp nhỏ phải mất một thời gian mới tiếp nhận khái niệm BI trong hình thức ban đầu của nó, nhưng giờ đây, bất kể quy mô doanh nghiệp nào, việc thu thập dữ liệu và tình báo doanh nghiệp đã trở thành một phần hữu cơ không thể tách rời của công ty.

Khai phá dữ liệu

Khai phá dữ liệu là một quy trình điện toán nhằm xác định các mô típ và xu hướng trong một tập dữ liệu. Khái niệm này được đưa ra từ thập niên 1990. Bằng cách phân tích dữ liệu theo cách khác với các phương pháp thông thường, người ta có thể thu được những kết quả bất ngờ nhưng rất hữu dụng. Sở dĩ khai phá dữ liệu có thể phát triển được là nhờ các công nghệ về cơ sở dữ liệu và kho dữ liệu, cho phép các công ty lưu trữ nhiều dữ liệu hơn nhưng vẫn thực hiện phân tích dữ liệu hợp lý. Một xu hướng kinh doanh chung

xuất hiện, trong đó các công ty bắt đầu "dự đoán" nhu cầu tiềm năng của khách hàng dựa trên việc phân tích các xu hướng mua hàng trong quá khứ.

TÌM KIẾM TRÊN MẠNG VỚI GOOGLE

Bước thay đổi quan trọng tiếp theo là internet. Trước nhu cầu tìm kiếm website trên mạng, Larry Page và Sergey Brin đã phát triển công cụ tìm kiếm Google có khả năng xử lý và phân tích dữ liệu lớn trong các máy tính phân tán. Điều đáng ngạc nhiên nằm ở chỗ, Google có thể đưa ra kết quả mà bạn muốn chỉ trong vòng vài giây. Điểm chính của hệ thống này là nó được "tự động hóa", "có khả năng điều chỉnh quy mô", và "có hiệu suất hoạt động cao". Bản đặc tả kỹ thuật về MapReduce năm 2004 đã thu hút sự quan tâm của rất nhiều kỹ sư, hình thành nên cả một đội quân tài năng nhằm đối phó với thách thức xử lý dữ liệu lớn. Vào cuối thập niên 2000, nhiều dự án phần mềm mã nguồn mở như Apache Hadoop và Apache Cassandra đã ra đời để đương đầu với thách thức này.

PHÂN TÍCH DỮ LIỆU LỚN TRÊN ĐÁM MÂY

Đầu những năm 2010 đánh dấu sự ra mắt của Amazon Redshift, một kho dữ liệu hoạt động trên đám mây, và Google BigQuery, kho dữ liệu xử lý các yêu cầu truy vấn dữ liệu trên hàng nghìn máy chủ cửa Google.

Cả hai đều giúp làm giảm đáng kể chi phí cũng như loại bỏ bớt những khó khăn và rắc rối trong quá trình xử lý dữ liệu lớn. Ngày nay, mọi công ty và kể cả các công ty khởi nghiệp, vốn thường không có ngân sách để thực hiện những phân tích như vậy, đều đã có thể lặp lại các chu trình PDCA (lập kế hoạch – thực hiện – kiểm tra – điều chỉnh) nhanh chóng nhờ sử dụng các công cụ dữ liệu lớn.

Chúng ta hãy cùng suy ngẫm một chút về quá trình vận động phát triển và tiến hóa của Dữ liệu – có cấu trúc, phi cấu trúc (bán cấu trúc, siêu dữ liệu) và Dữ liệu lớn. Theo quan điểm của tôi, tất cả đều là những khái niệm đã tồn tại xung quanh chúng ta từ khá lâu, tuy nhiên chúng ta mới có thể nhận ra, ghi lại, và sử dụng chúng trong vài thập niên trở lại đây, sau khi xuất hiện các công nghệ và hệ thống mới.

DỮ LIỆU CÓ CẤU TRÚC

Dữ liệu có cấu trúc đề cập đến các loại dữ liệu có mức độ tổ chức cao, chẳng hạn thông tin trong một cơ sở dữ liệu quan hệ. Khi thông tin mang tính cấu trúc cao và dự đoán được, thì các kết quả tìm kiếm thông tin có thể được tổ chức dễ dàng hơn và hiển thị theo nhiều cách thức sáng tạo hơn.

Dữ liệu có cấu trúc trước tiên phụ thuộc vào việc tạo ra một mô hình dữ liệu – tức mô hình về các loại

dữ liệu kinh doanh sẽ được ghi lại và cách thức lưu trữ, xử lý, và truy cập chúng. Mô hình này bao gồm lĩnh vực, loại hình, hoặc bất kỳ sự giới hạn nào liên quan đến dữ liệu được thu thập dựa trên các yêu cầu kinh doanh hoặc nguyên tắc kinh doanh cho trước. Dữ liệu có thể là số, tiền tệ, chữ cái, tên, ngày tháng, địa chỉ, và dữ liệu có thể được khoanh vùng để giới hạn với những thuật ngữ như ông, bà, hay tiến sĩ...

Lợi thế của dữ liệu có cấu trúc là có thể dễ dàng nhập, lưu trữ, truy vấn, và phân tích chúng. Đã từng có lúc, do chi phí cao và những hạn chế trong hiệu suất lưu trữ, ghi nhớ, và xử lý mà các cơ sở dữ liệu quan hệ và bảng tính sử dụng dữ liệu có cấu trúc là những phương tiện duy nhất để quản lý dữ liệu hiệu quả. Mọi dữ liệu không khớp với một cấu trúc có mức độ tổ chức chặt chẽ cho trước sẽ phải chuyển sang hình thức lưu trữ trên giấy và được cất trong tủ hồ sơ.

SQL (Structured Query Language – ngôn ngữ truy vấn có cấu trúc) là một ngôn ngữ lập trình được tạo ra để quản lý và truy vấn dữ liệu có cấu trúc trong các hệ thống quản lý cơ sở dữ liệu quan hệ. Ngôn ngữ này ban đầu do tập đoàn IBM phát triển vào đầu thập niên 1970, và sau đó phiên bản thương mại được công ty Relational Software (bây giờ là Oracle Corporation) phát triển.

Dữ liệu có cấu trúc là một bước cải tiến lớn so với các hệ thống phi cấu trúc sử dụng giấy, nhưng cuộc

sống không phải bao giờ cũng nằm gọn lỏn trong những chiếc hộp bé xíu. Kết quả là, dữ liệu có cấu trúc vẫn luôn phải cần đến sự hỗ trợ của phương pháp lưu trữ bằng giấy hay microfilm[1]. Do hiệu suất công nghệ tiếp tục được cải thiện, và chi phí có chiều hướng đi xuống, nên có thể đưa vào các hệ thống điện toán cả dữ liệu phi cấu trúc và dữ liệu bán cấu trúc.

Các ngân hàng bán lẻ là một trong những thể chế lâu đời nhất trong thế giới kinh doanh hiện đại. Trước khi tiến hành số hóa hoạt động và sử dụng các hệ thống tiên tiến để tối ưu hóa quy trình nghiệp vụ, các ngân hàng sử dụng những cuốn sổ có kích thước lớn để ghi lại và quản lý dữ liệu bằng phương pháp thủ công (nghĩa là dùng giấy và bút). Ngày ấy, các ngân hàng một lúc quản lý nhiều loại sổ đăng ký để phục vụ cho các mục đích khác nhau, bao gồm hồ sơ khách hàng, giao dịch tiền bạc, tài khoản vay... và trong hầu hết các trường hợp, mối liên hệ giữa tất cả các cuốn sổ này là một số chung, gọi là số tài khoản, được gán riêng cho từng khách hàng. Cách đây chưa lâu, các ngân hàng bán lẻ thường làm sổ ngân hàng bằng giấy giao cho khách hàng để theo dõi tất cả các hoạt động được thực hiện trên tài khoản ngân hàng của họ với

---

[1] Microfilm (hay còn gọi là vi phim): Loại film được dùng để chụp ảnh dữ liệu bằng phương pháp thu nhỏ. (ND)

số tài khoản giống nhau (thực ra hiện nay, sổ ngân hàng giấy vẫn còn được dùng phổ biến ở những quốc gia đang phát triển).

Tiếp đến là thời kỳ các hoạt động ngân hàng chuyển từ hồ sơ sổ sách bằng giấy sang máy vi tính. Nhiều quy trình nghiệp vụ được số hóa trong khi vẫn giữ nguyên những thuật ngữ và khái niệm hiện hành trong lĩnh vực ngân hàng – số tài khoản được dùng làm khóa phân biệt (khóa chính/khóa tổng hợp) và các loại bảng dữ liệu khác nhau được lập ra để thay thế cho các loại sổ đăng ký bằng giấy. Cũng trong khoảng thời gian này, từ thập niên 1970 đến thập niên 1980, những quy trình tương tự cũng diễn ra ở các lĩnh vực ngành nghề khác.

Những thuật ngữ như DBMS, RDBMS, Đồ thị ER, ERP... ra đời; những khái niệm và công nghệ này được nhiều lĩnh vực sử dụng rộng rãi từ thập niên 1970. Từ trang tính cho đến ERP, các dữ liệu doanh nghiệp phổ biến được đưa vào nhiều loại bảng biểu khác nhau, và các bảng này lại được kết nối với nhau thông qua các khóa chỉ mối quan hệ. Tính đến lúc này, dữ liệu được lưu trữ và sử dụng theo một phương pháp có trật tự được xác định từ trước.

DBMS: Hệ thống quản lý cơ sở dữ liệu (Database Management System – DBMS) là một ứng dụng phần mềm máy tính tương tác với người dùng, với các ứng dụng khác, và với bản thân cơ sở dữ liệu để thu thập và phân tích dữ liệu. Một DBMS đa mục đích được thiết kế nhằm định nghĩa, tạo ra, truy vấn, cập nhật, và quản lý các cơ sở dữ liệu.

RDBMS: Hệ thống quản lý cơ sở dữ liệu quan hệ (Relational Database Management System – RDBMS) là hệ thống quản lý cơ sở dữ liệu (DBMS) dựa trên mô hình quan hệ được phát triển bởi E. F. Codd thuộc phòng thí nghiệm nghiên cứu San Jose của IBM. Năm 2016, nhiều cơ sở dữ liệu được sử dụng phổ biến đều dựa trên mô hình cơ sở dữ liệu quan hệ kiểu này.

Đồ thị ER: Đồ thị thực thể - mối quan hệ (Entity-Relationship Diagram – ERD) là một kỹ thuật lập mô hình dữ liệu minh họa bằng đồ họa các thực thể trong một hệ thống thông tin và mối quan hệ giữa các thực thể đó. ERD là mô hình khái niệm và đại diện về dữ liệu, được dùng để mô tả cơ sở hạ tầng của khung thực thể.

ERP: Hoạch định nguồn lực doanh nghiệp (Enterprise Resource Planning – ERP) là một quy trình được các doanh nghiệp (thường là doanh nghiệp hoạt động trong lĩnh vực sản xuất) sử dụng để quản lý và tích hợp các khía cạnh quan trọng trong hoạt động kinh doanh của mình. Hệ thống thông tin quản lý ERP tích hợp các phần như lập kế hoạch, mua sắm, hàng lưu kho, bán hàng, marketing, tài chính, và nhân sự.

https://en.wikipedia.org/wiki/Relational_database_management_system

Dữ liệu có cấu trúc chỉ bất kỳ dữ liệu nào nằm trong một trường cố định trong một hồ sơ hoặc tệp tin, bao gồm cả dữ liệu chứa trong các cơ sở dữ liệu quan hệ và các trang tính. Dữ liệu được lưu trữ trong

các cấu trúc dữ liệu được xác định trước và dữ liệu được cấu hình trước tương tự như vậy chủ yếu được truy cập và sử dụng bằng SQL – ngôn ngữ truy vấn có cấu trúc. SQL được sử dụng để truy cập các dữ liệu được tổ chức theo logic ở các bảng khác nhau thông qua việc sử dụng các khóa chung (hay các mối quan hệ). Trước năm 1995, việc lưu trữ và sử dụng dữ liệu có cấu trúc tương đối dễ dàng vì dữ liệu này có cấu trúc đã biết được xác định từ trước và khối lượng dữ liệu cũng tương đối thấp. Các công ty như Oracle Corporation và IBM đã cung cấp nền tảng, công nghệ, và công cụ để truy cập và phân tích loại dữ liệu này.

Trước đây, người ta có thể dễ dàng gọi bất kỳ dữ liệu nào được lưu trữ trong các bảng và hàng trong một cấu trúc có trật tự được xác định từ trước là dữ liệu nói chung – bây giờ, nó được gọi là dữ liệu có cấu trúc nhằm phân biệt với dữ liệu phi cấu trúc (chúng ta sẽ bàn đến chủ đề này ở phần sau). Quy mô, hình thức, và lưu lượng của dữ liệu có cấu trúc bao giờ cũng đúng với các kỳ vọng đã được xác lập từ trước; các ngôn ngữ lập trình SQL là sự lựa chọn đúng đắn để thực hiện các thao tác hoặc phân tích trên dữ liệu có cấu trúc.

Trước khi chuyển sang bàn về dữ liệu phi cấu trúc, thiết nghĩ cần lưu ý rằng hầu hết các dữ liệu trước thập niên 1990 đều được tạo ra bên trong các tổ chức và sử dụng nội bộ. Kể từ đó, các nguồn dữ

liệu đã thay đổi rất nhiều và hiện nay, một lượng lớn dữ liệu đang được ghi lại là những luồng dữ liệu ngoài, chủ yếu do khách hàng và các thiết bị mà họ sử dụng tạo ra, chẳng hạn điện thoại di động, và đến từ hoạt động sử dụng các cổng mạng xã hội mở như Facebook và Twitter.

DỮ LIỆU PHI CẤU TRÚC

Giống với trường hợp của dữ liệu có cấu trúc, dữ liệu phi cấu trúc cũng không phải một khái niệm mới mẻ – bạn có nhớ các tờ báo ngày xưa được in ra như thế nào không? Trước khi máy vi tính xuất hiện, trang báo được sắp chữ trên các loại máy thủ công theo trình tự ngược – đây quả là một nhiệm vụ vô cùng mệt mỏi nếu xét đến việc in ấn bây giờ diễn ra đơn giản như thế nào. Khác với ngày nay, thời đó các thợ sắp chữ cho báo không hề được cung cấp một cấu trúc cho trước nào về cách thức trình bày trang báo, ngoại trừ tiêu đề trang, thường cũng chính là tên gọi của tờ báo. Có ngày trang báo được dàn theo hai cột, có ngày lại chia thành ba cột, năm khối, và hai hình ảnh. Cấu trúc tờ báo thay đổi thường xuyên tùy thuộc vào số lượng nội dung, kiểu chữ, và hình ảnh được chọn dùng. Thách thức lớn đối với thợ sắp chữ là làm sao để khớp dữ liệu (nội dung tin tức) vào hình thức phù hợp nhất dựa trên cơ sở là cấu trúc của tờ báo trong

ngày hôm đó. Tương tự như vậy, dữ liệu phi cấu trúc có đủ mọi loại hình dạng và quy mô, nên việc lưu trữ, phân tích, và sử dụng dữ liệu phi cấu trúc giống như cách áp dụng với dữ liệu có cấu trúc là cả một vấn đề.

Khi khái niệm về cơ sở dữ liệu ra đời, hầu hết chúng ta đều chỉ tập trung vào khía cạnh dễ lưu trữ và sử dụng của dữ liệu, vì thế dữ liệu phi cấu trúc bị "bỏ qua". Vấn đề không phải là dữ liệu phi cấu trúc không hữu ích, mà là giá trị tạo ra từ dữ liệu có cấu trúc thường bền vững hơn trong hoàn cảnh và năng lực kỹ thuật đương thời. Lưu lượng dữ liệu phi cấu trúc gia tăng từ cuối thập niên 1990, khi nhiều người tiêu dùng bắt đầu tạo dữ liệu trên mạng internet – thông qua mạng lưới các blog, diễn đàn, cổng mạng xã hội, và ở các địa điểm cũng như thiết bị khác. Công nghệ mới đầy tính sáng tạo này cùng với sự thay đổi trong hành vi của người dùng đã mở ra một cơ hội lớn cho các tổ chức; dĩ nhiên, kèm theo đó là một thách thức. Việc sử dụng các công nghệ khi đó để lưu trữ và phân tích dữ liệu phi cấu trúc sẽ khả thi nếu khối lượng và tốc độ của luồng dữ liệu này không gia tăng theo cấp số nhân như vậy.

Ngày nay, hầu hết chúng ta đều sử dụng mạng xã hội và dùng nhiều thiết bị khác nhau để kết nối mạng internet. Khi người tiêu dùng đăng tải một tin nhắn lên Facebook, tải một hình ảnh lên Instagram, hay sử dụng bất kỳ ứng dụng di động nào, dấu chân số

của họ đều sẽ được ghi lại và lưu trữ ở nhiều địa điểm khác nhau. Ngoài ra, các công ty cũng bắt đầu lưu trữ và sử dụng dữ liệu liên quan đến các hoạt động sử dụng di động, dữ liệu cảm biến số, và các dạng dữ liệu nhật ký khác với hy vọng sẽ hiểu rõ hơn về hành vi của khách hàng/người tiêu dùng, thị trường, và các cơ hội.

Nhiều người cho rằng dữ liệu số phi cấu trúc, vốn là một yếu tố quan trọng tạo thành một phần lớn trong dữ liệu lớn, ra đời cùng với sự xuất hiện của Web 2.0, hay cuộc cách mạng về mạng xã hội – tuy nhiên, không nên cho rằng đây là một nhận định đúng hoàn toàn. Dữ liệu có cấu trúc bất ngờ, hay dữ liệu phi cấu trúc, trong hình hài kỹ thuật số bắt đầu xuất hiện dưới hình thức văn bản email và các file nhật ký khác, tiếp đến là vào thập niên 1990, khi xuất hiện các cổng thông tin thảo luận dành cho người tiêu dùng, hay các diễn đàn. Xuất phát điểm của hoạt động tạo nội dung từ đám đông là những cuộc trao đổi đa luồng giữa nhiều người tiêu dùng khác nhau. Về sau, các cổng mạng xã hội như MySpace, Facebook, Twitter... đã mang đến những cách thức dễ dàng hơn để một lượng lớn người dùng có thể tha hồ tạo, thảo luận, và chia sẻ nội dung với nhiều hình thức đa dạng, bao gồm văn bản, hình ảnh, đường dẫn, và video. Cùng lúc đó, mạng internet và các công nghệ di động cũng không

ngừng được cải thiện để tiếp cận với nhiều người hơn trên thế giới, từ đó càng thúc đẩy mạnh mẽ khối lượng dữ liệu phi cấu trúc theo nhiều cách khác nhau.

Theo ước đoán, có tới 90% dữ liệu trong bất kỳ tổ chức nào đều là dữ liệu phi cấu trúc. Và tốc độ tăng trưởng của dữ liệu phi cấu trúc trong các tổ chức cao hơn nhiều so với tốc độ tăng trưởng của dữ liệu có cấu trúc.

Ngoài nội dung do người tiêu dùng tạo ra trên các trang mạng xã hội và nhiều website khác, các hoạt động của người tiêu dùng trong không gian số cũng gia tăng đáng kể, từ đó tạo ra khối lượng dữ liệu khổng lồ cho các công ty hoạt động đằng sau – chẳng hạn, các dữ liệu về hành vi nhấp chuột trên website hay dữ liệu về thói quen sử dụng thiết bị di động, tất cả đều là những dữ liệu hữu ích giúp doanh nghiệp cải tiến sản phẩm và dịch vụ của mình. Nhưng thách thức thực sự không chỉ giới hạn ở việc xử lý bản chất phi cấu trúc (sự đa dạng) của luồng dữ liệu mới mẻ này; quy mô (khối lượng) và tốc độ (vận tốc) tạo ra dữ liệu cũng là những vấn đề đau đầu khiến nhiệm vụ này càng trở nên khó khăn hơn bao giờ hết. Thật may mắn, Google và Yahoo! đã nhìn ra được thách thức này từ sớm và bắt tay vào tìm cách giải quyết vấn đề.

Nhiều tổ chức tin rằng dữ liệu phi cấu trúc hàm chứa những thông tin có thể giúp họ đưa ra quyết

sách đúng đắn hơn. Điều đáng tiếc nằm ở chỗ, phân tích dữ liệu phi cấu trúc thường là một nhiệm vụ hết sức khó khăn. Các tổ chức có thể tìm đến sự hỗ trợ của nhiều giải pháp phần mềm có khả năng đọc văn bản phi cấu trúc và rút ra những thông tin có ý nghĩa. Lợi ích chính của các công cụ này là khả năng thu thập được các thông tin khả thi giúp doanh nghiệp thành công trong một môi trường cạnh tranh. Ví dụ về các "dữ liệu phi cấu trúc" bao gồm sách, nhật ký, tài liệu, siêu dữ liệu, hồ sơ y khoa, file âm thanh, video, dữ liệu tương tự, hình ảnh, file, và văn bản phi cấu trúc, chẳng hạn phần nội dung của email, trang web, hay tài liệu xử lý văn bản.

Do khối lượng dữ liệu phi cấu trúc đang phát triển với tốc độ rất nhanh, nên nhiều doanh nghiệp cũng bắt đầu tìm đến các giải pháp công nghệ nhằm giúp họ quản lý và lưu trữ dữ liệu phi cấu trúc tốt hơn. Đó có thể là các giải pháp phần cứng hoặc phần mềm cho phép họ tận dụng hiệu quả nhất dung lượng lưu trữ của mình.

Ngoài dữ liệu có cấu trúc và phi cấu trúc, còn có một loại dữ liệu thứ ba thực ra là sự kết hợp của cả hai loại trên, dữ liệu bán cấu trúc. Dữ liệu bán cấu trúc là các thông tin không nằm trong cơ sở dữ liệu quan hệ nhưng có một số thuộc tính có tổ chức khiến việc phân tích chúng dễ dàng hơn. Ví dụ về dữ

liệu bán cấu trúc bao gồm các tài liệu XML và các cơ sở dữ liệu NoSQL.

Thuật ngữ Dữ liệu lớn có mối liên hệ mật thiết với dữ liệu phi cấu trúc. Dữ liệu lớn chỉ những tập dữ liệu cực lớn, khó thực hiện phân tích bằng các công cụ truyền thống. Dữ liệu lớn là sự kết hợp của tất cả các loại hình cấu trúc dữ liệu, từ dữ liệu phi cấu trúc, dữ liệu bán cấu trúc đến siêu dữ liệu; tuy vậy, theo ước tính, có đến 90% dữ liệu lớn là dữ liệu phi cấu trúc. Nhiều công cụ được thiết kế để phân tích dữ liệu lớn có thể xử lý dữ liệu phi cấu trúc.

Trong ngữ cảnh này, dữ liệu phi cấu trúc cũng bao gồm dữ liệu bán cấu trúc và siêu dữ liệu. Dữ liệu bán cấu trúc là dữ liệu có cấu trúc một phần, chẳng hạn dữ liệu email. Một số phần được xác định từ trước, chẳng hạn như địa chỉ email của người gửi và người nhận, thời gian gửi và nhận email... luôn luôn tuân theo một hình thức giống nhau, nhưng nội dung của email có thể bao gồm văn bản đơn giản, đường dẫn liên kết, hình ảnh, và các loại tệp tin đính kèm khác nhau. Tương tự, siêu dữ liệu, tức dữ liệu về dữ liệu chẳng hạn dữ liệu viết trong các thẻ XML, cũng có thể được cấu hình trước một phần. Ở đây chúng ta không cần đi sâu tìm hiểu về dữ liệu bán cấu trúc và siêu dữ liệu, bởi phần lớn chúng đều có thể được xếp vào nhóm lớn là dữ liệu phi cấu trúc, điều đó phù hợp với đề tài của cuốn sách này.

Nhiều lĩnh vực ngành nghề mang tính truyền thống như nông nghiệp, bảo hiểm... hiện cũng đang tích cực tạo ra những khối lượng dữ liệu khổng lồ, sau đó được lưu trữ và phân tích nhằm rút ra những kiến giải ngầm đằng sau.

DỮ LIỆU LỚN

"Dữ liệu lớn" là thuật ngữ bao trùm dùng để chỉ loại hình dữ liệu khổng lồ, đa dạng, có tốc độ tăng trưởng rất nhanh, vì quá lớn nên khó có thể lưu trữ trong các máy chủ kiểu cũ, vì mang tính phi cấu trúc nên khó có thể phân tích bằng các hệ thống BI kiểu cũ, và vì quá nhanh nên khó có thể giữ trong một kho dữ liệu tĩnh như trước đây. Giá trị của dữ liệu lớn sẽ bị hạn chế rất nhiều nếu doanh nghiệp không thể phân tích nó và rút ra những kiến giải khả thi nhằm xây dựng và triển khai các chiến lược hữu hiệu.

Mặc dù thuật ngữ "Dữ liệu lớn" tương đối mới mẻ, song hành vi thu thập và lưu trữ nhiều khối lượng lớn thông tin để phục vụ quá trình phân tích đã tồn tại từ rất lâu rồi. Khái niệm này bắt đầu được phổ biến vào đầu thập niên 2000, khi chuyên gia phân tích Doug Laney đưa ra một định nghĩa về dữ liệu lớn gồm ba chữ V và hiện nay đã được công nhận rộng rãi:

Volume (Khối lượng): Các tổ chức thu thập dữ liệu từ nhiều nguồn khác nhau, bao gồm các giao dịch

kinh doanh, mạng xã hội, và thông tin từ hệ thống thiết bị cảm biến hay dữ liệu từ máy đến máy. Khác với trước kia, khi việc lưu trữ khối lượng dữ liệu lớn là cả một vấn đề, các công nghệ hiện đại như Hadoop đã làm giảm đáng kể chi phí nguồn lực và quá trình tải.

Velocity (Vận tốc): Luồng dữ liệu đổ về với tốc độ cao chưa từng có trong lịch sử, và cần xử lý kịp thời luồng dữ liệu này. Từ điện thoại di động và các thẻ RFID[1] đến thiết bị cảm biến, internet và hệ thống đo lường thông minh, có rất nhiều nguồn dữ liệu khác nhau cùng hoạt động, từ đó càng làm gia tăng nhu cầu phải xử lý lượng lớn dữ liệu gần như theo thời gian thực.

Variety (Sự đa dạng): Dữ liệu có nhiều hình dạng khác nhau – từ dữ liệu có cấu trúc dạng số trong các cơ sở dữ liệu truyền thống đến dữ liệu phi cấu trúc trong các tài liệu văn bản, email, video, file âm thanh, bảng báo giá chứng khoán, và các giao dịch tài chính.

Ở một số trường hợp, dữ liệu lớn còn có thêm hai khía cạnh nữa:

Variability (Tính khả biến): Ngoài vận tốc và sự đa dạng ngày càng gia tăng của dữ liệu, các luồng

---

[1]RFID (Radio Frequency Identification): Nhận dạng qua tần số vô tuyến, một công nghệ kết nối sóng vô tuyến để tự động xác định và theo dõi các thẻ nhận dạng gắn vào vật thể. (ND)

chảy dữ liệu còn có thể mang tính thiếu nhất quán cao độ, thi thoảng lại có những đợt cao trào. Trên mạng xã hội đang có xu hướng gì nổi bật? Việc quản lý các lượng tải dữ liệu cao trào theo ngày, theo mùa, và theo sự kiện có thể là một thách thức rất lớn, đặc biệt với dữ liệu phi cấu trúc.

Complexity (Sự phức tạp): Dữ liệu ngày nay đến từ nhiều nguồn khác nhau, nên rất khó thực hiện kết nối, khớp, làm sạch, và biến đổi dữ liệu ở các hệ thống. Tuy nhiên, cần kết nối và thiết lập mối tương quan giữa các mối quan hệ, hệ thống thứ bậc, và nhiều mối liên kết dữ liệu khác nhau, nếu không dữ liệu của bạn có thể sẽ nhanh chóng phát triển vượt quá tầm kiểm soát.

Khối lượng dữ liệu hiện đang được tạo ra và lưu trữ trên toàn cầu đã vượt quá khả năng hình dung của chúng ta, và sẽ còn tiếp tục gia tăng. Điều đó nghĩa là càng có nhiều tiềm năng rút ra những kiến giải quan trọng từ các thông tin của doanh nghiệp – nhưng mới chỉ có một phần nhỏ dữ liệu được phân tích. Điều này có ý nghĩa gì đối với các doanh nghiệp? Họ có thể làm gì để sử dụng tốt hơn nguồn thông tin thô vẫn đều đặn chảy về tổ chức của mình hằng ngày?

Tầm quan trọng của dữ liệu lớn không nằm ở việc bạn có trong tay bao nhiêu dữ liệu, mà là bạn làm gì với nó. Bạn có thể lấy dữ liệu từ bất kỳ nguồn nào rồi

tiến hành phân tích để tìm ra những câu trả lời nhằm 1) cắt giảm chi phí, 2) giảm thời gian thực hiện, 3) phát triển sản phẩm mới và tối ưu hóa sản phẩm/dịch vụ, và 4) xây dựng quy trình ra quyết định thông minh. Khi kết hợp dữ liệu lớn với công nghệ phân tích tiên tiến, bạn có thể hoàn thành những nhiệm vụ liên quan đến kinh doanh, ví dụ:

1. Xác định nguyên nhân gốc rễ của các thất bại, vấn đề, và sai sót gần như theo thời gian thực.
2. Tạo ra các coupon khuyến mãi tại thời điểm mua hàng dựa trên thói quen mua sắm của khách hàng.
3. Tính toán lại mức độ rủi ro trong các danh mục đầu tư chỉ trong vài phút.
4. Xác định các hành vi sai trái trước khi chúng gây ảnh hưởng đến tổ chức.

Hầu hết các cuốn sách và công trình nghiên cứu bàn về Dữ liệu và/hoặc Phân tích đều bỏ qua sự khác biệt cốt lõi mang tính nền tảng giữa dữ liệu mới ngày nay (dữ liệu lớn/dữ liệu phi cấu trúc) và dữ liệu cũ trong quá khứ (dữ liệu có cấu trúc), và ở đây, chúng tôi không nói đến những thuộc tính hiển nhiên đã đề cập trong phần trước. Về cơ bản, các nhà lãnh đạo, quản lý trong doanh nghiệp và đội ngũ kỹ thuật của

họ đã tạo ra sự khác biệt này, từ lâu trước khi dữ liệu được các doanh nghiệp tạo ra, lưu trữ, và phân tích.

Điểm khác biệt là *sự kiểm soát đối với nguồn dữ liệu* – các doanh nghiệp trước đây đã tạo ra một môi trường các quy trình nghiệp vụ và hoạt động thu thập dữ liệu vận hành theo cơ chế đóng, chịu sự kiểm soát nghiêm ngặt đến mức trước cả khi quá trình thu thập dữ liệu bắt đầu, doanh nghiệp đã biết họ sẽ thu thập loại hình dữ liệu hay thông tin gì, từ đâu, và họ sẽ lưu trữ cũng như xử lý các dữ liệu này như thế nào. Các chuyên gia hệ thống đã xác định từ trước rằng họ sẽ thu thập "dữ liệu gì", "từ đâu", và "theo hình thức nào", và sau đó phân tích chúng cùng với cấu trúc, quy mô, và lưu lượng mà họ đặt ra. Thực ra, những khía cạnh như độ dài của một giá trị dữ liệu cũng được xác định từ trước để mọi dữ liệu đầu vào đều tuân theo các tiêu chí giống nhau được định sẵn ngay từ đầu trong quá trình thiết kế hệ thống và cơ sở dữ liệu. Đây là quy trình mà các doanh nghiệp bấy lâu nay vẫn làm theo, có lẽ bởi những hạn chế về mặt kỹ thuật. Nhiều doanh nghiệp còn thiết lập ranh giới xung quanh mình, và họ hầu như chỉ xử lý các tập dữ liệu vốn rất quen thuộc đến từ những nguồn đã biết trong một môi trường khép kín.

Trong nhiều trường hợp, khi dữ liệu đã biết được xác định từ trước là những dữ liệu duy nhất tồn tại

trong tổ chức, các vị lãnh đạo gặp phải khó khăn khi tìm kiếm những kiến giải chính từ các tập dữ liệu na ná nhau. Do đó, vô hình trung chúng lại hạn chế khả năng đổi mới sáng tạo trong doanh nghiệp cũng như khả năng lĩnh hội được những kiến thức sâu sắc hơn về hành vi của khách hàng và các điều kiện kinh doanh. Trong những trường hợp như vậy, các nhà quản lý cũng phải phụ thuộc rất nhiều vào đội ngũ kỹ thuật mới có thể tìm hiểu về tình hình kinh doanh.

Chúng tôi cho rằng chính cách thu thập dữ liệu trong môi trường kiểm soát cao độ như thế đã hạn chế khả năng khai thác các cơ hội mới, trí tưởng tượng, sự đổi mới sáng tạo, và triển vọng kiến tạo giá trị.

Đối lập với dữ liệu cũ, dữ liệu mới đã phá vỡ những ranh giới này, và các doanh nghiệp cũng đã bắt đầu tiếp nhận (dù chủ động hay bị động) lượng lớn dữ liệu từ các nguồn khác nhau, phần lớn đều không nằm trong tầm kiểm soát của họ, và phát triển cách tiếp cận thụ động để xử lý chúng. Một số doanh nghiệp đã nhân cơ hội này để thúc đẩy hoạt động đổi mới sáng tạo và thực hiện những bước tiến lớn trong khi các doanh nghiệp khác còn đang thử nghiệm hoặc đắn đo. Chúng ta gọi đây là "dữ liệu mới" vì dữ liệu có khối lượng lớn (phi cấu trúc) không dễ xác định như dữ liệu có cấu trúc cũ. Và giờ đây, khi học cách xử lý

luồng dữ liệu mới đến từ những nguồn mới, chúng ta cũng đồng thời có thể xử lý dữ liệu phi cấu trúc do các hệ thống nội bộ tạo ra, chẳng hạn các file nhật ký hệ thống vốn trước đây vẫn bị các hệ thống cũ bỏ qua.

Dưới đây là một số ví dụ về các loại hình dữ liệu mới có thể thu thập ngay từ bây giờ. Thực ra, đây cũng là những dữ liệu mà các doanh nghiệp hiện đang lưu trữ và phân tích nhằm rút ra thêm nhiều giá trị nữa cho tất cả các bên liên quan.

Dữ liệu khách hàng

Doanh nghiệp có thể thu thập được rất nhiều thông tin, bằng cách trực tiếp hoặc gián tiếp, từ khách hàng để qua đó cải thiện các sản phẩm/dịch vụ của họ – dưới đây là một số ví dụ và các nguồn dữ liệu khả thi.

1. Thông tin phản hồi về sản phẩm hoặc dịch vụ – đây là đầu mục thông tin tối cần thiết mà mọi doanh nghiệp đều nên quan tâm. Thông tin phản hồi của khách hàng có thể được thu thập theo nhiều cách thức khác nhau:

    a. Khảo sát khách hàng: Khảo sát sau mua hàng hoặc thư/email phản hồi là một trong nhiều cách thu thập ý kiến phản hồi của khách hàng. Các công ty như TrustPilot và Reevoo cung cấp dịch vụ thu thập thông tin phản hồi cho

các doanh nghiệp. Các công ty này cũng chia sẻ thông tin phản hồi của khách hàng nhằm giúp các khách hàng tiềm năng có thêm thông tin để đưa ra quyết định hợp lý dựa trên trải nghiệm của khách hàng trước đó.

b. Lắng nghe mạng xã hội: Khách hàng chia sẻ rất cởi mở và thẳng thắn về các sản phẩm, dịch vụ, và thương hiệu trên mạng xã hội – doanh nghiệp có thể lắng nghe và tham gia vào các cuộc trao đổi này. Bạn có thể tìm thấy rất nhiều ví dụ trên Twitter hay Facebook, trong đó người tiêu dùng hoặc khách hàng đưa ra những ý kiến phản hồi tiêu cực về nhiều sản phẩm khác nhau, từ phần mềm chống virus đến điện thoại di động.

Nhiều thương hiệu coi hoạt động lắng nghe mạng xã hội là một công cụ chủ động thu thập thông tin phản hồi của khách hàng nhằm thực hiện những bước điều chỉnh/cải thiện cần thiết. Domino's Pizza là một thương hiệu như vậy, và chúng ta sẽ cùng nhau bàn đến trường hợp này ở phần sau của cuốn sách.

c. Ý kiến phản hồi trực tiếp bằng miệng hoặc viết tay do khách hàng cung cấp ngay tại cửa hàng, qua bưu điện hoặc email.

d. Những ý kiến phàn nàn do khách hàng gửi qua website hoặc email.

e. Thông tin phản hồi được chia sẻ ở các diễn đàn thảo luận.

2. Dữ liệu tại điểm bán (Point of Sale – PoS): Thông thường, hệ thống điểm bán lẻ sẽ bao gồm một quầy thanh toán tiền mặt với máy tính, màn hình, ngăn kéo quầy thu ngân, máy in hóa đơn, màn hình hiển thị cho khách hàng, và máy quét mã vạch; ngoài ra, đại đa số các hệ thống PoS cũng được trang bị đầu đọc thẻ ghi nợ/thẻ tín dụng.

Các hệ thống tại điểm bán thu thập dữ liệu về sản phẩm và chất lượng sản phẩm, tần suất mua sản phẩm... từ mỗi khách hàng, và những dữ liệu này sau đó có thể được sử dụng trong quá trình phân tích và dự đoán hành vi mua sắm của khách hàng. Cũng có thể sử dụng những dữ liệu này để phục vụ việc lập hồ sơ khách hàng, hoặc hỗ trợ doanh nghiệp xây dựng mối quan hệ riêng với từng khách hàng.

Có nhiều ví dụ cho thấy dữ liệu PoS có thể bổ sung giá trị cho doanh nghiệp như thế nào. Chẳng hạn, chuỗi siêu thị Target ở Mỹ sử dụng dữ liệu PoS để dự đoán nhu cầu của những người sắp làm cha mẹ bằng cách phân tích xu

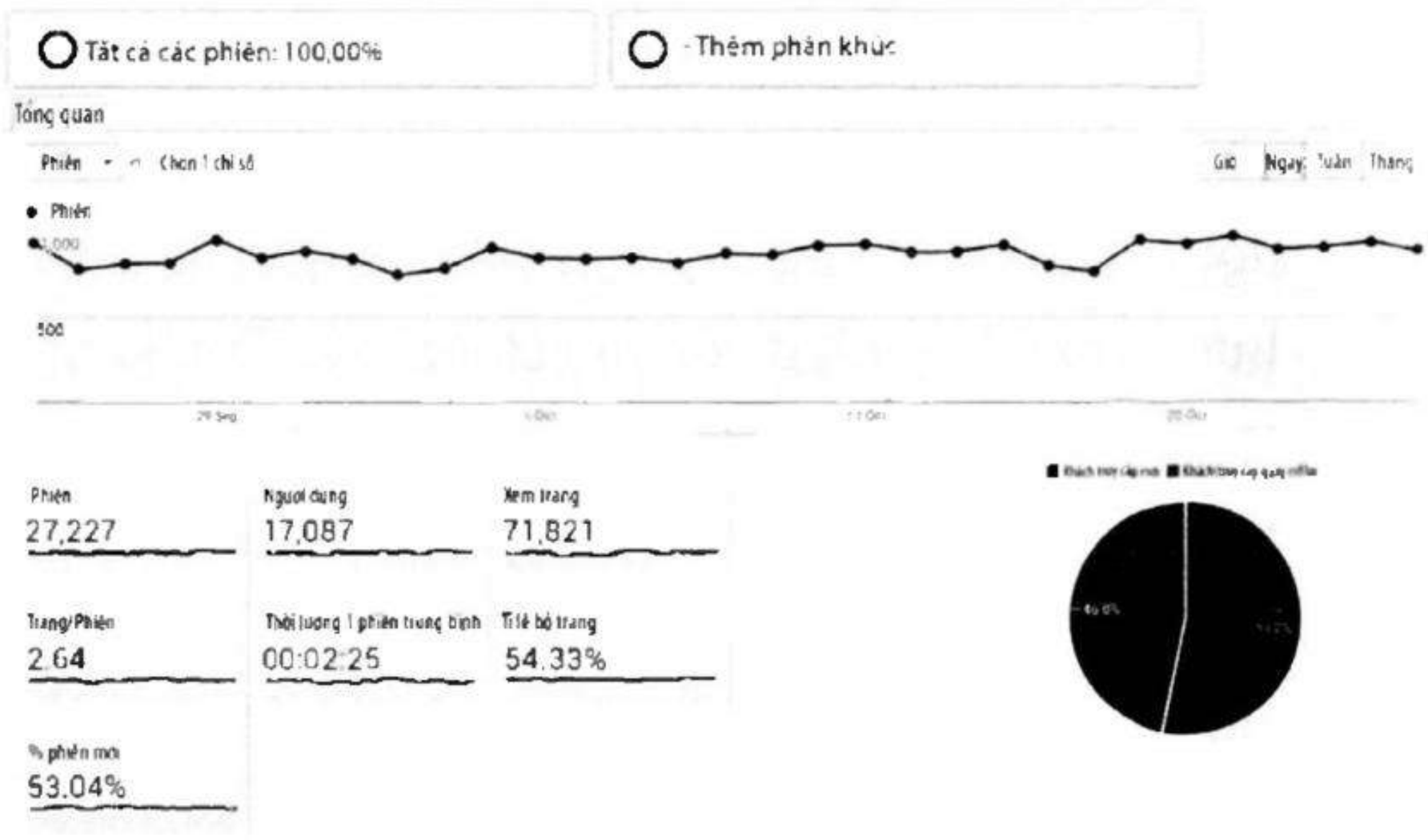
hướng mua sắm của họ dựa trên lịch sử mua sắm. Chuỗi siêu thị Tesco ở Anh cũng rất nổi tiếng với việc sử dụng dữ liệu PoS và dữ liệu từ thẻ trung thành để gia tăng doanh số bán hàng, cải tiến quy trình, và cung cấp dịch vụ tốt nhất cho khách hàng.

3. Dữ liệu tại cửa hàng hoặc dữ liệu sử dụng website: Hiện nay, công nghệ có thể giúp doanh nghiệp xác định các hoạt động của khách hàng/khách truy cập trên website hoặc trong các cửa hàng vật lý của họ. Với sự hỗ trợ của dữ liệu này, doanh nghiệp có thể thực hiện những thay đổi trong việc cung cấp sản phẩm/dịch vụ theo hướng phù hợp nhất với nhu cầu của khách hàng. Dựa trên thông tin về hoạt động của người dùng thu thập được, doanh nghiệp có thể điều chỉnh để hoàn thiện thiết kế cũng như bố cục trình bày của cửa hàng và website.

Google Analytics là một trong những dịch vụ phổ biến nhất đối với các nhà bán lẻ trong lĩnh vực thương mại trực tuyến khi họ muốn theo dõi các hoạt động diễn ra trong cửa hàng trực tuyến của mình. Với sự hỗ trợ của Google Analytics, doanh nghiệp có thể thu thập được rất nhiều thông tin giá trị, chẳng hạn lhách truy cập cũ/mới, các trang được đọc

nhiều nhất, các phần phổ biến trên website, các sản phẩm được ưa chuộng... Thông tin này giúp doanh nghiệp cải tiến giao diện, cách thức trình bày, thông điệp, và sản phẩm/dịch vụ của mình, qua đó khiến khách hàng hài lòng hơn, tỉ lệ chuyển đổi thành khách hàng cao hơn, và nhiều lợi ích khác nữa (xem Hình 2.1).

Hình 2.1. Một bảng trình bày dữ liệu về các lượt truy cập website



Tương tự như Google Analytics, Retail Next là một trong số ít các công ty đã phát triển công nghệ để giúp các doanh nghiệp theo dõi và học hỏi từ khách hàng/khách ghé thăm cửa hàng vật lý của họ. Các thông tin thu thập được có thể giúp doanh nghiệp thực hiện nhiều thay đổi, chẳng hạn điều chỉnh bố

cục trình bày ở cửa hàng, hoặc thay đổi vị trí dán tấm áp phích của các chiến dịch.

Dữ liệu về nhân viên

Các dữ liệu liên quan đến đội ngũ nhân viên từ lâu đã mang đến giá trị lớn cho doanh nghiệp, và trước đây, dữ liệu loại này chỉ tồn tại (hoặc được xác định) ở một số nguồn ít ỏi. Ngày nay, các doanh nghiệp có thể tiếp cận nhiều loại hình dữ liệu từ nhiều nguồn khác nhau, có liên quan đến nhiều chiều kích khác nhau trong mối quan hệ giữa nhân viên và đơn vị tuyển dụng.

1. Dữ liệu về sự hài lòng của nhân viên: Có nhiều cách thu thập thông tin về sự hài lòng và ý kiến phản hồi của nhân viên về tổ chức nơi họ làm việc. Từ các bảng khảo sát và ý kiến phản hồi trực tiếp bằng miệng đến các cuộc trao đổi trên mạng xã hội, doanh nghiệp có thể thu thập thông tin về ý kiến và quan điểm của nhân viên để xem họ hài lòng tới mức nào khi làm việc cho doanh nghiệp, và có thể làm gì để khiến họ hài lòng.
2. Dữ liệu về hoạt động/năng suất của nhân viên: Các doanh nghiệp ngày nay được tiếp cận với nhiều phương tiện thu thập và phân

tích dữ liệu liên quan đến năng suất làm việc của nhân viên. Từ những chiếc điện thoại di động do công ty sở hữu và phát cho nhân viên sử dụng, hệ thống email, đến việc cấp quyền truy cập vào mạng riêng ảo VPN[1], doanh nghiệp sẽ có thể thu thập dữ liệu về nhân viên và đưa ra những hành động dựa trên thông tin rút ra được từ các dữ liệu đó.

Chẳng hạn, từ các báo cáo đội ngũ bán hàng cập nhật hằng ngày, từ dữ liệu theo dõi về vị trí xe và các hoạt động trên điện thoại thông minh của họ, doanh nghiệp có thể biết được nhân viên của mình sử dụng thời gian đi gặp khách hàng tiềm năng có hiệu quả hay không. Ngoài ra, bạn cũng có thể thiết lập mô hình tính toán sự chênh lệch giữa tỉ lệ chuyển đổi thành khách hàng trên thực tế so với tỉ lệ ước tính của nhân viên bán hàng. Mô hình này (vốn có thể được xác định trong bản thân dữ liệu) sau đó có thể được liên hệ với các giao dịch chắc chắn sẽ diễn ra trong thời gian tới, qua đó dự đoán được doanh số bán hàng tương lai và doanh thu vào sổ.

---

[1]Mạng riêng ảo VPN: Một mạng riêng, thường dùng để kết nối các máy tính của một tổ chức lại với nhau thông qua mạng internet công cộng. (ND)

Năm 2013, CEO mới bổ nhiệm của Yahoo! là Marissa Mayer trở thành tâm điểm chú ý (và gây tranh cãi một chút) khi bà ra lệnh cấm toàn bộ nhân viên của công ty làm việc từ xa. Bà bị chỉ trích rất nhiều vì nước cờ này, nhưng chúng tôi muốn bạn chú ý đến phương pháp bà sử dụng để đi đến quyết định đó. Giống như nhiều công ty công nghệ khác, các nhân viên của Yahoo! khi làm việc ở nhà sẽ phải đăng nhập từ xa vào mạng cá nhân an toàn của họ, do đó, khi họ không làm việc từ xa, người ta sẽ biết họ đang làm việc ở văn phòng. Marissa để ý thấy rằng vào những thời điểm bãi đỗ xe của công ty tương đối vắng, nhiều nhân viên lại không đăng nhập từ xa vào hệ thống của công ty. Dựa trên hai thông tin này, bà kết luận rằng nhiều nhân viên có thể đã xin làm việc ở nhà nhưng lại không làm việc, vì vậy bà quyết định hủy bỏ hoàn toàn chính sách làm việc từ xa nhằm cải thiện năng suất. Vấn đề nằm ở chỗ, bà có thể truy cập vào các dữ liệu giúp bà nhìn ra bức tranh lớn.

3. Các dữ liệu về sở thích và khả năng sáng tạo của nhân viên: Những cuộc trao đổi của nhân viên trên các diễn đàn trong và ngoài công ty, trên mạng xã hội, và trong các email cũng có thể cung cấp rất nhiều thông tin về mối quan

tâm và kiến thức chuyên môn của họ. Các nhà quản lý và lãnh đạo trong doanh nghiệp có thể sử dụng các thông tin này để tìm ra những người có tinh thần cách tân, những nhà lãnh đạo tư tưởng, và những người có khả năng gây ảnh hưởng nhằm hỗ trợ và quảng bá cho thương hiệu của công ty. Có thể dễ dàng sắp xếp dữ liệu loại này bên trong tổ chức với sự hỗ trợ của đội ngũ IT.

Khái niệm tương đối mới mẻ "người ủng hộ là nhân viên" ra đời từ khoảng năm 2012. Ý tưởng ở đây là doanh nghiệp cung cấp cho nhân viên các thông điệp marketing để nhân viên tận dụng mạng xã hội của họ nhằm lan truyền thông tin về thương hiệu công ty. Doanh nghiệp phân tích dữ liệu từ các hoạt động trên mạng xã hội của nhân viên, xác định ra những người có tầm ảnh hưởng lớn nhất trong số đó, và sử dụng họ để quảng bá các thông điệp về công ty trên mạng xã hội rộng lớn. Sự ủng hộ của nhân viên sẽ giúp các công ty xây dựng được một thương hiệu mạnh mẽ hơn, đồng thời góp phần thu hút được nhân tài đến với công ty.

Dữ liệu về tình hình cạnh tranh

Trước đây, việc phân tích đối thủ cạnh tranh được thực hiện dựa trên dữ liệu thị trường lấy từ các cơ quan

chính phủ (chẳng hạn Cơ quan Thuế và Hải quan hay Cục Quản lý Doanh nghiệp), các hãng nghiên cứu thị trường, hay các bộ phận nghiên cứu trong nội bộ doanh nghiệp. Trong suốt nhiều thập niên, việc giám sát tin tức về đối thủ cạnh tranh được thực hiện bằng phương pháp thủ công. Ngày nay, có thể sử dụng công nghệ để giám sát thương hiệu bằng cách đưa các ứng dụng phần mềm "bò" vào hàng triệu trang web, mạng xã hội, và các cổng thông tin trực tuyến khác và tự động tạo ra những luồng dữ liệu khổng lồ.

Với sự hậu thuẫn của đổi mới sáng tạo trong công nghệ, khách hàng ngày nay cũng bắt đầu tham gia vào quá trình tạo nguồn dữ liệu cho hoạt động phân tích thương hiệu và đối thủ cạnh tranh – khách hàng công khai thể hiện quan điểm của họ về các sản phẩm và dịch vụ trên mạng xã hội, các diễn đàn trao đổi, và các cổng thông tin chuyên đăng tải những đánh giá về sản phẩm, và họ cần rất ít thời gian để thực hiện hoạt động này. Ngoài ra, vì những cuộc trao đổi đó diễn ra trong môi trường độc lập, nên nhận định của khách hàng hoặc người dùng về sản phẩm hoặc dịch vụ thường sát với thực tế hơn.

TÓM TẮT

Giống với bất kỳ khía cạnh nào khác trong cuộc sống, quyết định đúng vào đúng thời điểm có thể mang lại

một tác động lớn đến sự phát triển của doanh nghiệp. Để tránh sai sót hoặc sơ sẩy, các doanh nghiệp cố gắng ra quyết định dựa trên dữ liệu và lý trí ở mức tối đa. Chúng tôi từng chứng kiến một số quyết định thông minh ra đời dựa trên sự hậu thuẫn của dữ liệu cùng tác động mà chúng mang lại, và thấy dữ liệu có đóng góp to lớn như thế nào đến sự phát triển của doanh nghiệp. Chúng tôi cũng nhìn lại chặng đường đã qua của dữ liệu và các mối liên hệ với kinh doanh – tình báo doanh nghiệp và các quyết định dựa trên dữ liệu không phải khái niệm gì mới mẻ, nhưng trước đây chúng không được các nhà quản lý ưa chuộng như bây giờ. Một trong những lý do giải thích cho hiện tượng này là sự ra đời của các công nghệ hiện đại và khả năng tiếp cận với những dữ liệu đúng đắn và có ý nghĩa vào đúng thời điểm. Trong chương này, trước khi bàn đến dữ liệu lớn và các loại hình/hợp phần của nó, chúng tôi đã đề cập đến dữ liệu, khoa học phân tích, thống kê, và điện toán. Chúng tôi cũng điểm nhanh những hợp phần dữ liệu thiết yếu, chẳng hạn kho dữ liệu, RDBMS, khai phá dữ liệu, và tình báo doanh nghiệp. Tiếp đến, chúng tôi nói về Dữ liệu lớn và các hợp phần của nó – bao gồm dữ liệu có cấu trúc và phi cấu trúc – và giải thích vì sao trong thời gian gần đây dữ liệu phi cấu trúc lại có sự tăng trưởng mạnh mẽ đến vậy, sau đó là một số ví dụ chỉ ra rằng dữ liệu phi cấu trúc đang được sử dụng trong một số

khía cạnh chính của đời sống kinh doanh như khách hàng, nhà cung cấp, và nhân viên. Với sự ra đời của các công nghệ dữ liệu lớn như Hadoop và Hive, có thể khẳng định rằng chúng ta đã tiến một bước khá xa, nhưng vẫn còn nhiều không gian cho các hoạt động phát minh và cải thiện ở lĩnh vực này trong tương lai. Ở các chương tiếp theo, chúng tôi sẽ bàn đến những khái niệm về kiến tạo giá trị trong doanh nghiệp, đồng thời chỉ ra vì sao có thể coi dữ liệu nói chung (và dữ liệu lớn nói riêng) là một lợi thế lớn đối với mọi doanh nghiệp trong thời đại này.

3

KIẾN TẠO GIÁ TRỊ TỪ DỮ LIỆU LỚN

Từ *giá trị* trong cách nói chuyện hằng ngày và trong bối cảnh kinh doanh có thể mang nhiều ý nghĩa khác nhau. Do đó, cần phải quyết định và định nghĩa về tầm nhìn mà chúng ta sẽ áp dụng trong cuốn sách này, sau đó mới đi vào tìm hiểu xem dữ liệu lớn có thể tác động ra sao đến giá trị. Từ *giá trị* xuất phát từ từ *valere* trong tiếng Latin và có thể hiểu là *tầm quan trọng, giá trị hay tính hữu dụng của một thứ gì đó.* Từ góc độ tiền bạc, nó thường có nghĩa là *giá trị vật chất hoặc giá trị tiền bạc của một thứ gì đó.* Trong cuốn sách này, chúng tôi đứng ở góc nhìn của tổ chức với đối tượng độc giả mục tiêu là những nhà quản lý trong các tổ chức thương mại, muốn kiến tạo giá trị theo khía cạnh tiền bạc. Có một thực tế không chỉ đúng trong mối liên hệ với dữ liệu lớn mà còn đúng trong nhiều khía cạnh khác của đời sống kinh doanh

là không phải lúc nào cũng có thể đo lường giá trị trực tiếp theo góc độ tiền bạc, do đó đôi khi phải cần đến các phương pháp đo lường mang tính đại diện để đánh giá về giá trị. Trong chương này, chúng tôi sẽ bàn đến những yếu tố thúc đẩy nói chung đối với quá trình kiến tạo giá trị thông qua một mẫu Báo cáo Kết quả Kinh doanh và Báo cáo Tình hình Tài chính cho một doanh nghiệp thương mại giả tưởng. Nhưng trước hết, xin lưu ý rằng đây là cuốn sách viết về dữ liệu lớn, không phải sách về tài chính và kế toán, vì thế, những độc giả đã có kiến thức tốt về tài chính và kế toán có thể đọc lướt qua chương này. Chương này đưa ra những ví dụ chung để dẫn nhập vào vấn đề kiến tạo giá trị từ dữ liệu lớn, và các chương tiếp theo sẽ bàn đến công nghệ cùng những trường hợp ví dụ điển hình liên quan đến hoạt động đầu tư vào dữ liệu lớn.

CÁC YẾU TỐ THÚC ĐẨY GIÁ TRỊ TRONG TỔ CHỨC THƯƠNG MẠI

Theo thời gian, giá trị được tạo ra nếu doanh thu vượt quá chi phí tạo doanh thu. Đây là một kiến thức cơ bản về kinh doanh và khi bàn đến chuyện giá trị được tạo ra như thế nào, có lẽ chúng ta cũng nên tìm về những điều cơ bản. Doanh thu là huyết mạch của mọi hoạt động kinh doanh và chúng ta cần phải chú ý đến nó. Tuy vậy, nói bao giờ cũng dễ hơn làm. Đặc điểm

nổi bật của doanh thu là không chắc chắn, và những gì diễn ra trên thị trường có thể thay đổi kỳ vọng của chúng ta về doanh thu tương lai. Ở khía cạnh ngược lại, chi phí thường – nhưng không phải luôn luôn – mang tính chắc chắn tương đối. Cũng cần lưu ý rằng doanh nghiệp có sự kiểm soát ở mức độ nhất định đối với chi phí. Như vậy, cơ sở để chúng ta suy nghĩ về vấn đề kiến tạo giá trị là doanh thu không chắc chắn và chi phí tương đối chắc chắn. Câu hỏi cần lưu tâm ở đây là mức độ ảnh hưởng của dữ liệu lớn đối với sự không chắc chắn của doanh thu và tác động của nó đến chi phí. Bao hàm trong chi phí không chỉ có khoản bỏ ra để tạo doanh thu, mà còn là chi phí của nguồn vốn đầu tư dùng để tạo doanh thu. Hệ số lợi nhuận trên vốn đầu tư (return on invested capital – ROIC) có hai hợp phần: số lượng vốn đầu tư tương ứng với biên lợi nhuận (thu nhập/doanh thu). Giá trị được tạo ra nếu biên lợi nhuận gia tăng mà vốn đầu tư không tăng, hoặc nếu vốn đầu tư giảm trong khi biên lợi nhuận giữ nguyên. Ở góc độ khác, biên lợi nhuận có hai hợp phần thúc đẩy giá trị: một là doanh thu, và một là chi phí. Vốn đầu tư cũng có hai hợp phần: một là vốn lưu động dùng trong các hoạt động của doanh nghiệp, chẳng hạn sản phẩm lưu kho hoặc sự chênh lệch về thời gian từ lúc nhận được thanh toán của khách hàng đến lúc thanh toán cho các nhà cung cấp; hợp phần thứ hai là các tài sản cố định dùng để

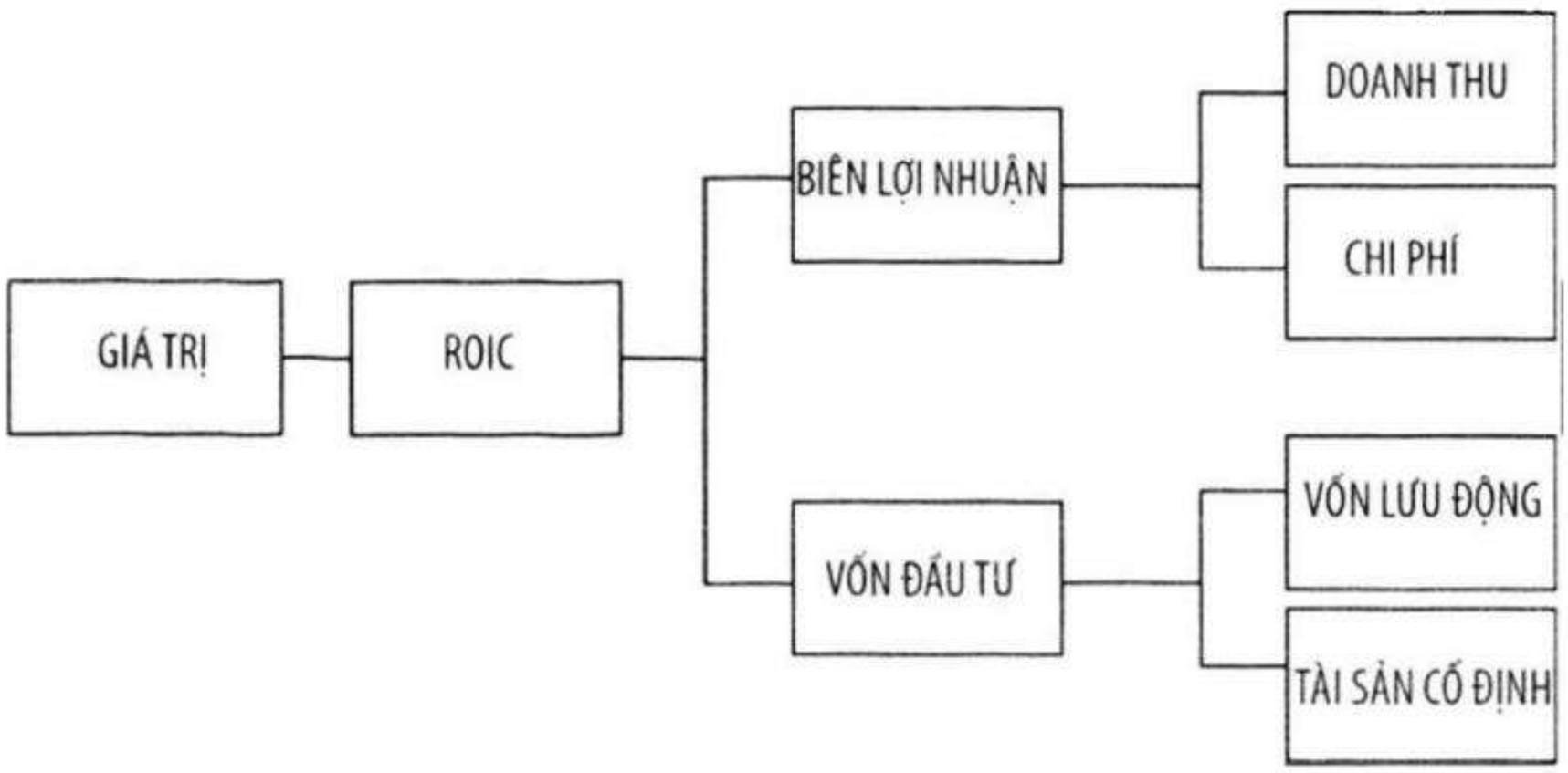
tạo doanh thu, chẳng hạn nhà xưởng và máy móc dùng để vận hành doanh nghiệp. Hình 3.1 minh họa các yếu tố thúc đẩy giá trị trong quá trình kiến tạo giá trị. Do đó, nhiệm vụ hàng đầu là phải tìm hiểu xem việc đầu tư vào công nghệ phân tích dữ liệu lớn sẽ tác động đến yếu tố thúc đẩy giá trị nào, và liệu phần doanh thu bổ sung đó có vượt quá chi phí đầu tư vào dữ liệu lớn hay không.

Để tìm hiểu giá trị được tạo ra như thế nào, chúng ta phải nắm rõ một hoạt động như đầu tư vào dữ liệu lớn tác động đến yếu tố thúc đẩy giá trị nào. Dữ liệu lớn có giúp mang lại sản phẩm và dịch vụ *nhanh hơn, tốt hơn, rẻ hơn* không? Nó có làm gia tăng doanh thu mà không gia tăng chi phí hay giữ nguyên doanh thu nhưng giảm bớt chi phí không? Hay dữ liệu lớn có làm giảm vốn lưu động hay tài sản cố định không? Dĩ nhiên, có thể có sự kết hợp giữa các yếu tố thúc đẩy doanh thu cơ bản, nhưng điều cốt lõi là phải hiểu sự đầu tư vào các hoạt động dữ liệu lớn sẽ tác động đến yếu tố thúc đẩy giá trị nào.

Cũng cần nhận thức rằng các yếu tố thúc đẩy giá trị có nhiều đặc điểm khác nhau, và chúng không có cùng mức độ như nhau về khả năng dự đoán được và độ chính xác trong những giả định làm nền tảng cho các ước tính về việc kiến tạo giá trị. Bạn nên lưu ý đến điều này khi đánh giá một khoản đầu tư vào dữ liệu lớn

(hoặc bất kỳ khoản đầu tư nào khác). Bảng 3.1 minh họa mức độ không chắc chắn đối với các yếu tố thúc đẩy giá trị. Ở phần cuối chương, chúng tôi sẽ quay lại bảng này để bàn về khả năng của các hoạt động đầu tư vào dữ liệu lớn trong việc thay đổi hoặc giảm bớt mức độ không chắc chắn trong quyết định đầu tư.

HÌNH 3.1 Các yếu tố thúc đẩy trong quá trình kiến tạo giá trị



BẢNG 3.1 Mức độ không chắc chắn khi thu hút giá trị

Yếu tố thúc đẩy giá trị	*Hệ thống phân bậc trong kiến tạo giá trị*	*Mức độ không chắc chắn khi thu hút giá trị*
Doanh thu	Khó dự đoán và khó định lượng	Đây là yếu tố thúc đẩy giá trị có mức độ không chắc chắn cao nhất, bởi nó đưa ra những giả định dựa trên sự tăng trưởng về số lượng khách hàng hoặc sự gia tăng về giá thành ở một khối lượng doanh số cho trước.
Chi phí	Có thể dự đoán và có thể định lượng	Nhìn chung chi phí mang tính nội bộ và có mức độ chắc chắn cao hơn so với những giả định dựa trên doanh thu gia tăng.
Vốn lưu động	Có thể dự đoán và có thể định lượng	Vốn lưu động nhìn chung nằm dưới sự kiểm soát của doanh nghiệp và do đó, có thể đưa ra những dự đoán về nó với mức độ chắc chắn cao hơn so với doanh thu.
Tài sản cố định	Có thể dự đoán nhưng có thang thời gian dài hơn	Nhìn chung, doanh nghiệp nắm rõ về giá trị của các tài sản cố định mà họ sở hữu. Nếu giá trị được tạo ra bằng cách giảm bớt (loại bỏ) tài sản cố định, thì sẽ phát sinh sự không chắc chắn về giá trị của tài sản khi bán lại và thời gian cần bỏ ra để thực hiện việc loại bỏ tài sản đó.

Để giúp bạn đọc hiểu rõ hơn về các yếu tố thúc đẩy giá trị, chúng tôi đã tạo ra một mẫu bảng báo cáo

thu nhập và báo cáo tài chính dùng cho một doanh nghiệp giả định. Chúng tôi sẽ dựa vào hai bảng này để đưa ra các ví dụ về tiềm năng kiến tạo giá trị khi áp dụng dữ liệu lớn theo nhiều cách khác nhau trong doanh nghiệp giả định trên.

Mẫu bảng báo cáo thu nhập và báo cáo tài chính

Bảng báo cáo thu nhập là xuất phát điểm để tìm hiểu về việc giá trị được tạo ra ở đâu và như thế nào. Chúng tôi sẽ sử dụng mẫu biểu này để minh họa cách dữ liệu lớn thúc đẩy quá trình kiến tạo giá trị.

Trong các ví dụ dưới đây, chúng tôi sẽ lập bảng báo cáo thu nhập và báo cáo tài chính cho một doanh nghiệp quy mô vừa với doanh số bán hàng 100 triệu đô-la và thu nhập ròng 10 triệu đô-la. Chi phí bỏ ra để mua các loại hàng hóa phục vụ cho việc sản xuất các sản phẩm của công ty này là 40 triệu đô-la, như vậy biên lợi nhuận gộp sẽ là 60 triệu đô-la, hay 60% trong ví dụ của chúng tôi. Chúng tôi lựa chọn doanh nghiệp có quy mô nhỏ hơn với mô hình tài chính tương đối đơn giản để phục vụ mục đích minh họa thay vì lập mô hình cho một số công ty sản xuất ra lượng dữ liệu lớn khổng lồ như Amazon và Facebook, vốn là những doanh nghiệp lớn có độ phức tạp cao. Chúng tôi tin rằng tiềm năng đầu tư tạo giá trị của các doanh

nghiệp nhỏ cũng hấp dẫn như các doanh nghiệp lớn hơn khi các doanh nghiệp cung cấp dịch vụ phân tích dữ liệu phát triển và tăng trưởng. Quy mô đầu tư vào công nghệ phân tích dữ liệu lớn tương đối đa dạng nhưng nhìn chung đều ở mức cao.

Chẳng hạn, chi phí đầu tư vào phân tích dữ liệu lớn của một công ty viễn thông lên đến hàng triệu đô-la mỗi năm – đây là những con số có thực, cho thấy mức độ đầu tư cần thiết vào một dự án dữ liệu lớn.

Đầu tư vốn	15 triệu đô-la
Đầu tư vào Hadoop	2 triệu đô-la
Chi phí vận hành IT/kho dữ liệu	3 triệu đô-la
Chi phí duy trì 30 chuyên gia phân tích dữ liệu và đội ngũ nhân viên phục vụ	3 triệu đô-la

Đây là những khoản đầu tư đáng kể ngay cả đối với một công ty viễn thông lớn. Sở dĩ mức độ đầu tư này được phê duyệt không chỉ vì đó là một công ty lớn, mà còn bởi mức độ dữ liệu được tạo ra hằng ngày thông qua mạng lưới và cơ sở khách hàng rộng lớn của công ty. Một số lĩnh vực ngành nghề khác cũng có những đặc điểm tương tự, chẳng hạn ngân hàng, bảo hiểm, và các công ty logistics với những đội xe lớn chuyên chở nhiều gói hàng tới nơi cần giao – thực ra bất kỳ tổ chức nào có số lượng khách hàng và số lượng đơn vị sản phẩm bán ra ở mức cao đều được. Một loại

hình tổ chức đặc biệt phù hợp là các doanh nghiệp có sự hiện diện đáng kể trên internet. Các công ty như Google và Facebook tạo ra một lượng lớn dữ liệu số để có thể thực hiện các hoạt động phân tích trên đó; nhưng ngoài ra, các tổ chức nhỏ hơn, chẳng hạn các công ty thương mại điện tử, cũng tạo ra được một lượng dữ liệu vừa đủ để nhận được những lợi ích từ hoạt động phân tích dữ liệu lớn.

Trong ví dụ của mình, chúng tôi giả định rằng khoản đầu tư rơi vào phần mua dịch vụ dữ liệu lớn, chi phí đầu tư cho tài sản cố định tương đối hạn chế hơn so với ví dụ về công ty viễn thông ở trên. Việc phát triển và thiết lập hệ thống phân tích thường do một công ty tư vấn dữ liệu lớn cung cấp, do đó phần chi phí đầu tư ở đây sẽ được chi trả trong năm thứ nhất, bên cạnh đó là các khoản chi phí hằng năm để thuê các dịch vụ hỗ trợ và phân tích của nhà cung cấp dịch vụ phân tích.

Bảng báo cáo thu nhập (Bảng 3.2) giúp chúng ta xác định được các yếu tố tác động đến doanh thu và chi phí trong quá trình kiến tạo giá trị. Nhưng chúng ta cũng cần cân nhắc đến số vốn cần thiết để vận hành doanh nghiệp, đến việc dữ liệu lớn có thể thúc đẩy quá trình kiến tạo giá trị ra sao bằng cách tác động đến số vốn sử dụng trong doanh nghiệp, chẳng hạn vốn lưu động và các tài sản cố định. Để tính toán

giá trị được tạo ra, chúng tôi sử dụng Lợi nhuận Hoạt động ròng Sau thuế (Net Operating Profit After Tax – NOPAT) bởi đây là khoản thu nhập trước khi tính đến phần thu nhập cho các nhà cung cấp vốn, tức là trước khi thanh toán lãi suất vay cho ngân hàng và trả cổ tức cho cổ đông đầu tư vào công ty.

Ví dụ về bảng báo cáo tài chính (Bảng 3.3) được trình bày tương đối đơn giản và nó cho thấy những hợp phần chính tạo nên tài sản có và tài sản nợ. Vốn lưu động là tài sản ngắn hạn trừ đi nợ ngắn hạn, tức là các loại tài sản có và tài sản nợ có thể được nhanh chóng chuyển thành tiền mặt. Trong ví dụ của chúng tôi, vốn lưu động là:

- Tiền mặt + Khoản phải thu + Hàng tồn kho – Khoản phải trả (5+25+20) – (25) = 25 triệu đô-la.
- Số vốn cần thiết để hỗ trợ cho doanh nghiệp là tiền vay ngân hàng và tài sản sở hữu (50+75) xấp xỉ 125 triệu đô-la.

BẢNG 3.2 Mẫu bảng báo cáo thu nhập
(giá trị tính theo đơn vị triệu đô-la)

Doanh số bán hàng	100	
Giá vốn hàng bán (COGS), nhân lực và nguyên vật liệu	(40)	
Biên lợi nhuận gộp (100-40)	60	60%
Chi phí vận hành:	(35)	
– Chi phí vận hành chung và chi phí quản lý hành chính (5)		
– Chi phí kinh doanh (15)		
– Dịch vụ khách hàng (5)		
– Nghiên cứu và phát triển (10)		
Khấu hao	(10)	
Thu nhập kinh doanh	15	15%
– Thuế	(3)	
Lợi nhuận Hoạt động ròng Sau thuế (NOPAT)	12	12%
– Chi phí lãi suất	(2)	
Thu nhập ròng	10	10%
– Lợi nhuận giữ lại (6)	(6)	
– Cổ tức (4)	(4)	

BẢNG 3.3 Mẫu bảng báo cáo tài chính

Tài sản			
Tiền mặt	5	Vốn lưu động	
Khoản phải thu	25	Vốn lưu động	Hàng hóa đã gửi hóa đơn nhưng chưa nhận được thanh toán
Hàng tồn kho	20	Vốn lưu động	
Tài sản ròng, nhà xưởng và trang thiết bị	100	Tài sản cố định	
Tổng tài sản	150		
Nợ			
Khoản phải trả	25	Vốn lưu động	Hàng hóa vật tư đã nhận nhưng chưa thanh toán
Vay ngân hàng	50	Vốn	
Cổ phần	75	Vốn	Vốn cổ phần 50 đô-la và lợi nhuận giữ lại tích lũy 25 đô-la
Tổng nợ	150		

Trong ví dụ mẫu, chúng tôi đã ước tính chi phí vốn[1] là 10%, bao gồm cả chi phí nợ và vốn cổ phần.

Giá trị tạo ra được tính theo công thức lấy hệ số lợi nhuận trên vốn đầu tư (ROIC) từ lợi nhuận hoạt

[1] Chi phí vốn: Chỉ các khoản chi phí cố định, bỏ ra một lần để mua nhà xưởng, trang thiết bị sử dụng trong quá trình sản xuất hàng hóa hoặc cung cấp dịch vụ. Nói cách khác, đây là tổng chi phí cần thiết để đưa một dự án vào trạng thái có thể vận hành về mặt thương mại. (ND)

động ròng sau thuế (NOPAT) chia cho vốn đầu tư (invested capital – IC).

ROIC = NOPAT / IC

Vốn đầu tư (IC) là 125 triệu đô-la lấy từ vốn cổ phần, lợi nhuận giữ lại, và vay ngân hàng.

NOPAT là 12 triệu đô-la.

Trong ví dụ của chúng tôi, ROIC hay giá trị được tạo ra là $12/$125 = 9,6%, nói cách khác, ROIC thấp hơn chi phí vốn 10%. Với điều kiện hiện tại, doanh nghiệp này tạo ra được lợi nhuận nhưng không tạo ra giá trị. Nghĩa là bất kỳ khoản đầu tư nào vào các hoạt động dữ liệu lớn đều phải mang về lợi nhuận hơn 10% mới được coi là tạo ra giá trị và cải thiện tỉ suất 9,6% ROIC hiện tại.

Từ góc nhìn đầu tư, thường có một độ trễ thời gian giữa các khoản đầu tư vốn và những thay đổi sau đó về doanh thu và/hoặc chi phí. Điều này nghĩa là chúng ta phải tính đến giá trị thời gian của tiền bạc. Tiền bạc ngày hôm nay có giá trị cao hơn tiền bạc trong tương lai. Do đó, chúng tôi muốn giới thiệu giá trị hiện tại ròng (net present value – NPV) làm phương pháp phân tích dòng tiền di chuyển khỏi doanh nghiệp (dưới hình thức hoạt động đầu tư và chi phí) và dòng tiền chảy vào doanh nghiệp (doanh thu hoặc những sự cắt giảm trong vốn đầu tư), và sử dụng công cụ này để đánh giá giá trị của các quyết định đầu tư. Giá trị hiện tại ròng dương cho thấy doanh thu dự kiến đến từ một dự án hoặc một khoản

đầu tư (xét theo giá trị tiền tệ hiện tại) cao hơn chi phí dự kiến (cũng theo giá trị tiền tệ hiện tại). Nhìn chung, một khoản đầu tư có giá trị NPV dương sẽ có lãi, còn giá trị NPV âm sẽ là một khoản lỗ ròng. Ngoài việc tính toán chi phí và doanh thu của khoản đầu tư vào dữ liệu lớn theo giá trị tiền mặt, bạn cũng cần tính đến tỉ lệ khấu hao khi tính giá trị hiện tại ròng. Ở ví dụ vừa nêu, chúng tôi ước tính chi phí vốn là 10%; tuy nhiên, đối với tất cả những dự án, trong đó có các dự án về dữ liệu lớn, bao giờ cũng tồn tại rủi ro không đạt được kết quả như dự tính ban đầu. Vì vậy, chúng ta cần phải bổ sung một yếu tố để bù đắp cho các rủi ro đó. Vẫn ở ví dụ trên, chúng tôi sẽ đưa vào một tỉ lệ khấu hao là 15%, nghĩa là bên cạnh các khoản chi phí vốn, chúng tôi còn ước lượng phần bù rủi ro là 5%.

Trong các ví dụ đã nêu, chúng tôi lấy một khoản đầu tư vào các dịch vụ dữ liệu lớn trong năm thứ nhất là 0,4 triệu đô-la, và chi phí vận hành hằng năm là 0,1 triệu đô-la. Vì dữ liệu lớn là một công nghệ có sức phát triển nhanh, nên chúng tôi giả định rằng tuổi thọ của dự án đầu tư đó là 5 năm; sau 5 năm, dự án này không có giá trị còn lại và cũng không góp thêm tiền mặt cho doanh nghiệp đầu tư.

Tuy nhiên, không phải tất cả các khía cạnh trong một doanh nghiệp đều có thể đo lường trực tiếp từ góc độ tiền tệ. Vì thế, chúng ta cần phải cân nhắc đến các phương tiện khác khi đo lường quá trình kiến tạo giá trị.

GIÁ TRỊ THỊ TRƯỜNG VÀ CÁC GIÁ TRỊ PHI TÀI CHÍNH

Giá trị là một trong những thuật ngữ mà chúng ta nghe nhắc đến thường xuyên, thậm chí bản thân chúng ta cũng sử dụng rất nhiều. Từ *giá trị* có thể có nhiều ý nghĩa khác nhau, và hãy bắt đầu bằng cách đưa ra định nghĩa riêng về giá trị trong bối cảnh này.

Trong khía cạnh dữ liệu và thông tin, từ *giá trị* có thể được định nghĩa theo nhiều cách khác nhau.

- Giá trị đối với xã hội: Một xã hội mở và sự chia sẻ thông tin, chẳng hạn thông tin được sử dụng như thế nào để hỗ trợ cho Mùa xuân Ả-rập, một chuỗi các cuộc biểu tình, nổi dậy, và phản kháng chống lại chính phủ ở khắp khu vực Trung Đông vào đầu năm 2011.
- Giá trị đối với chính phủ: Các chính phủ sử dụng ngày càng nhiều dữ liệu và các công nghệ phân tích để cải thiện hiệu suất, cung cấp thông tin, và giao tiếp với công dân. Cục thuế nước Anh là cơ quan rất tích cực sử dụng dữ liệu lớn, bao gồm mạng xã hội, để điều tra về các trường hợp gian lận thuế có thể xảy ra.
- Giá trị đối với bạn và những người thân thiết xung quanh bạn: Các ứng dụng như Facebook, WhatsApp, và các mạng xã hội khác mang

đến dòng tương tác liên tục theo thời gian thực giữa bạn với gia đình và bạn bè.

- Giá trị tài chính đối với các tổ chức thương mại.

Cuốn sách này nhằm mục đích tìm hiểu về cách giá trị được tạo ra trong các tổ chức thương mại, và chúng tôi bắt đầu bằng việc tìm hiểu giá trị ở cấp độ cao nhất. Trong trường hợp tổ chức thương mại, đây chính là giá trị doanh nghiệp[1] của tổ chức. Theo các phương pháp định giá doanh nghiệp truyền thống, người ta nhân giá trị của một cổ phiếu với số lượng cổ phiếu cộng với giá trị của các khoản nợ – đây được gọi là giá trị doanh nghiệp. Với các công ty được niêm yết công khai, cổ phiếu được giao dịch trên sàn chứng khoán, và rất dễ tính ra giá trị thị trường của các công ty này. Với các công ty tư nhân, giá trị cổ phiếu khó tính toán hơn vì không có thị trường đại chúng cho cổ phiếu đó. Một cách khác để định giá doanh nghiệp là sử dụng giá trị ghi sổ của doanh nghiệp bởi con số này thể hiện các khoản đầu tư vốn từng được thực hiện theo thời gian. Khi so sánh số liệu tính toán giá trị ghi sổ và giá trị thị trường theo thời gian, chúng ta sẽ thấy nổi lên một câu hỏi thú vị: Nên định giá doanh

[1]Giá trị doanh nghiệp (enterprise value): Là một thước đo về tổng giá trị của một công ty, không những bao gồm giá trị vốn hóa thị trường của công ty mà tính đến cả các khoản nợ ngắn và dài hạn cũng như số tiền mặt trong bảng cân đối tài chính của công ty đó. (ND)

nghiệp như thế nào và yếu tố nào thúc đẩy sự định giá này? Các nghiên cứu của hãng tư vấn Accenture cho thấy giá trị ghi sổ đã giảm từ mức bằng 80% giá trị thị trường xuống còn 25% trong một giai đoạn kéo dài 20 năm. Ý nghĩa rút ra là các khoản đầu tư vào tài sản hữu hình được ghi lại trong bảng cân đối kế toán của doanh nghiệp ngày nay ít có tác động đến các yếu tố thúc đẩy giá trị hơn so với trong quá khứ. Một trong những cách giải thích khả thi ở đây là: các yếu tố phi tài chính ngày nay có tác động lớn hơn đến giá trị thị trường, và các biện pháp kế toán truyền thống không thể nắm bắt được thông tin này trong các yếu tố thúc đẩy giá trị. Các yếu tố phi tài chính thường là về Kinh tế, Xã hội, và Quản trị (Economic, Social, Governance – ESG). Những nghiên cứu về các yếu tố ESG cho thấy có sự liên hệ giữa chúng với giá trị doanh nghiệp của một công ty. Dữ liệu lớn là dữ liệu lớn có thể thúc đẩy quá trình kiến tạo giá trị trong các yếu tố ESG.

Các giá trị phi tài chính cốt lõi trong khung ESG là mối quan hệ với khách hàng, xã hội, sự đổi mới sáng tạo, môi trường, nguồn vốn con người, và vấn đề quản trị doanh nghiệp. Nếu chia nhỏ các yếu tố này hơn nữa, chúng ta sẽ xác định được đâu là điểm mà các phân tích dữ liệu lớn có thể chuyển thành những yếu tố thúc đẩy giá trị phi tài chính, tức là tạo ra giá trị (xem Bảng 3.4).

BẢNG 3.4 Các yếu tố phi tài chính thúc đẩy quá trình kiến tạo giá trị

Các yếu tố phi tài chính cốt lõi	*Quan hệ khách hàng*	*Xã hội*	*Đổi mới sáng tạo*	*Môi trường*	*Nguồn vốn con người*	*Quản trị doanh nghiệp*
Các chỉ số chính	Sự hài lòng của khách hàng	Nhận thức của công chúng; Quản lý chuỗi cung ứng	Phát triển sản phẩm và dịch vụ	Tác động đến môi trường	Sự gắn kết của nhân viên	Sự liêm chính về đạo đức
Các yếu tố ESG	Giữ chân khách hàng; Sự trung thành của khách hàng; Sự tín nhiệm; Danh tiếng; Giá cả, chất lượng của sản phẩm/ dịch vụ	Đầu tư vào cộng đồng; Đối thoại với các bên liên quan; Phủ sóng truyền thông	Các sản phẩm và dịch vụ mới; Chi phí cho hoạt động R&D; Giá trị của các bằng sáng chế; Số lượng sản phẩm mới	Hiệu suất sử dụng năng lượng; Giảm bớt rác thải; Tái chế; Giảm ô nhiễm từ hoạt động vận tải	Sức khỏe và an toàn; Tuyển dụng và giữ chân nhân viên; Đào tạo; Quản lý hiệu suất làm việc; Bình đẳng và đa dạng	Báo cáo và sự minh bạch; Nguyên tắc hành xử có đạo đức; Thành phần của hội đồng quản trị; Bình đẳng và đa dạng

Mô hình ESG cung cấp một khung giúp chúng ta tìm hiểu xem dữ liệu lớn có thể tạo ra giá trị như thế nào, và sau đó liên kết thông tin này với bảng báo cáo thu nhập và bảng cân đối kế toán của doanh nghiệp. Có thể dễ dàng ước tính các khoản đầu tư và chi phí liên quan đến phân tích và các hoạt động về dữ liệu lớn. Việc ước tính doanh thu tiềm năng sẽ khó hơn một chút. Do đó, chúng ta cần phải xác định những gì đại diện cho các yếu tố thúc đẩy giá trị và đưa ra những nhận định về giá trị tài chính của những đại diện này. Chúng tôi đã tính đến một số mối quan hệ chịu ảnh hưởng của các hoạt động đầu tư vào dữ liệu lớn, bên cạnh đó chúng tôi cũng thừa nhận rằng còn có các mối quan hệ quan trọng khác nữa.

Dữ liệu lớn là một hệ quả tự nhiên của quá trình số hóa mảng bán hàng và các dòng chảy sản phẩm trong chuỗi cung ứng. Khi internet trở thành một cuộc chuyển mình lớn trong cách thức giao tiếp của con người và doanh nghiệp, các số liệu dự đoán về sự phát triển trong lĩnh vực thương mại điện tử cũng ở cấp số nhân. Những dự đoán sớm này hầu hết đều đúng, ngoại trừ thang thời gian dài hơn so với những dự đoán ban đầu. Thông thường, đó là sự hội tụ của ba yếu tố: điều mà công nghệ có thể mang lại, những khoản đầu tư vào cơ sở hạ tầng, và sự chấp nhận cũng như sự thay đổi về hành vi của người tiêu dùng.

Ngày nay, thương mại điện tử và bán lẻ trực tuyến là một lực lượng lớn đằng sau sự phát triển và nhu cầu đối với dữ liệu lớn. Anh hiện đang dẫn đầu với 17% doanh số bán lẻ trực tuyến, tiếp đến là Mỹ với 13%, Đức 12%, Pháp và Thụy Điển 9%. Theo ước tính, các nhà bán lẻ ở Anh sẽ gửi đi 1,2 tỉ kiện hàng, được giao từ các nhà kho chỉ rộng 40m². Hoạt động giao hàng bằng xe tải tăng 4% trong giai đoạn từ năm 2015 đến 2016 trong khi lưu lượng giao thông nói chung chỉ tăng 1%. Sự dịch chuyển về cách thức mua sắm đã thúc đẩy sự phát triển của công nghệ phân tích dữ liệu lớn bởi vì ngay từ ban đầu, phần lớn thông tin đều đã ở dạng số. Một nhà bán lẻ như Tesco có 3.500 cửa hàng ở Anh, với 40.000 sản phẩm tạo ra 100.000 điểm dữ liệu để giám sát và theo dõi. Một hệ quả tiếp theo của sự phát triển trong lĩnh vực thương mại điện tử là việc phân phối hàng hóa với đội quân gồm hàng nghìn xe tải chạy ngang dọc trên khắp mọi nẻo đường của đất nước. Theo ước tính của ngành giao thông vận tải, hiện đang thiếu khoảng 45.000 tài xế xe tải. Có lẽ giải pháp sẽ là loại xe tải không người lái, và như vậy, nhu cầu về năng lực dữ liệu càng lớn hơn nữa để có thể giám sát loại xe này khi chúng di chuyển khắp nơi. Các đơn vị chuyển phát hàng như FedEX và DHL vừa là công ty vận tải lại vừa là trung tâm dữ liệu.

CÁC KHOẢN ĐẦU TƯ VÀO NƠI DỮ LIỆU LỚN CÓ THỂ TẠO RA GIÁ TRỊ

Đối với các tổ chức thương mại, quyết định tạo ra nhiều giá trị nhất mà một công ty có thể thực hiện là tăng giá. Theo một nghiên cứu do hãng tư vấn quản lý McKinsey thực hiện đối với 2.348 công ty, khi mọi điều kiện khác đều giống nhau, chỉ cần doanh nghiệp gia tăng 1% các hoạt động sau đây, lợi nhuận họ thu về cũng sẽ gia tăng theo:

- Tăng giá 1%: tăng lợi nhuận 11%
- Giảm chi phí biến đổi 1%: tăng lợi nhuận 7,8%
- Tăng số lượng khách hàng 1%: tăng lợi nhuận 3,3%
- Cắt giảm chi phí cố định 1%: tăng lợi nhuận 2,3%

Nhiều doanh nghiệp thường không khỏi ngạc nhiên khi phát hiện ra rằng giá trị cao nhất được tạo ra khi khách hàng sẵn sàng chi trả nhiều hơn cho sản phẩm/dịch vụ hơn là khi có thêm nhiều khách hàng cùng chi trả một mức giá. Giảm chi phí biến đổi, tức chi phí sản xuất, cũng là một phương án tương đối hấp dẫn, còn cắt giảm chi phí cố định là hành động ít mang lại lợi nhuận nhất mà một doanh nghiệp có thể thực hiện.

Các khoản đầu tư vào dữ liệu lớn thường tập trung vào khách hàng – hoặc gia tăng số lượng khách hàng và doanh thu bán hàng, hoặc gia tăng doanh thu trên mỗi khách hàng. Kết quả là, doanh nghiệp sẽ có thể tiến hành phân khúc cơ sở khách hàng tốt hơn. Các công ty viễn thông có số lượng khách hàng lớn và có sự tương tác khá thường xuyên với khách hàng vì họ thường vào mạng internet trên điện thoại thông minh, đồng thời, nhà cung cấp dịch vụ viễn thông cũng nắm rõ thông tin về vị trí của khách hàng. Do đó, tiềm năng tạo giá trị ở đây là rất lớn. Theo cách làm phổ biến hiện nay trong ngành này, họ sử dụng khoảng 6-7 phân khúc khách hàng khác nhau, chẳng hạn phân khúc người dùng dữ liệu tích cực, người thường xuyên đi lại giữa các quốc gia, người có xu hướng sớm tiếp nhận công nghệ mới... Chỉ số tài chính chủ đạo trong lĩnh vực viễn thông là Chi phí thu nạp khách hàng (Subscriber Acquisition Costs – SAC), tỉ lệ rời bỏ (tỉ lệ phần trăm số khách hàng không gia hạn hợp đồng) và Doanh thu trung bình trên người dùng (Average Revenue Per User – ARPU). Những yếu tố thúc đẩy giá trị mà ai cũng có thể nhìn thấy là giảm SAC, giảm tỉ lệ rời bỏ, và gia tăng ARPU. Vì thế, ngành viễn thông đang đầu tư mạnh tay vào các công nghệ phân tích dữ liệu lớn nhằm hiểu rõ hơn về cơ sở khách hàng của mình, đưa ra những quyết sách mang lại lợi nhuận liên quan đến việc xác định khách hàng nào nên bỏ

tiền thực hiện các chiến dịch marketing, nhóm khách hàng nào tốn nhiều chi phí phục vụ, và làm thế nào để khuyến khích khách hàng sử dụng nhiều hơn nữa.

Bản chất của việc đầu tư vào dữ liệu lớn để giải quyết vấn đề khách hàng và doanh thu thường mang tính bấp bênh hơn nhiều so với việc tiết kiệm chi phí. Kết quả đầu tư phụ thuộc hoàn toàn vào độ chính xác của các giả định đằng sau quyết định được đưa ra. Một thách thức khác cần cân nhắc là sự sẵn có của các dữ liệu chính xác dùng để thiết lập mốc tiêu chuẩn và so sánh mốc này với các phương án đầu tư thay thế khác – đây gọi là thử nghiệm A/B. Nhìn chung, nếu doanh nghiệp không nắm vững dữ liệu khách hàng, quyết định đầu tư chỉ giống như bắn tên trong bóng tối. Khi đó, vấn đề này lại trở thành tình huống cổ điển về việc con gà có trước hay quả trứng có trước, nghĩa là việc đầu tư vào dữ liệu lớn để cải thiện độ chính xác của dữ liệu trở thành yêu cầu tiên quyết để thu thập được các thông tin và dữ liệu cần thiết nhằm hỗ trợ cho quyết định đầu tư vào dữ liệu lớn.

Có một số yếu tố đại diện có thể được áp dụng để nắm chắc hơn về các giả định nhằm xác định tiềm năng kiến tạo giá trị. Trong phần tiếp theo, chúng tôi sẽ cung cấp các ví dụ cho thấy việc đầu tư vào dữ liệu lớn có thể tạo ra giá trị như thế nào. Chúng tôi sẽ bắt đầu với việc xác định các phân khúc khách hàng sẵn

sàng chi trả mức giá cao hơn và gia tăng mức độ chi tiêu trên khách hàng, và chúng tôi sẽ sử dụng ví dụ về ngành viễn thông để minh họa cách thiết lập một số hoạt động kiến tạo giá trị liên quan đến khách hàng.

Kiến tạo giá trị từ nguồn doanh thu bổ sung từ khách hàng hiện tại – giá cao hơn

Đây là một khía cạnh quan trọng trong hoạt động tạo giá trị. Trong nhiều doanh nghiệp, chi phí thu nạp khách hàng tương đối cao, do đó phần doanh thu bổ sung từ cơ sở khách hàng hiện tại có thể mang lại một khoản lợi nhuận rất lớn. Điều này đặc biệt đúng khi biên lợi nhuận gộp cao (tức chi phí sản xuất sản phẩm hoặc dịch vụ ở mức thấp). Trong lĩnh vực viễn thông, chi phí phục vụ thêm một khách hàng hầu như không đáng kể, bởi mạng lưới viễn thông đã được xây dựng, và trong chừng mực nhà mạng không cần phải bổ sung công suất do số lượng khách hàng gia tăng, biên lợi nhuận gộp gần đạt mức 100%. Khó khăn khi sử dụng dữ liệu lớn là làm sao để xác định được các phân khúc khách hàng ít nhạy cảm về giá hơn, chẳng hạn những người đón đầu trào lưu. Bằng cách bán chéo hoặc bán thêm các sản phẩm/dịch vụ cao cấp hơn cho các khách hàng hiện tại, chỉ số doanh thu trên khách hàng có thể sẽ còn được đẩy cao hơn nữa. Những thị trường có biên lợi nhuận gộp

cao thường là sân chơi dành cho các doanh nghiệp số như Facebook và Google, mô hình kinh tế của họ vận hành theo hướng lợi nhuận tăng dần theo sự tăng trưởng và quy mô, và việc gia tăng thị phần thường mang lại lợi nhuận rất cao.

Quay trở lại với ví dụ mẫu của chúng ta, nguồn doanh thu bổ sung từ việc gia tăng 1% giá thành sản phẩm đối với 20% khách hàng hiện tại mang lại mức doanh thu trung bình là 102 đô-la. Điều này sẽ bổ sung thêm 2 triệu đô-la vào doanh thu và thêm 1,2 triệu đô-la tiền mặt ở mức 60% biên lợi nhuận gộp. Giả sử chi phí marketing và bán hàng tăng thêm 0,1 triệu đô-la. Chỉ số NPV trên khoản đầu tư này sẽ là 3,0 triệu đô-la.

Bảng 3.5 cho thấy sức mạnh của việc tăng giá. Tính đến thời điểm này, đây là yếu tố thúc đẩy mạnh mẽ nhất đối với việc kiến tạo giá trị. Tính giá cao hơn cho khách hàng hiện tại là một biện pháp có sức hấp dẫn lớn, và đối với các công ty viễn thông, nhiệm vụ trọng tâm của bộ phận marketing là gia tăng chỉ số ARPU (doanh thu trung bình trên khách hàng).

BẢNG 3.5 Tạo giá trị bổ sung từ khách hàng hiện tại

(Đơn vị: đô-la)

	Đầu tư	*Chi phí marketing bổ sung*	*Chi phí vận hành*	*Phần gia tăng của dòng tiền do doanh thu gia tăng*	*Dòng tiền thuần*	*NPV*
Năm 1	0,4 triệu	0,1 triệu	0,1 triệu	1,2 triệu	0,6 triệu	
Năm 2		0,1 triệu	0,1 triệu	1,2 triệu	1,0 triệu	
Năm 3		0,1 triệu	0,1 triệu	1,2 triệu	1,0 triệu	
Năm 4		0,1 triệu	0,1 triệu	1,2 triệu	1,0 triệu	
Năm 5		0,1 triệu	0,1 triệu	1,2 triệu	1,0 triệu	3,0 triệu

Bằng cách xác định các phân khúc khách hàng có xu hướng chi tiêu cao, chẳng hạn những người sớm tiếp nhận sản phẩm, các khoản đầu tư vào phân tích dữ liệu lớn có thể mang lại nhiều kết quả rất hấp dẫn.

Tạo giá trị từ sự hài lòng của khách hàng

Khách hàng có thể tác động đến quá trình kiến tạo giá trị, hoặc thông qua sự tăng trưởng doanh thu nhờ bán thêm được sản phẩm hay dịch vụ, hoặc thông qua cắt giảm chi phí. Chúng ta hãy bắt đầu với việc cắt giảm chi phí sản xuất, bán hàng, và bảo dưỡng sản phẩm. Trong hầu hết các doanh nghiệp, khách hàng đều có nhiều hơn một lần tương tác với sản phẩm và dịch vụ. Áp lực cạnh tranh gia tăng khiến việc tạo sự khác biệt cho sản phẩm trên thị trường ngày càng khó khăn hơn; vì thế, trong nhiều lĩnh vực ngành nghề, yếu tố dịch vụ trong hành trình khách hàng có

khả năng làm gia tăng cả chi phí lẫn cơ hội giữ chân khách hàng, và có thể củng cố mối quan hệ với khách hàng. Trong lĩnh vực viễn thông, việc phân tích hiệu suất vận hành của mạng lưới là điều kiện then chốt để doanh nghiệp có thể nâng cao chất lượng dịch vụ. Cần một lượng rất lớn dữ liệu và thông tin để quản lý mạng lưới và các hoạt động phân tích dữ liệu lớn được sử dụng nhằm kết hợp dữ liệu về hiệu suất theo thời gian thực với những dự đoán về các sự kiện trong tương lai. Chẳng hạn, một sự kiện bóng đá lớn với hàng nghìn người hâm mộ đổ ra đường sẽ đòi hỏi chất lượng mạng lưới tốt. Về mặt vận hành, nhà cung cấp dịch vụ viễn thông có thể xác định sự hiện diện của các khách hàng mang lại giá trị cao ở một trận đấu bóng đá và do đó cung cấp dịch vụ ưu tiên cho họ. Việc này cần phải được thực hiện tự động, và chỉ có thể thực hiện thông qua các thuật toán dự đoán kết hợp với khả năng xác định từng khách hàng.

Câu hỏi đặt ra là chi phí phục vụ có thể tác động ra sao đến chi phí vận hành trong bảng báo cáo thu nhập mẫu. Trong ví dụ ở Bảng 3.6, chúng tôi ước tính rằng khoản đầu tư vào dữ liệu lớn sẽ giúp bộ phận dịch vụ khách hàng phục vụ được thêm 10% khách hàng mỗi ngày, và rằng chi phí phục vụ khách hàng là 5 triệu đô-la mỗi năm, chủ yếu là phần chi phí cho nhân sự (xem bảng báo cáo thu nhập mẫu). Nếu nhờ cải thiện năng suất mà nhu cầu nhân sự giảm đi 10%, khi đó doanh nghiệp sẽ tiết kiệm được 0,5 triệu đô-la

chi phí nhân sự hằng năm. Tuy nhiên, việc tái cơ cấu bộ phận dịch vụ sẽ tốn 0,3 triệu đô-la trong Năm 1. Dường như trong ví dụ này, những cải thiện về hiệu suất lại khiến doanh nghiệp mất khoảng 0,3 triệu đô-la trong Năm 1 (tiết kiệm 0,5 triệu đô-la – đầu tư 0,4 triệu đô-la – tái cơ cấu 0,3 triệu đô-la – chi phí vận hành dữ liệu lớn 0,1 triệu đô-la), và sau đó tiết kiệm 0,4 triệu đô-la trong các năm tiếp theo bằng cách đầu tư vào dữ liệu lớn. Khi tính điểm hòa vốn, chúng ta thấy rằng khoản đầu tư này sẽ hoàn vốn sau 1,6 năm. Khi tính giá trị hiện tại ròng với tỉ lệ khấu hao 15%, chúng ta sẽ có chỉ số NPV là 0,73 triệu đô-la.

Ví dụ này là một khoản đầu tư hợp lý vì chi phí đầu tư và số tiền tiết kiệm được tương đối chính xác và có thể dự đoán. Doanh nghiệp nên bật đèn xanh để thực hiện dự án đầu tư này.

BẢNG 3.6 Tạo giá trị từ sự hài lòng của khách hàng

(Đơn vị: đô-la)

	Đầu tư	*Tái cơ cấu*	*Chi phí vận hành*	*Tiết kiệm*	*Dòng tiền thuần*	*NPV*
Năm 1	0,4 triệu	0,3 triệu	0,1 triệu	0,5 triệu	-0,3 triệu	
Năm 2			0,1 triệu	0,5 triệu	0,4 triệu	
Năm 3			0,1 triệu	0,5 triệu	0,4 triệu	
Năm 4			0,1 triệu	0,5 triệu	0,4 triệu	
Năm 5			0,1 triệu	0,5 triệu	0,4 triệu	0,73 triệu

Tạo giá trị từ việc giảm chi phí thu nạp khách hàng

Một cách để tính chi phí thu nạp khách hàng (SAC) là cộng tổng các loại chi phí liên quan đến bán hàng, chẳng hạn chi phí vận hành cửa hàng và các kênh phân phối khác, chi phí liên quan đến marketing và khuyến mãi, các chương trình giảm giá và các hoạt động khác nhằm giành được một khách hàng. Lấy số liệu này chia cho số lượng khách hàng, khi đó chúng ta sẽ có một số liệu đại diện cho chi phí thu nạp khách hàng. Điều mà nội bộ giới marketing đều biết rõ là một phần lớn chi phí marketing sẽ bị lãng phí, nhưng không ai biết hoạt động nào gây lãng phí. Vấn đề chúng ta đang cố gắng giải quyết là làm thế nào để nhắm mục tiêu khách hàng chính xác hơn và giảm bớt chi phí thu nạp khách hàng mới.

Sau đó, có thể đánh giá hiệu quả của những khoản đầu tư vào dữ liệu lớn thông qua phần chi phí SAC cắt giảm được hoặc tích của số lượng khách hàng tăng thêm và doanh thu trên mỗi khách hàng. Có thể giảm chỉ số SAC bằng cách nhắm mục tiêu khách hàng tốt hơn, hoặc thực hiện phân khúc khách hàng hợp lý hơn. Nếu vận dụng cách thức đầu tư như mô tả trong trường hợp ví dụ về việc phục vụ thêm 10% khách hàng, mọi chuyện có thể sẽ diễn ra như sau. Chi phí bán hàng của doanh nghiệp trong ví dụ mẫu của

chúng tôi là 5 triệu đô-la, trong đó chi phí marketing trực tiếp là 1 triệu đô-la/năm. Với 1 triệu khách hàng thực hiện giao dịch mua sắm, chi phí marketing là 1 đô-la/khách hàng. Việc xây dựng một kho kiến thức vừa đủ về phân khúc khách hàng gần như bất khả thi nếu chỉ dựa vào các phương pháp phân khúc truyền thống dựa trên tiêu chí thu nhập và thông tin nhân khẩu học. Sử dụng các tập dữ liệu khổng lồ từ hàng loạt nguồn khác nhau, chẳng hạn thói quen mua sắm, các hoạt động trên mạng xã hội..., dữ liệu lớn sẽ cung cấp những phân tích có thể cải thiện độ chính xác của việc khoanh vùng marketing cho các nhóm mục tiêu khác nhau trong cơ sở khách hàng. Trong ví dụ của chúng tôi, chi phí marketing trên mỗi khách hàng là 1 đô-la/năm trên cơ sở doanh thu trung bình là 100 đô-la/khách hàng. Sử dụng chiến lược marketing đẩy[1] qua email, khoản chi phí 1% cho marketing dường như chưa đủ cao để bảo đảm thành công cho những khoản đầu tư lớn. Tuy nhiên, do tỉ lệ khách hàng rời bỏ trong ví dụ của chúng tôi là gần 50%, nên công ty này mỗi năm phải thay thế một nửa cơ sở khách hàng của mình, nghĩa là mỗi năm họ phải tìm kiếm thêm

[1]Marketing đẩy (push marketing): Một chiến lược quảng bá trong đó doanh nghiệp cố gắng giới thiệu về sản phẩm của mình trực tiếp với khách hàng, bất kể họ có mong muốn hay quan tâm tới sản phẩm hay không. (ND)

500.000 khách hàng mới. Câu hỏi đặt ra là, liệu một khoản đầu tư trị giá 0,4 triệu đô-la vào công nghệ phân tích dữ liệu lớn, cách phân khúc khách hàng hiệu quả hơn, và khả năng nhắm mục tiêu cho các thông điệp marketing tốt hơn có tạo ra giá trị không? Có thể tạo ra giá trị gì thông qua khả năng phân khúc khách hàng tốt hơn và nhắm mục tiêu đúng vào các phân khúc khách hàng có khuynh hướng mua hàng cao hơn? Phân khúc khách hàng tiềm năng nào có khả năng trở thành khách hàng trung thành cao hơn so với nhóm khách hàng chỉ tìm đến mua một lần?

Trong ví dụ ở Bảng 3.7, việc hiểu rõ cơ sở khách hàng hơn và khả năng thực hiện marketing trực tiếp đúng mục tiêu hơn có thể khiến tỉ lệ rời bỏ giảm xuống, ví dụ từ 50% xuống còn 30%, nghĩa là 200.000 khách hàng sẽ quay trở lại. Do đó, chi phí marketing trực tiếp có thể giảm 0,2 triệu đô-la (200.000 khách hàng x 1 đô-la chi phí marketing/khách hàng). Giá trị được tạo ra có thể là một khoản đầu tư ròng trị giá 0,3 triệu đô-la trong Năm 1 và một khoản tiết kiệm 0,1 triệu đô-la trong bốn năm tiếp theo. Giả sử khoản tiết kiệm này được duy trì trong suốt vòng đời của dự án đầu tư trên, thì giá trị hiện tại ròng sẽ được tính với NPV là 0,1 triệu đô-la. **Đây là khoản đầu tư không nên theo đuổi vì NPV gần như bằng 0.**

BẢNG 3.7 Tạo giá trị từ việc giảm chi phí thu nạp khách hàng

(Đơn vị: đô-la)

	Đầu tư	*Chi phí vận hành*	*Tiết kiệm nhờ giảm chi phí marketing*	*Dòng tiền thuần*	*NPV*
Năm 1	0,4 triệu	0,1 triệu	0,2 triệu	-0,3 triệu	
Năm 2		0,1 triệu	0,2 triệu	0,1 triệu	
Năm 3		0,1 triệu	0,2 triệu	0,1 triệu	
Năm 4		0,1 triệu	0,2 triệu	0,1 triệu	
Năm 5		0,1 triệu	0,2 triệu	0,1 triệu	0,1 triệu

Như Bảng 3.7 đã chỉ ra, không phải cứ đầu tư nhằm mục tiêu cắt giảm chi phí là sẽ đạt được thành quả. Tuy vậy, chi phí thu nạp khách hàng thường khá cao, nên hãy cùng tìm hiểu về việc tăng sự hiệu quả cho các khoản chi phí marketing và các hoạt động tiếp cận khách hàng mục tiêu.

Tạo ra giá trị nhờ thực hiện marketing hiệu quả hơn – thu nạp thêm nhiều khách hàng

Một cách khác để thực thi chiến lược giảm chi phí marketing thông qua việc giảm chi phí thu nạp khách hàng là duy trì mức chi phí marketing nhưng thu hút thêm khách hàng. Gia tăng cơ sở khách hàng là tham vọng của hầu hết mọi tổ chức, và nhìn chung, marketing là cách giúp họ đạt được tham vọng này.

Chúng tôi tiếp tục sử dụng ví dụ trên (xem Bảng 3.8), chi phí marketing vẫn là 1 triệu đô-la, và do khả năng nhắm mục tiêu cho thông điệp marketing được cải thiện, nên có thể gia tăng số lượng khách hàng lên khoảng 10.000 người, từ đó gia tăng doanh thu thêm 1 triệu đô-la với mức chi tiêu trung bình là 100 đô-la/khách hàng. Với biên lợi nhuận gộp 60% cho mức doanh thu trung bình từ một khách hàng là 100 đô-la, kết quả là dòng tiền mặt tăng thêm 0,6 triệu đô-la. Chỉ số NPV của khoản đầu tư trị giá 0,4 triệu đô-la trong năm năm là 1,33 triệu đô-la.

Bằng cách giữ nguyên chi phí marketing và gia tăng hiệu quả cũng như khả năng nhắm mục tiêu trong marketing, một khoản đầu tư vào phân tích dữ liệu lớn có thể mang lại NPV là 1,33 triệu đô-la. Như vậy, đây là một dự án đầu tư hợp lý. Nó cho thấy rằng doanh thu là yếu tố thúc đẩy giá trị mạnh hơn so với cắt giảm chi phí; tuy nhiên, cắt giảm chi phí lại có tính chắc chắn hơn so với việc dự đoán mức doanh thu cao hơn.

BẢNG 3.8 Tạo giá trị nhờ thực hiện marketing hiệu quả hơn

(Đơn vị: đô-la)

	Đầu tư	*Chi phí vận hành*	*Gia tăng dòng tiền nhờ gia tăng doanh thu*	*Dòng tiền thuần*	*NPV*
Năm 1	0,4 triệu	0,1 triệu	0,6 triệu	0,1 triệu	
Năm 2		0,1 triệu	0,6 triệu	0,5 triệu	
Năm 3		0,1 triệu	0,6 triệu	0,5 triệu	
Năm 4		0,1 triệu	0,6 triệu	0,5 triệu	
Năm 5		0,1 triệu	0,6 triệu	0,5 triệu	1,33 triệu

Tạo giá trị từ giá trị giữ chân khách hàng

Việc giữ chân khách hàng thường liên quan đến cách họ nhìn nhận về giá trị của sản phẩm và dịch vụ mà doanh nghiệp cung cấp. Khi sản phẩm hoặc dịch vụ có giá trị cao đối với khách hàng, và họ phải mất chi phí nếu chuyển đổi sang nhà cung cấp khác thì khả năng giữ chân khách hàng thường ở mức cao. Chìa khóa là tạo ra chi phí chuyển đổi sao cho khách hàng cảm thấy bất tiện hoặc "tốn kém" khi mua sản phẩm/dịch vụ từ đối thủ cạnh tranh của doanh nghiệp. Chẳng hạn, ngày nay, người nông dân là một doanh nhân công nghệ cao, với các loại máy kéo và máy gặt được kiểm soát bằng thiết bị định vị GPS, có thể tìm ra cung đường di chuyển hiệu quả nhất về mặt chi phí khi gặt. Theo cách nói trong dữ liệu lớn thì có thể ví đây là giá trị thông tin mang đến cho khách hàng,

chẳng hạn cung cấp cho người nông dân thông tin về giá cả thị trường các mặt hàng nông sản mà anh ta cung cấp, dự báo thời tiết, và xu hướng thị trường. Những thông tin này có thể mang lại giá trị cao cho người nông dân. Trong trường hợp mặt hàng khoai tây, việc sử dụng dữ liệu và kiến thức nông học có thể gia tăng sản lượng từ 30% lên 50%, tùy thuộc vào loại đất – đây là sự kiến tạo giá trị quan trọng đối với nông dân. Một khía cạnh khác trong đó sự tối ưu hóa mang lại lợi ích không chỉ cho người nông dân mà còn cho xã hội nói chung là việc tiết giảm sử dụng hóa chất và thuốc trừ sâu. Ở Mỹ có khoảng 2,2 triệu nông trại, và một nông trại trung bình dành ra 111.000 đô-la để mua các loại hóa chất; do đó bất kỳ sự cắt giảm nào trong nhóm chi phí này cũng vừa góp phần cải thiện khả năng sinh lãi cho người nông dân lại vừa có những sự đóng góp quan trọng cho môi trường.

Cơ hội máy tinh hóa hoạt động nông nghiệp và tối ưu hóa việc sử dụng hóa chất và thuốc trừ sâu đẩy hai nhóm nhà cung cấp khác nhau vào thế đối đầu trong một cuộc chiến tranh về dữ liệu. Chẳng hạn, một bên là công ty hóa chất nông nghiệp Monsanto, và một bên là nhà sản xuất trang thiết bị nông nghiệp John Deere. Monsanto nhận ra rằng dữ liệu có thể giúp gia tăng sản lượng hoa màu trên toàn thế giới thêm 30% thông qua việc cung cấp những lời khuyên về trồng trọt dựa trên dữ liệu thu thập được – và tạo ra

một sự tác động về mặt kinh tế trị giá lên tới khoảng 20 tỉ đô-la. Tuy nhiên, việc thu thập dữ liệu và sau đó là trồng trọt và gặt hái cần đến những công cụ nông nghiệp tinh vi và có kết nối với dữ liệu. Người nông dân cần đến những thiết bị cảm ứng, chẳng hạn để đo lường độ mặn và các đặc điểm khác của đất, và những ứng dụng như Hệ thống Định vị Toàn cầu (Global Positioning System – GPS) và Hệ thống Thông tin Địa lý (Geography Information System – GIS) để theo dõi những thay đổi trong môi trường. Nhà sản xuất thiết bị nông nghiệp John Deere cung cấp các hệ thống thu thập, phân tích, và diễn giải dữ liệu đến từ những bộ cảm biến gắn trên các trang thiết bị nông nghiệp cũng như dữ liệu đến từ các nguồn khác. Một khi John Deere đã nắm được dữ liệu về các xu hướng lịch sử và thông tin về một trang trại cụ thể, thì nếu người nông dân sở hữu trang trại đó muốn chuyển sang sử dụng nhà cung cấp trang thiết bị nông nghiệp khác, anh ta sẽ phải mất khá nhiều thời gian mới có thể xây dựng lại cơ sở dữ liệu này. Một điểm gây tranh cãi trong quá trình phát triển dữ liệu lớn: ai mới là chủ sở hữu thực sự của dữ liệu. John Deere hay người nông dân? Tùy thuộc vào câu trả lời mà phương pháp tính toán giá trị sẽ khác nhau.

Nhìn nhận sự việc từ góc độ của John Deere, việc kiến tạo giá trị cũng giống như các ví dụ đã nêu ở trên. Đối với người nông dân, việc kiến tạo giá trị là giá trị

của thặng dư từ phía khách hàng[1], nghĩa là giá trị của thông tin để có thêm lợi nhuận trừ đi bất kỳ khoản chi phí nào liên quan đến việc thu thập dữ liệu.

Quản lý chuỗi cung ứng

Chúng ta vừa nhìn vào các hoạt động của doanh nghiệp theo hướng xuôi dòng, chủ yếu tập trung vào khách hàng. Đối với nhiều tổ chức, việc quản lý chuỗi cung ứng mang lại lợi thế cạnh tranh cho họ bằng cách tiết giảm chi phí nhờ gia tăng hiệu suất hoạt động, bảo đảm rằng sản phẩm luôn có sẵn ở đúng nơi, đúng thời điểm.

Quản lý hiệu quả chuỗi cung ứng là một khía cạnh then chốt trong hiệu suất hoạt động của doanh nghiệp. Các chuỗi cung ứng đang ngày càng dài hơn và phức tạp hơn, do đó dữ liệu, dòng chảy dữ liệu, và các quy trình ra quyết định có khả năng mang lại những tác động lớn đến tiềm năng kiến tạo giá trị. Cách tạo giá trị chính trong quản lý chuỗi cung ứng là tránh để mất doanh số do hết hàng trong kho và do dự trữ các mặt hàng có tốc độ bán chậm trong kho. Gã khổng lồ trong lĩnh vực quản lý hàng tồn kho và quản lý chuỗi cung ứng là nhà bán lẻ lớn nhất thế giới,

---

[1]Thặng dư từ phía khách hàng (hay thặng dư của người tiêu dùng): Là hiện tượng xảy ra khi mức giá người tiêu dùng bỏ ra để mua một sản phẩm/dịch vụ thấp hơn mức giá họ sẵn sàng chi trả. (ND)

Walmart. Walmart vận hành một cơ sở dữ liệu lên tới 40 petabyte ghi lại các giao dịch bán hàng trong tuần trước đó. Trong bối cảnh khối lượng giao dịch lớn, môi trường kinh doanh vô cùng năng động, với các chương trình khuyến mãi mua 1 tặng 2, chương trình giảm giá, và đối thủ cạnh tranh không ngừng điều chỉnh giá cả, việc quản lý doanh thu và biên lợi nhuận xuyên suốt toàn bộ hệ thống các siêu thị của Walmart sẽ là bất khả thi nếu không có phân tích dữ liệu lớn. Ngoài dữ liệu bán hàng, Walmart còn thu thập các luồng dữ liệu từ 200 nguồn khác, chẳng hạn dữ liệu khí tượng học (doanh số bán kem gia tăng khi xuất hiện sóng nhiệt), dữ liệu kinh tế (trong túi người tiêu dùng có bao nhiêu tiền), dữ liệu mạng xã hội (ai đang nói gì về Walmart), và dữ liệu về giá xăng (giá xăng tăng nghĩa là người dân sẽ chuyển tới mua sắm ở các cửa hàng tuy nhỏ hơn nhưng ở gần nhà họ). Nhìn từ góc độ phân tích, chúng ta thấy một điểm thú vị nằm ở thách thức của việc kết hợp dữ liệu bán hàng hằng tuần vốn luôn biến động nhanh chóng và các dữ liệu về xu hướng diễn ra chậm hơn, chẳng hạn thời tiết, tình hình kinh tế, và những thay đổi trong giá xăng dầu.

Cũng có thể sử dụng chuỗi cung ứng để tạo ra sự khan hiếm. Zara là tên thương hiệu của Inditex, nhà sản xuất quần áo Tây Ban Nha. Inditex được

Amancio Ortega thành lập năm 1972, và có mô hình chuỗi cung ứng tương đối khác so với đối thủ chính của họ là H&M. H&M thiết kế và đặt hàng quần áo từ các nhà sản xuất ở tiểu lục địa Ấn Độ, thường là 6-9 tháng trước khi vào mùa chính. Để làm được điều này, bên mua cần có kỹ năng mua sắm và hoạch định tuyệt vời, nhưng nếu bên mua ước tính quá cao doanh số của một số mặt hàng, thì họ sẽ phải thực hiện những chương trình giảm giá lớn mới có thể bán hết loạt hàng của mùa trước, trước khi quần áo cho mùa sau được trưng bày ở các cửa hàng. Zara có cách làm khác hẳn. Khi Ortega mới khởi nghiệp trong lĩnh vực bán lẻ, ông tìm nguồn cung ứng quần áo từ tất cả các doanh nghiệp nhỏ, thậm chí từ các thợ may riêng lẻ ở Tây Ban Nha, thuê họ may quần áo theo đúng các yêu cầu kỹ thuật của ông. Điều này nghĩa là khi công ty dần phát triển, đội ngũ ở Zara phải học cách quản lý hiệu quả hàng nghìn nhà cung cấp. Ngày nay, Inditex được tích hợp theo chiều dọc, và phần lớn hoạt động sản xuất diễn ra ở Tây Ban Nha. Khoảng cách gần và sự hợp tác chặt chẽ mở ra một phương thức hết sức khác biệt để vận dụng dữ liệu trong chuỗi cung ứng. Zara có 7.000 cửa hàng hoạt động ở 91 quốc gia, nhưng họ vẫn có thể thiết kế, sản xuất, và chuyển giao một loạt quần áo mới sau 4-6 tuần. Các nhà thiết kế của Zara tạo ra khoảng 40.000 thiết kế mới, trong đó họ chọn ra 10.000 thiết kế để đưa vào sản xuất,

tạo ra hơn 300.000 đơn vị lưu kho mới (SKU). Chỉ có thể quản lý được những thông tin này thông qua dữ liệu, mà quan trọng hơn là thông qua việc chia sẻ dữ liệu giữa các nhà thiết kế, đơn vị sản xuất, và nhà quản lý phụ trách chuỗi cung ứng. Dữ liệu về điểm bán hàng điện tử (EPOS)[1] từ các cửa hàng luôn sẵn sàng phục vụ nhu cầu tìm hiểu của tất cả các bộ phận và đơn vị trong công ty này. Do đó, doanh nghiệp có thể nhanh chóng đưa ra các quyết định dựa trên luồng dữ liệu theo thời gian thực về mặt hàng gì đang được bán ở đâu. Giá trị được tạo ra từ khả năng xử lý khối lượng lớn dữ liệu nằm ở phần vốn lưu động trên bảng cân đối tài chính. Zara có thể giảm đáng kể số lượng sản phẩm quần áo không bán được trong chuỗi cung ứng và trong các cửa hàng. Tỉ lệ hàng không bán được trong lĩnh vực may mặc là 17-20%, trong khi với Zara, tỉ lệ này chỉ ở mức dưới 10%. Do vậy, họ ít phải thực hiện những chương trình giảm giá lớn để dọn kho cho quần áo của mùa mới, từ đó gia tăng biên lợi nhuận gộp và khả năng sinh lợi. Đôi khi người ta nói rằng Zara có khả năng dự đoán những thay đổi nhanh chóng trong ngành thời trang tốt hơn, nhưng thực ra, chính khả năng phản ứng trước những biến

---

[1]Điểm bán hàng điện tử: Là hệ thống sử dụng máy vi tính để ghi lại thông tin doanh số ở các cửa hàng bán lẻ với công cụ chính là máy quét laser ở quầy thu ngân để đọc mã vạch trên các gói hàng. (ND)

động trong ngành thời trang xuyên suốt quá trình vận hành của họ mới là yếu tố chính mang lại thành công. Tất cả những điều này đều dựa trên việc quản lý luồng chảy dữ liệu và thông tin theo thời gian thực.

Theo nghiên cứu của hãng tư vấn Deloitte, những thành tựu phát triển chính của dữ liệu lớn và công nghệ phân tích tiên tiến trong các chuỗi cung ứng hiện đang được tích hợp vào (1) các công cụ tối ưu hóa, (2) dự đoán nhu cầu, (3) hoạch định kinh doanh tích hợp, và (4) hợp tác và phân tích rủi ro. Báo cáo trên cũng chỉ ra rằng năng lực phân tích hiện đang được coi là năng lực quan trọng thứ năm, nhưng được kỳ vọng sẽ trở thành năng lực quan trọng nhất đối với hiệu suất hoạt động của chuỗi cung ứng trong tương lai. Nghiên cứu của Accenture phát hiện ra rằng có ba yếu tố thực hành chính góp phần củng cố tỉ suất lợi nhuận trên đầu tư (ROI) khi đầu tư vào công nghệ phân tích dữ liệu lớn trong chiến lược chuỗi cung ứng của doanh nghiệp. Thứ nhất, ROI sẽ tăng trưởng mạnh nếu doanh nghiệp đưa ra các quyết định đầu tư vào dữ liệu lớn từ tầm nhìn bao quát toàn diện và xuyên suốt tổ chức của mình; thứ hai, họ xác định rõ ràng các vấn đề và cơ hội mà một dự án đầu tư vào công nghệ phân tích dữ liệu lớn muốn nhắm đến, thường thì họ sẽ làm việc này dựa trên sự hậu thuẫn của các dự án thử nghiệm nhỏ nhằm xác nhận giả thiết ban đầu; và thứ ba, họ đưa hoạt động phân tích

dữ liệu lớn trở thành một phần trong các hoạt động thường nhật của công ty chứ không coi đó là một bộ phận chức năng riêng biệt trong tổ chức. Một số dữ liệu chính từ báo cáo của Accenture cho thấy:

Các kết quả từ hoạt động đầu tư vào dữ liệu lớn:

- Cải thiện ít nhất 10% chất lượng dịch vụ khách hàng và hoàn thành đơn hàng (46% người trả lời khảo sát)
- Thời gian phản ứng để khắc phục các vấn đề phát sinh trong chuỗi cung ứng được rút ngắn lại với hiệu quả cao hơn (41% người trả lời khảo sát)
- Gia tăng ít nhất 10% hiệu suất hoạt động của chuỗi cung ứng (36% người trả lời khảo sát)

Theo Accenture, việc đưa phân tích dữ liệu lớn vào hoạt động hằng ngày của công ty đã góp công lớn trong việc rút ngắn chu kỳ từ đơn đặt hàng đến khâu bàn giao sản phẩm, đồng thời gia tăng hiệu suất hoạt động của chuỗi cung ứng.

Các doanh nghiệp kinh doanh sản phẩm thường có chuỗi cung ứng vừa dài vừa phức tạp cho những yếu tố đầu vào trong quá trình sản xuất. Trong 30 năm qua, phần lớn hoạt động sản xuất linh kiện đã di chuyển từ các địa điểm gần thị trường người dùng ở Mỹ và Tây Âu sang châu Á và Trung Quốc. Việc vận

chuyển những linh kiện đơn giản từ các nhà máy sản xuất ở châu Á tới các dây chuyền sản xuất ở Mỹ và châu Âu ngày càng dễ dàng và ít tốn kém hơn. Trong 15 năm qua, số lượng sản phẩm được sản xuất và lắp ráp ở châu Á gia tăng, kết quả là, phần nhiều giá trị bổ sung diễn ra trong chuỗi cung ứng thay vì ở phía chủ sở hữu sản phẩm cuối cùng. Xu hướng này đã góp phần nâng cao kỹ năng và kiến thức về sản xuất ở châu Á, và trong trường hợp của Apple, chiếc iPhone được tạo ra từ nhiều linh kiện do hàng loạt đơn vị chuyên về linh kiện điện tử ở khắp châu Á cung cấp, và thành phẩm được lắp ráp tại công ty Foxconn ở Trung Quốc, sau đó được chuyển tới khách hàng ở khắp nơi trên thế giới. Sự phát triển của các công ty sản xuất công nghệ cao ở châu Á cũng tạo ra một thế hệ mới các công ty xông xáo, bắt đầu cạnh tranh với các công ty truyền thống của phương Tây trong lĩnh vực cung cấp dịch vụ viễn thông (ví dụ, Huawei từ Trung quốc), ô tô (ví dụ, Hyundai từ Hàn Quốc), điện thoại di động (Samsung), và nhiều công ty khác. Cạnh tranh gia tăng dẫn đến cạnh tranh về sản phẩm, đồng thời dẫn đến cạnh tranh về hiệu suất hoạt động của chuỗi cung ứng với trọng tâm là tốc độ của dòng chảy thông tin và sự điều phối nhịp nhàng giữa nhà cung cấp với đơn vị vận chuyển. Các chuỗi cung ứng được quản lý tốt có thể giúp làm giảm lượng hàng tồn kho, giảm hiện tượng hàng hóa lỗi thời, từ đó cải thiện giá

trị nhờ giảm vốn lưu động. Tiềm năng tạo giá trị nhờ giảm giá vốn hàng hóa trong bảng cân đối lỗ lãi có thể rất hấp dẫn. Những lợi ích này đối lập với các chi phí và rủi ro trong các chuỗi cung ứng phức tạp, khó quản lý. Quay trở lại với ví dụ mẫu của chúng ta (xem Bảng 3.9), giá vốn hàng bán (COGS) giảm từ 40% xuống còn 39%, tất cả các điều kiện khác như nhau, sẽ giúp gia tăng dòng tiền thêm 1 triệu đô-la từ khoản đầu tư trị giá 0,4 triệu đô-la trong Năm 1, và đạt được chỉ số NPV là 2,67 triệu đô-la.

Trong các thị trường hàng hóa tiêu dùng nhanh (FMCG), sự khác biệt giữa lợi nhuận tốt và kết quả kém nằm ở chỗ nhu cầu có thể được dự đoán chính xác tới mức nào, và nguồn cung sản phẩm nên nhắm vào đâu để giảm thiểu tình trạng hết hàng và hàng hóa không được giao tới tay người dùng cuối.

BẢNG 3.9 Tạo giá trị từ khả năng quản lý chuỗi cung ứng hiệu quả

(Đơn vị: đô-la)

	Đầu tư	*Chi phí vận hành*	*Gia tăng dòng tiền nhờ giảm COGS*	*Dòng tiền thuần*	*NPV*
Năm 1	0,4 triệu	0,1 triệu	1,0 triệu	0,5 triệu	
Năm 2		0,1 triệu	1,0 triệu	0,9 triệu	
Năm 3		0,1 triệu	1,0 triệu	0,9 triệu	
Năm 4		0,1 triệu	1,0 triệu	0,9 triệu	
Năm 5		0,1 triệu	1,0 triệu	0,9 triệu	2,67 triệu

Một công ty lớn và nổi tiếng trong lĩnh vực sản xuất kem đánh răng hiện đang sử dụng dữ liệu lớn để bảo đảm rằng sản phẩm của họ luôn luôn có mặt trên các kệ hàng. Trong ví dụ này, họ đạt được thành công không chỉ nhờ năng lực phân tích của bản thân, mà còn nhờ vai trò của các đơn vị cung cấp dữ liệu khác trong chuỗi quyết định. Hình 3.2 mô tả các nguyên tắc đằng sau dòng chảy thông tin.

Nguồn dữ liệu chính ở đây là dữ liệu về điểm bán hàng điện tử (EPOS) từ các nhà bán lẻ hoạt động rải rác ở khắp nơi trên cả nước, những nơi sản phẩm của nhà sản xuất này được bày bán. Nhà sản xuất kem đánh răng có hai lựa chọn: hoặc mua dữ liệu EPOS trực tiếp từ các nhà bán lẻ rồi tự tổng hợp dữ liệu, hoặc mua dữ liệu từ các công ty tổng hợp dữ liệu EPOS như Nielsen hay Kantar Worldpanel. Do chi phí mua và tự tổng hợp dữ liệu của nhà bán lẻ thường tương đối cao, nên phương án mang hiệu quả kinh tế cao hơn là mua dữ liệu từ nhà tổng hợp dữ liệu, vì trong gói sản phẩm họ cung cấp có cả thông tin về doanh số và giá cả của các thương hiệu kem đánh răng cạnh tranh. Dĩ nhiên, các thương hiệu kem đánh răng đối thủ có thể cũng mua những dữ liệu này, do đó lợi thế chỉ nghiêng về phía công ty nào có khả năng tốt hơn trong việc phân tích dữ liệu và đưa ra những quyết định về chính sách định giá,

chiến dịch marketing, và quản lý chuỗi cung ứng. Thực ra, tương tự như Zara, điều then chốt ở đây là khả năng chia sẻ dữ liệu với tất cả các khâu trong chuỗi cung ứng và khả năng ra quyết định, đặc biệt là ra quyết định nhanh hơn đối thủ cạnh tranh. Đây vừa là thách thức vừa là cơ hội bởi doanh nghiệp sẽ có được cái nhìn toàn diện, diễn biến theo thời gian thực về giá cả, khối lượng, và dòng chảy sản phẩm trong chuỗi cung ứng. Khối lượng dữ liệu và khả năng ra quyết định nhanh chóng là một trong những thách thức lớn của các hoạt động đầu tư vào dữ liệu lớn. Đây là lúc các phần mềm trực quan hóa như Tableau phát huy tác dụng; chúng có thể hỗ trợ các quyết định được đưa ra hằng tuần, thậm chí hằng ngày, có tác động đến lợi nhuận, và do đó tác động đến quá trình kiến tạo giá trị cho nhà sản xuất kem đánh răng. Tuy nhiên, đừng xem thường mức độ khó khăn trong việc ra quyết định xem cần trực quan hóa thông tin nào và như thế nào. Chỉ nội những quyết định này cũng có thể làm thay đổi quyết định được đưa ra và kết quả sau đó. Nhìn chung, khi một số người cùng nhau ra quyết định, các hình ảnh trực quan như biểu đồ và đồ thị thường sẽ có tác dụng tốt hơn so với hình thức trình bày bằng bảng số liệu.

Nếu một doanh nghiệp có khả năng dự đoán nhu cầu tốt hơn, nhờ đó có thể duy trì sự hiện diện liên tục

của sản phẩm trên các kệ hàng, thì xét trong bối cảnh ví dụ mẫu của chúng tôi (xem Bảng 3.10), họ có thể gia tăng doanh thu 2%, hay 2 triệu đô-la. Dòng tiền tăng thêm 1,2 triệu đô-la sẽ mang lại biên lợi nhuận gộp là 60%. Việc dòng tiền gia tăng 0,6 triệu đô-la đại diện cho chỉ số NPV là 3,34 triệu đô-la.

Như vậy, chúng ta vừa bàn đến các ví dụ tính toán giá trị được tạo ra trên cơ sở một tập hợp các giả định. Phần tiếp theo sẽ bàn về đổi mới sáng tạo trên cương vị một yếu tố thúc đẩy trong quá trình kiến tạo giá trị, đồng thời chỉ ra rằng cần có khả năng ra quyết định nhanh chóng và hiệu quả để hấp thụ được giá trị mà dữ liệu lớn mang lại.

HÌNH 3.2 Dòng chảy dữ liệu
trong lĩnh vực kinh doanh kem đánh răng



BẢNG 3.10 Tạo giá trị nhờ cải thiện chất lượng dự đoán

(Đơn vị: đô-la)

	Đầu tư	*Chi phí vận hành*	*Gia tăng dòng tiền nhờ gia tăng doanh thu*	*Dòng tiền thuần*	*NPV*
Năm 1	0,4 triệu	0,1 triệu	1,2 triệu	0,7 triệu	
Năm 2		0,1 triệu	1,2 triệu	1,1 triệu	
Năm 3		0,1 triệu	1,2 triệu	1,1 triệu	
Năm 4		0,1 triệu	1,2 triệu	1,1 triệu	
Năm 5		0,1 triệu	1,2 triệu	1,1 triệu	3,34 triệu

Khi bản thân dữ liệu là nguồn kiến tạo giá trị

Chúng ta đã tìm hiểu dữ liệu cung cấp thông tin ra sao để giúp doanh nghiệp đưa ra được những quyết định có khả năng kiến tạo giá trị. Tuy nhiên, có những công ty gặt hái được nhiều thành công xuất sắc trong đó phần kiến tạo giá trị của họ lại nằm ngay ở bản thân dữ liệu. Trong ví dụ về nhà sản xuất kem đánh răng đã nêu, chúng tôi có nhắc đến AC Nielsen và vai trò của họ trên cương vị một đơn vị cung cấp dữ liệu. Trên thực tế, AC Nielsen và một loạt các công ty khác như IMS Health đều phục vụ những thị trường rất khác nhau trên một cương vị chung là nhà cung cấp dữ liệu. Một điều thú vị là tất cả các công ty này đều được thành lập từ gần 100 năm trước, điều đó cho thấy, dữ liệu vẫn luôn luôn là một tài sản có giá trị. AC Nielsen được Arthur C. Nielsen thành lập năm 1923 tại Chicago, và nhiệm vụ của họ là cung cấp

cho các chuyên gia marketing những thông tin khách quan, đáng tin cậy về hiệu quả của các chiến dịch marketing. Đây vẫn là mục tiêu chủ yếu của họ ngày nay, nhưng thay vì cách làm trước kia là hỏi khách hàng xem họ nghĩ gì về các chiến dịch marketing, bây giờ hầu hết mọi dữ liệu đều được thu thập bằng các phương tiện điện tử thông qua nhiều nguồn khác nhau, sau đó được phân tích, đóng gói, và bán lại cho khách hàng, để khách hàng sử dụng các dữ liệu này trong quá trình ra quyết định. IMS Health kiến tạo giá trị cho bản thân và cho khách hàng của mình bằng cách sở hữu một trong những nguồn dữ liệu toàn diện nhất trong lĩnh vực chăm sóc sức khỏe. Giá trị ở đây là kiến tạo nên một mạng lưới nguồn dữ liệu như các nhà sản xuất thuốc, hãng dược phẩm, nhà bán lẻ, bệnh viện, và bất kỳ tổ chức nào khác mua và sử dụng thuốc cùng các loại sản phẩm tiêu dùng trong lĩnh vực chăm sóc sức khỏe. IMS Health có hơn 29.000 nguồn dữ liệu để họ đối chiếu, phân tích, và cung cấp trong hơn 10.000 báo cáo khác nhau rồi đem bán cho khách hàng ở khắp nơi trên thế giới.

Đổi mới sáng tạo – các quyết định dựa trên dữ liệu sử dụng dữ liệu lớn

Gần đây nổi lên một cuộc tranh luận khá gay gắt giữa các nhà kinh tế học và các nhà khoa học máy tính về việc hoạt động đầu tư vào IT và mới đây nhất

là vào các công nghệ phân tích dữ liệu lớn có thể thúc đẩy khả năng sinh lãi tới mức độ nào. Năm 1993, nhà nghiên cứu Erik Brynjofsson thuộc Viện Công nghệ Massachusetts đã giới thiệu khái niệm "nan đề của năng suất), chỉ ra rằng tuy hoạt động đầu tư vào IT có sự gia tăng vào thập niên 1980, nhưng không mang lại tác động có thể đo lường nào đến sự tăng trưởng của tổng sản phẩm quốc nội (GDP) hay các thước đo khác. Phát hiện này cũng được Robert Gordon xác nhận trong cuốn sách của ông mang tựa đề *The Rise and Fall of American Growth* (tạm dịch: Sự phát triển và thoái trào của sức tăng trưởng Mỹ) (2016). Gordon cho rằng sự phát triển về năng suất làm việc ở các doanh nghiệp Mỹ đã chững lại kể từ thập niên 1970, dẫu rằng họ đã bỏ ra những khoản đầu tư lớn trong mảng công nghệ IT. Những nghiên cứu mới đây nhất của Brynjolfsson về quá trình ra quyết định dựa trên dữ liệu (data driven decision-making – DDD) đã tìm thấy những bằng chứng cho thấy thực ra, DDD có mối liên hệ với mức gia tăng 5-6% trong sản lượng và năng suất, nhưng việc này không giải thích được thông qua các yếu tố đầu vào và cách sử dụng IT truyền thống. Brynjolfsson cho rằng bản thân hoạt động đầu tư vào IT không quan trọng, điều quan trọng là vận dụng dữ liệu thu thập được để phục vụ cho quá trình ra quyết định.

Công ty NESTA ở Vương quốc Anh đã thực hiện khảo sát đối với 500 công ty tại Anh và hỏi xem họ ra quyết định như thế nào để phát triển doanh số bán hàng. Chỉ có 18% trường hợp người tham gia khảo sát trả lời rằng công ty họ sử dụng dữ liệu và các kỹ thuật phân tích cho quá trình ra quyết định. 43% cho biết họ phần lớn hoặc hoàn toàn chỉ dựa vào kinh nghiệm và trực giác khi ra quyết định. Theo nghiên cứu này, các công ty sử dụng dữ liệu:

- Có xác suất cao gấp đôi trong việc chạy các dự án thử nghiệm trước để tìm hiểu xem điều gì có tác dụng, điều gì không.
- Có xác suất cao gấp ba lần trong việc sử dụng dữ liệu khách hàng để phát triển chiến lược kinh doanh.
- Có xác suất cao gấp 25% trong việc khẳng định rằng họ ra mắt sản phẩm và dịch vụ trước các đối thủ của mình.

Nghiên cứu này gọi các tổ chức tích cực sử dụng dữ liệu trong quá trình ra quyết định là các "datavore" (người cuồng dữ liệu). Phân tích của NESTA cho thấy nhiều doanh nghiệp sử dụng các báo cáo về xu hướng và tình báo doanh nghiệp (BI). Lỗ hổng lớn nhất là cách thức hiển thị dữ liệu – 84% datavore sử dụng phương pháp trực quan hóa và các bảng tổng kết,

trong khi số lượng công ty khác áp dụng các kỹ thuật trên chỉ là 49%. Dữ liệu cũng chỉ ra rằng cả hai nhóm này, datavore và phần còn lại, vẫn tương đối ít sử dụng các thử nghiệm và phân tích thống kê như phân tích hồi quy, do đó những công ty biết cách sử dụng dữ liệu tinh tế hơn để ra quyết định sẽ có cơ hội đạt được lợi thế trước đối thủ.

Phát hiện trên cùng với nghiên cứu của Brynjolfsson về DDD cho thấy có nhiều cơ hội lớn cho các công ty biết vận dụng quá trình ra quyết định dựa trên dữ liệu như phân tích dữ liệu lớn, các công ty này chắc chắn sẽ bỏ xa đối thủ trên thị trường. Theo nghiên cứu của NESTA, 46,4% các công ty DDD cho biết khách hàng trực tuyến có ảnh hưởng lớn đến hiệu suất hoạt động của họ, trong khi con số này ở các công ty hạn chế sử dụng dữ liệu trong quá trình ra quyết định là 12%. Vấn đề mấu chốt là, thành công trong kinh doanh là một trò chơi của lợi thế cạnh tranh tương đối, và người giành phần thắng là những doanh nghiệp sở hữu cơ sở dữ liệu tốt hơn, nhưng quan trọng hơn cả là có các quy trình ra quyết định tốt hơn.

Nghiên cứu của hãng nghiên cứu thị trường Forrester giúp chúng ta hiểu rõ hơn về cách các nhà bán lẻ vật lộn giải quyết nhu cầu về dữ liệu, và đối phó với thách thức xử lý khối lượng lớn dữ liệu.

"Các nhà bán lẻ bị choáng ngợp trước dữ liệu," công ty chuyên về dữ liệu khách hàng, Dunhummby, nhận định. Trong nghiên cứu này, 28% nhà bán lẻ cho biết việc thiếu khả năng sử dụng những dữ liệu sẵn có là một rào cản ngăn doanh nghiệp phát triển trở thành doanh nghiệp lấy khách hàng làm trung tâm; 66% nói rằng vấn đề chính là làm sao để đạt được một tầm nhìn duy nhất về khách hàng. Sự kết hợp giữa các cửa hàng vật lý với hoạt động bán hàng trực tuyến, thường được gọi là kinh doanh đa kênh, là một cơn đau đầu lớn đối với nhà bán lẻ, và 29% nhà bán lẻ tham gia trong nghiên cứu của Forrester cho hay: sự xung đột giữa chủ sở hữu các kênh bán hàng khác nhau là yếu tố ảnh hưởng đến sự tiến bộ, đồng thời làm giảm giá trị tiềm năng có thể được tạo ra. Những thông tin này cho thấy sự hợp tác và lối tư duy toàn diện, nhất quán cũng là yếu tố thúc đẩy quá trình kiến tạo giá trị, nhất là trong bối cảnh khối lượng thông tin đồ sộ và phức tạp ngày nay. Mặc dù không thuộc phạm vi đề tài của cuốn sách này, nhưng cũng xin nói thêm: rõ ràng các quyết định của cấp lãnh đạo và quản lý sẽ có tác động to lớn đến kết quả của các hoạt động đầu tư vào dữ liệu lớn.

Gọi vốn cộng đồng (crowdsourcing) và những viên gạch LEGO – nhiều người trong chúng ta đã lớn lên cùng với những món đồ chơi LEGO, tha hồ khám

phá và mặc sức để trí tưởng tượng bay cao thông qua những trò chơi xây dựng với LEGO. Điều đặc biệt với trường hợp này không chỉ là sự trường tồn của một sản phẩm đơn giản như viên gạch LEGO, mà còn ở chỗ công ty này đã nắm lấy cơ hội mà dữ liệu lớn mang lại để xây dựng và củng cố cộng đồng người dùng trực tuyến, lấy đó làm nguồn ý tưởng và thông tin phản hồi chính để đưa vào quy trình phát triển sản phẩm mới của mình. Ngày nay, LEGO có khoảng 150 nhóm người dùng và 100.000 người hâm mộ tích cực ở độ tuổi trưởng thành thường xuyên đóng góp ý tưởng và ý kiến phản hồi về các thiết kế và sản phẩm mới. Nếu một thiết kế trong cộng đồng người dùng nhận được 10.000 ý kiến tán thành, và nếu thiết kế đó có thể thương mại hóa, nhà phát minh sẽ nhận được 1% tổng giá trị doanh số ròng cho thiết kế sản phẩm đó. Năm 2011, dự án Minecraft nhận được 10.000 ý kiến tán thành từ khắp nơi trên thế giới trong vòng 48 giờ và thu hút được 30.000 lượt thích trên các trang Facebook. Trong suốt những năm qua, người hâm mộ tuổi trưởng thành của thương hiệu này đã tải lên mạng hơn 300.000 tác phẩm mà họ sáng tạo với đồ chơi LEGO, và đăng tải hơn 4,5 triệu bức ảnh, bản vẽ, và hướng dẫn. Với quy mô của cộng đồng này, việc giám sát và duy trì sự tham gia của họ sẽ thực sự khó khăn nếu thiếu những phân tích dữ liệu lớn như phân tích tâm lý trong cộng đồng và phân tích dữ liệu phi

cấu trúc của hàng nghìn cuộc trao đổi diễn ra không ngừng trên mạng internet. Giá trị được tạo ra ở đây rất dễ thấy là doanh thu bán hàng từ các thiết kế sản phẩm mới đã có ít nhất 10.000 khách hàng tiềm năng ngay từ khi chưa ra đời. Cũng nhờ cách tiếp cận này, nguồn lực phát triển sản phẩm của LEGO từ vài trăm nhân viên nội bộ nhân lên thành hàng chục nghìn người hâm mộ LEGO ở khắp nơi trên thế giới.

Một khoản chi phí lớn mà xã hội phải gánh là chăm sóc sức khỏe và chăm sóc người cao tuổi. Theo nghiên cứu của hãng tư vấn McKinsey, nếu ngành chăm sóc sức khỏe Hoa Kỳ sử dụng dữ liệu lớn để tăng cường hiệu suất hoạt động và chất lượng, thì ngành này có thể tạo ra nguồn giá trị lên tới 300-450 tỉ đô-la. McKinsey cho rằng các yếu tố thúc đẩy giá trị sử dụng phân tích dữ liệu lớn có thể mang lại những giá trị như sau:

- 70-100 tỉ đô-la từ các chương trình phòng bệnh khoanh vùng mục tiêu và giám sát việc tuân thủ chế độ điều trị của bệnh nhân với sự hỗ trợ của dữ liệu
- 90-100 tỉ đô-la từ việc hợp lý hóa các phương pháp điều trị đã được chứng thực và tăng cường sự phối hợp giữa các nhà cung cấp

- 50-70 tỉ đô-la từ việc sắp xếp môi trường chăm sóc bệnh nhân phù hợp và giảm tỉ lệ bệnh nhân tái nhập viện
- 50-100 ti đô-la từ những hoạt động đổi mới sáng tạo trong khía cạnh thanh toán và sự minh bạch hóa hoạt động
- 40-70 tỉ đô-la từ việc tăng tốc quá trình khám phá trong lĩnh vực nghiên cứu và phát triển và cải thiện chất lượng các chương trình thử nghiệm

Điều đáng lưu ý là khoảng 160-200 tỉ đô-la giá trị được tạo ra có liên quan đến cách phát triển và vận dụng hiệu quả các phương thức điều trị và các loại thuốc mới, trong khi phần còn lại của giá trị được tạo ra (140-240 tỉ đô-la) lại liên quan đến sự phối hợp giữa các bên tham gia vào hệ thống y tế, trong đó khả năng chia sẻ dữ liệu dường như là một rào cản lớn. Báo cáo trên cho thấy những tiến bộ về công nghệ trong việc phân tích và trực quan hóa khối lượng lớn dữ liệu – đặc biệt là dữ liệu phi cấu trúc từ các hồ sơ bệnh nhân – mà dữ liệu lớn mang lại đã mở ra những cơ hội tiến bộ vốn không thể có trong 20 năm về trước.

Giá trị tạo ra trong hệ thống chăm sóc sức khỏe có thể được vận dụng theo hai cách. Bằng cách cắt giảm

các khoản chi phí, chính phủ có thể hạ thấp thuế suất hoặc sử dụng tiền thuế để bổ sung thêm nhiều phương pháp điều trị bệnh khác. Tại châu Âu, McKinsey ước tính rằng các chính phủ có thể đạt được 100 tỉ euro về mặt hiệu suất hoạt động. Để đạt được giá trị tiềm năng này, không những phải đổi mới sáng tạo trong các biện pháp đo lường và độ sẵn có của dữ liệu bệnh nhân, mà phải đổi mới sáng tạo trong cách xã hội xử lý khối lượng thông tin lớn cần được chia sẻ giữa các nhà cung cấp công nghệ, hệ thống chăm sóc sức khỏe, và các cơ quan chính phủ tham gia vào việc ra quyết định cũng như phân bổ ngân sách và nguồn lực.

Dữ liệu lớn đang tạo ra những tác động mạnh mẽ đến xã hội thông qua thu thập dữ liệu hiệu quả và với khối lượng lớn. Điều này ảnh hưởng đến chúng ta trên cương vị những công dân khi tương tác với các cơ quan chính phủ chẳng hạn như các cơ quan thuế và dịch vụ xã hội. Trong cuộc trao đổi giữa một thanh tra thuế cao cấp ở Đan Mạch và một tác giả của cuốn sách này về việc truy tìm những người trốn thuế thông qua không chỉ các dữ liệu chính thức mà cả các dữ liệu phi cấu trúc từ các mạng xã hội, chưa kể đến hoạt động của các cơ quan tình báo – không gian mạng đang dần trở thành chiến trường mới để bảo vệ an ninh quốc gia.

Như vậy, chúng tôi vừa trình bày một số cách kiến tạo giá trị. Điểm quan trọng là khi cân nhắc đầu tư vào dữ liệu lớn, chúng ta phải tập trung tìm hiểu những yếu tố thúc đẩy quá trình kiến tạo giá trị. Điểm quan trọng thứ hai là phải nắm rõ được mức độ không chắc chắn trong từng yếu tố thúc đẩy giá trị, trong đó doanh thu là yếu tố kém chắc chắn nhất trong số bốn yếu tố. Điều này trên thực tế có nghĩa là phải trình bày rõ ràng và công khai về các giả định làm nền tảng cho những tính toán khi quyết định nên hay không nên đầu tư vào công nghệ phân tích dữ liệu lớn. Chúng tôi thiết nghĩ sẽ hữu ích nếu trình bày lại bảng đã giới thiệu ở phần đầu chương này (Bảng 3.11).

BẢNG 3.11 Mức độ không chắc chắn khi thu hút giá trị

Yếu tố thúc đẩy giá trị	*Hệ thống phân bậc trong kiến tạo giá trị*	*Mức độ không chắc chắn khi thu hút giá trị*
Doanh thu	Khó dự đoán và khó định lượng	Đây là yếu tố thúc đẩy giá trị có mức độ không chắc chắn cao nhất, bởi nó đưa ra những giả định dựa trên sự tăng trưởng về số lượng khách hàng hoặc sự gia tăng về giá thành ở một khối lượng doanh số cho trước.
Chi phí	Có thể dự đoán và có thể định lượng	Nhìn chung chi phí mang tính nội bộ và có mức độ chắc chắn cao hơn so với những giả định dựa trên doanh thu gia tăng.

Vốn lưu động	Có thể dự đoán và có thể định lượng	Vốn lưu động nhìn chung nằm dưới sự kiểm soát của doanh nghiệp và do đó, có thể đưa ra những dự đoán về nó với mức độ chắc chắn cao hơn so với doanh thu.
Tài sản cố định	Có thể dự đoán nhưng có thang thời gian dài hơn	Nhìn chung, doanh nghiệp nắm rõ về giá trị của các tài sản cố định mà họ sở hữu. Nếu giá trị được tạo ra bằng cách giảm bớt (loại bỏ) tài sản cố định, thì sẽ phát sinh sự không chắc chắn về giá trị của tài sản khi bán lại và thời gian cần bỏ ra để thực hiện việc loại bỏ tài sản đó.

Lưu ý: Hãy nhớ rằng trên đây chỉ là các ví dụ nhằm mục đích minh họa, và cơ hội kiến tạo giá trị thông qua dữ liệu lớn trên thực tế là rất lớn và đa dạng.

TÓM TẮT

Ở chương này, chúng tôi vừa trình bày về việc kiến tạo giá trị và vai trò của hoạt động đầu tư vào dữ liệu lớn trong việc tạo ra lợi thế cạnh tranh cho doanh nghiệp trong môi trường kinh doanh dồi dào dữ liệu ngày nay. Khái niệm giá trị trong kinh doanh ra đời từ nhiều thế kỷ trước, nhưng ý nghĩa và tầm quan trọng của nó vẫn giữ nguyên cho đến bây giờ. Chúng tôi đã bàn về việc kiến tạo giá trị tài chính và phi tài chính, đồng thời giới thiệu các ví dụ cho thấy có thể tạo giá

trị như thế nào thông qua việc sử dụng các công nghệ phân tích dữ liệu lớn. Phương pháp chúng tôi sử dụng để tính toán tiềm năng kiến tạo giá trị là xác định một trong bốn yếu tố thúc đẩy giá trị trong khía cạnh tài chính của doanh nghiệp, và sử dụng giá trị hiện tại thuần của một khoản đầu tư vào dữ liệu lớn để đưa ra quyết định cuối cùng.

4

CÁC KỸ THUẬT VÀ GIẢI PHÁP DỮ LIỆU LỚN

Ở các chương trước, bạn đã cùng chúng tôi tìm hiểu về dữ liệu, dữ liệu lớn, khoa học phân tích, các cơ hội và quy trình kiến tạo giá trị. Mục đích là để bạn tìm hiểu về việc kiến tạo giá trị từ dữ liệu ở cấp độ cao. Chương này sẽ đưa bạn tới giai đoạn tiếp theo – việc nhận thức và hiểu rõ về dữ liệu cũng như tầm quan trọng của việc kiến tạo giá trị sẽ luôn luôn gắn với sự phân tích và đúc rút ra các thông tin giá trị và kiến giải từ số dữ liệu thu thập được. Ở chương này, chúng tôi sẽ bàn đến các kỹ thuật phân tích dữ liệu khác nhau, từ các mô hình thống kê truyền thống đến các công nghệ phân tích dự đoán và các kỹ thuật trực quan hóa dữ liệu mới nhất. Trong các chương trước, chúng tôi có nói rằng phân tích dữ liệu không phải là

khái niệm mới, nhưng phương pháp thực hiện phân tích nhằm rút ra và vận dụng các thông tin có ý nghĩa từ kết quả phân tích thì đã và đang thay đổi và cải thiện đáng kể – từ phân tích văn bản, ngữ nghĩa, trực quan hóa dữ liệu, cho đến phân tích dự đoán, chúng tôi sẽ giới thiệu một số kỹ thuật phổ biến nhất trong chương này. Một số độc giả có thể sẽ cho rằng chương này phần nào mang tính chuyên môn, nhưng xin cam đoan rằng chúng tôi đã nỗ lực hết sức mình để có thể trình bày đơn giản và dễ hiểu nhất.

PHÂN TÍCH DỮ LIỆU LỚN

90% của chiến tranh là thông tin.

– Napoleon Bonaparte

Trong cuộc họp hội đồng quản trị của một công ty phần mềm trị giá nhiều triệu đô-la hoạt động trên phạm vi toàn cầu có trụ sở chính đặt tại trung tâm thành phố London, các vị lãnh đạo đang loay hoay với một câu hỏi tưởng chừng như nhỏ nhặt nhưng lại hết sức quan trọng: Vì sao doanh thu của một dòng sản phẩm chiếm đến hơn 70% thị trường Anh lại liên tục giảm?

Khung cảnh phòng họp không có gì bất thường cả. Vị giám đốc kinh doanh trình ra những số liệu kinh doanh rất khả quan, tất cả đều trên đà khởi sắc,

đồng thời không quên đổ lỗi cho bộ phận sản phẩm và đổi mới sáng tạo vì đã không đưa ra được những sản phẩm có chất lượng tốt nhất như thị trường kỳ vọng. Người đứng đầu bộ phận đổi mới sáng tạo chìa ra tất cả những bằng cấp chứng chỉ về chất lượng sản phẩm mà họ mới được cấp, và nhắc lại về lộ trình sản phẩm mà tất cả đã thống nhất từ đầu năm – trên thực tế, ông đang làm đúng theo kế hoạch. Sau đó, người đứng đầu bộ phận đổi mới sáng tạo đưa ra ý kiến rằng có lẽ nguyên do nằm ở việc đội ngũ marketing đã không nghĩ ra được chương trình quảng bá phù hợp cho sản phẩm. Lúc này, giám đốc kinh doanh đứng về phía vị giám đốc đổi mới sáng tạo để lên tiếng chỉ trích đội marketing. Ông cho rằng đội marketing không tạo được danh sách những khách hàng tiềm năng có chất lượng đội bán hàng tiếp cận, và rằng các khách hàng tiềm năng đó đều không sẵn sàng mua các sản phẩm có giá trị cao của công ty. Hai người tiếp tục nhấn mạnh rằng đội marketing hiện chỉ tập trung quảng bá cho một sản phẩm mới ra mắt thuộc phân khúc giá thấp, tuy bán được với số lượng lớn nhưng lại mang về ít doanh thu. Bây giờ, đến lượt vị giám đốc marketing đứng lên làm rõ tình hình để không phải nhận toàn bộ trách nhiệm cho sự sụt giảm doanh thu của công ty. Mặc dù ông không nhất trí với nhiều điểm được nêu ra trong cuộc tranh luận này, nhưng ông cũng chỉ ra rằng tất cả các nỗ

lực marketing đều được thực hiện dựa trên kế hoạch đã thống nhất từ trước, và không có lý do gì để đội bán hàng không mang thêm doanh thu về cho công ty cả. Nhưng bên cạnh đó, ông cũng đưa ý kiến rằng có lẽ không ai trong căn phòng này được tiếp cận với những thông tin xác đáng, có thể giải thích cho số liệu doanh thu hiện tại. Vị giám đốc marketing yêu cầu đội IT trình bày chi tiết – có thể vẫn còn thiếu điều gì đó trong những dữ kiện và thực tế đang được trình bày, và cần phải điều tra làm rõ. Hội đồng quản trị quyết định cho vị lãnh đạo bộ phận, IT bốn tuần để trả lời và đưa ra những kiến giải, nếu có thể.

Vị giám đốc IT trao đổi với đội ngũ các chuyên gia phân tích của mình, lập ra một tiểu ban chuyên trách nhiệm vụ này, và yêu cầu họ xác định, kết nối, và phân tích tất cả các dữ liệu mà công ty đang thu thập từ nhiều nguồn khác nhau rồi báo cáo kết quả cho ông sau hai tuần nữa.

Đội phân tích tiếp xúc với các phòng ban chức năng khác, bao gồm kinh doanh, đổi mới sáng tạo, tài chính, và marketing để tìm hiểu xem họ lưu trữ các loại dữ liệu khác nhau ở đâu và như thế nào. Thách thức đầu tiên mà tiểu ban hành động mới thành lập gặp phải là các dữ liệu bị cô lập và không "giao tiếp" với nhau. Đây là một tình huống điển hình trong phần lớn các tổ chức, nơi thiếu vắng hoàn toàn luồng chảy

dữ liệu tích hợp giữa các bộ phận phòng ban chức năng khác nhau. Đội IT phải nhanh chóng xây dựng một nền tảng để đưa tất cả các dữ liệu về cùng một nơi và bắt tay vào ráp nối các thông tin để tìm hiểu xem họ có thể nhìn ra bất kỳ thông tin hoặc xu hướng nào trong các luồng dữ liệu khác nhau hay không. Tất cả đều được họ phân tích – từ các tài liệu hướng dẫn đội bán hàng qua điện thoại đến các thông điệp truyền thông, từ thông tin liên hệ của khách hàng tiềm năng trong hệ thống quản lý quan hệ khách hàng đến xu hướng tiêu dùng của các khách hàng mang lại giá trị lớn. Sau đó, công ty này phát hiện ra một số dữ kiện và ráp nối mọi việc lại với nhau, và họ bắt đầu thấy những lý do hợp lý để giải thích cho doanh thu sụt giảm của công ty. Một trong những điểm chính họ rút ra được sau quá trình phân tích là khoản phí cố định mà các khách hàng giá trị cao phải trả để duy trì hệ thống phần mềm kế thừa do công ty lắp đặt trên máy chủ của họ. Doanh thu từ việc duy trì các hệ thống đó chiếm 35% doanh thu của toàn công ty, và phần lớn số doanh thu này đều đóng góp trực tiếp vào lợi nhuận. Gần đây, do những thay đổi về điều kiện kinh tế, nhiều khách hàng đã phải nghĩ đến các ý tưởng nhằm cắt giảm chi phí, và một trong số đó là ngừng sử dụng hệ thống phần mềm kế thừa này để chuyển sang các sản phẩm phần mềm khác hoạt động trên nền tảng web và có chi phí thấp hơn. Các chuyên gia

phân tích cũng phát hiện ra rằng hồi đầu năm, công ty họ đã giới thiệu một sản phẩm SAAS (software as a service – phần mềm như một dịch vụ) cũng hoạt động trên nền tảng web, và đây là một trong những sản phẩm trọng tâm trong kế hoạch marketing hiện tại. Theo chiến lược thu nạp những khách hàng đầu tiên cho sản phẩm mới ra mắt này, đội ngũ bán hàng sẽ được thưởng lớn nếu tìm được khách hàng. Vì thế, họ cố gắng hết sức để tìm kiếm khách hàng mới cho sản phẩm này, vốn đòi hỏi chi phí ban đầu thấp, không cần ký hợp đồng dài hạn, và không có phí duy trì thường xuyên. Hiển nhiên, sản phẩm mới đã ăn lạm vào phần doanh thu kinh doanh các hệ thống phần mềm kế thừa và gây ảnh hưởng lớn đến phần doanh thu từ khoản phí duy trì hệ thống.

Một phát hiện khác có liên quan đến chức năng của bộ phận kinh doanh, khi lực lượng bán hàng phụ trách việc bán các hệ thống phần mềm kế thừa cũ cũng được giao phụ trách bán sản phẩm SAAS mới. Đội ngũ bán hàng qua điện thoại sẽ được thưởng nếu đặt được các cuộc hẹn với khách hàng tiềm năng, và dĩ nhiên, họ sẽ dễ dàng thu hút được sự quan tâm của khách hàng tiềm năng hơn nếu giới thiệu sản phẩm mới vì sản phẩm mới có mức giá thấp và không đòi hỏi nhiều sự cam kết từ phía khách hàng. Chu kỳ kinh doanh cho sản phẩm mới hoạt động trên nền

tảng web cũng được giảm từ mức trung bình 6 tháng xuống còn 1 tháng, và sự hài lòng của khách hàng ngày càng tăng lên do đây là một sản phẩm tinh gọn, dễ sử dụng, đòi hỏi chi phí thấp nhưng mang lại giá trị cao. Cuối cùng, tổng doanh thu đến từ hoạt động bán hàng giảm đi bởi vì phần lớn đội ngũ bán hàng và bán hàng qua điện thoại đều tập trung quảng bá cho sản phẩm mới giá thấp, và thậm chí ngay cả khi số lượng các giao dịch bán hàng tăng lên nhanh chóng, doanh thu tổng vẫn bị ảnh hưởng nặng nề. Cùng lúc đó, phí duy trì đối với các khách hàng cũ cũng giảm đi – sự thay đổi này không hoàn toàn do việc ra mắt hay tích cực quảng bá cho sản phẩm mới nhưng sự suy giảm doanh thu từ phí duy trì hệ thống cũng góp phần không nhỏ vào tình hình doanh thu tổng giảm sút.

Khi những phát hiện này được đem ra trình bày trước hội đồng quản trị công ty, các thành viên trong hội đồng có những cảm xúc trái ngược và lẫn lộn – thay vì đổ lỗi cho nhau, giờ đây họ đã nhận ra rằng mọi phòng ban đều đang làm việc hết sức mình, chỉ có điều họ làm việc vì bản thân chứ không phối hợp cùng nhau. Do vậy xuất hiện nhu cầu phải có một cách tiếp cận từ góc nhìn tập thể và một kế hoạch dài hạn, có thể mang lại giá trị ở cấp độ toàn công ty thay vì cấp độ các phòng ban riêng biệt.

Sau vài cuộc trao đổi, họ xác định được một nhu cầu rõ ràng: cần phải vạch ra một chiến lược mới, không chỉ để giải quyết vấn đề doanh thu trước mắt, mà còn nhằm giúp tất cả các bộ phận trong công ty phối hợp hoạt động nhịp nhàng với nhau. Cấu trúc định giá cho khách hàng hiện tại và cho các hệ thống kế thừa được xác định lại cho phù hợp với những thay đổi gần đây trong nền kinh tế, đồng thời giải quyết các nhu cầu của khách hàng nhằm xây dựng mối quan hệ làm ăn dài hạn hơn với họ. Lực lượng bán hàng được chia thành hai nhóm – một nhóm tập trung vào việc bán hệ thống kế thừa cũ, và một nhóm tập trung thúc đẩy doanh số cho sản phẩm SAAS có chi phí thấp và vận hành trên nền tảng web mới ra mắt. Ngoài ra, đội ngũ phát triển sản phẩm cũng được tham gia trong quá trình này để quyết định xem cần có những đặc điểm nào để khuyến khích khách hàng chuyển từ các sản phẩm chi phí thấp trên web sang các hệ thống tiên tiến có chi phí cao hơn. Dựa trên ý kiến phản hồi của đội ngũ phát triển sản phẩm, họ quyết định loại bỏ một số đặc điểm trong sản phẩm SAAS, và những khách hàng có nhu cầu sử dụng các đặc điểm cao cấp này sẽ được giới thiệu hệ thống tiên tiến có giá trị cao.

Đội marketing tiến hành phân khúc thị trường cho sản phẩm mới và bắt đầu quảng bá sản phẩm mới cho các doanh nghiệp không có kinh phí cho hệ

thống phần mềm kế thừa, và cũng không cần đến một hệ thống tiên tiến như vậy. Ngân sách cho chiến dịch marketing cũng được chia rõ cho các hoạt động quảng bá sản phẩm mới và sản phẩm cũ. Sau khi có dữ liệu về kết quả thực hiện chiến dịch, và tính toán tỉ suất ROI liên quan đến số lượng khách hàng mới cho sản phẩm SAAS, họ quyết định đẩy mạnh đầu tư vào việc quảng bá cho sản phẩm mới.

Việc tính toán doanh thu cho hai sản phẩm cũng được thực hiện minh bạch. Mỗi nhóm phụ trách từng sản phẩm mới và cũ đều được xác định chỉ số đo lường hiệu suất công việc riêng biệt và rõ ràng, với những mục tiêu phù hợp.

Trong vòng sáu tháng tiếp theo, công ty này đã ghi nhận nhiều kết quả tích cực, và doanh thu tổng cũng bắt đầu tăng. Trong trường hợp ví dụ này, những thông tin có ý nghĩa và những phát hiện chính đều đến từ quá trình phân tích dữ liệu – những thông tin này không có gì mới, nhưng đã được hoàn thiện hơn và có sự hỗ trợ của dữ liệu. Sau khi đội ngũ thực hiện đưa ra các kiến nghị và đề xuất, lãnh đạo tất cả các bộ phận đều chấp nhận và bắt tay vào triển khai, vì họ không chỉ nhìn được rõ ràng các xu hướng trong đó mà còn có thể kết nối các đầu mục thông tin để cùng nhau mang lại thành công cho tập thể.

Dữ liệu (hay dữ liệu lớn) không thể mang lại bất kỳ giá trị nào nếu nó tồn tại trong môi trường chân không – trường hợp ví dụ vừa bàn ở trên minh họa rõ ràng cho điểm này. Doanh nghiệp sẽ thu được giá trị thực sự của dữ liệu lớn khi việc xử lý luồng dữ liệu khổng lồ này cùng các kết quả phân tích chuyên sâu trang bị cho các nhà quản lý khả năng đưa ra những quyết định nhanh chóng và hiệu quả. Việc ra quyết định dựa trên bằng chứng như vậy giúp bảo đảm sự phát triển dài hạn cho công ty. Phân tích dữ liệu lớn là quá trình khám phá dữ liệu lớn để phát hiện ra những mô hình ẩn, những mối tương quan khả thi, cùng các thông tin hữu ích và/hoặc bị ẩn khác, có thể sử dụng để đưa ra những quyết định chính xác hơn. Với khoa học phân tích dữ liệu lớn, các nhà khoa học dữ liệu cũng như những người khác có thể phân tích nhiều khối lượng lớn dữ liệu đang chảy không ngừng một cách nhanh chóng và đơn giản – điều vốn không thể thực hiện nếu sử dụng các kỹ thuật phân tích và giải pháp tình báo doanh nghiệp kiểu cũ. Để hiểu rõ hơn về sự khác biệt giữa nhu cầu về phân tích dữ liệu lớn và phân tích kiểu cũ, hãy giả sử rằng công ty của bạn mỗi giờ đều tích lũy được hàng tỉ dòng dữ liệu với hàng trăm triệu cách kết hợp dữ liệu khác nhau trong nhiều kho dữ liệu và nhiều hình thức đa dạng. Nhiệm vụ của bạn là phải chạy các phân tích theo thời gian thực để ra được kết quả nhanh chóng nhằm

điều chỉnh hiệu quả hoạt động trực tuyến của cửa hàng – bạn không thể thực hiện loại hình phân tích này với các kỹ thuật phân tích kiểu cũ đâu. Việc phân tích luồng dữ liệu hoạt động với hiệu suất cao trong những trường hợp như thế này phải được thực hiện bằng các kỹ thuật phân tích dữ liệu lớn mới có thể phân định theo thời gian thực điều gì là quan trọng, điều gì không quan trọng.

Phân tích dữ liệu lớn kiểm tra nhiều khối lượng lớn dữ liệu để phát hiện ra những xu hướng ẩn, những mối tương quan, và rất nhiều thông tin sâu sắc khác. Lúc này, chúng ta đã hoàn toàn có thể phân tích những khối lượng lớn dữ liệu và cung cấp kết quả phân tích theo đúng hình thức kỳ vọng, gần như theo thời gian thực – dĩ nhiên, việc thực hiện các nhiệm vụ này sẽ chậm chạp và kém hiệu quả hơn nếu sử dụng các giải pháp tình báo doanh nghiệp truyền thống.

Phân tích dữ liệu lớn giúp các tổ chức kiểm soát và điều khiển được dữ liệu, sử dụng nó để tìm kiếm các cơ hội mới. Rốt cuộc, điều này sẽ giúp họ từng bước hoàn thiện cách thức vận hành tổ chức, giới thiệu ra thị trường những sản phẩm và dịch vụ mang tính cạnh tranh cao hơn, cho ra đời những sản phẩm mới mang tinh thần đổi mới sáng tạo cao, và tạo ra sự hài lòng cho khách hàng. Các tổ chức có thể tạo ra giá trị với dữ liệu lớn ở ba khía cạnh chính.

Xuất sắc trong vận hành

Các công nghệ dữ liệu lớn như Hadoop, lưu trữ đám mây, và phân tích dữ liệu làm giảm đáng kể các khoản chi phí liên quan đến việc lưu trữ những khối lượng dữ liệu khổng lồ. Ngoài ra, các hoạt động phân tích thực hiện trên các tập dữ liệu này cuối cùng sẽ giúp doanh nghiệp xác định được đâu là những điểm cần cải thiện trong nhiều quy trình nghiệp vụ khác nhau, từ đó giúp họ tiết kiệm được chi phí và sắp xếp hợp lý hơn các quy trình vận hành bên trong tổ chức.

Quá trình ra quyết định thông minh

Dù trong mảng kinh doanh, marketing hay thiết lập chiến lược, khi các quyết định được đưa ra nhanh chóng và có sự hậu thuẫn của dữ liệu, chúng sẽ mang đến những kết quả ấn tượng cho doanh nghiệp. Với năng lực xử lý nhanh chóng của cơ sở hạ tầng cùng các công cụ phân tích dữ liệu lớn, doanh nghiệp có thể thực hiện phân tích dữ liệu ngay lập tức, nhanh chóng rút ra được những thông tin có ý nghĩa và nhanh chóng ra quyết định dựa trên những gì họ vừa đúc rút được từ nguồn dữ liệu sẵn có. Như vậy, doanh nghiệp hầu như có thể tránh được hoàn toàn việc đoán mò, đồng thời gieo vào trong nền văn hóa của tổ chức những hạt giống của quá trình ra quyết định nhanh chóng và thông minh.

Đổi mới sáng tạo

Khác với trước đây, các doanh nghiệp ngày nay có thể tiếp cận và thu thập các quan điểm, ý kiến phản hồi, và đánh giá của khách hàng về các sản phẩm hoặc dịch vụ họ đang dùng hoặc về giá trị họ mong muốn nhận được từ các sản phẩm hoặc dịch vụ đó. Không chỉ có vậy, các doanh nghiệp bây giờ còn xác định được những thông tin chưa được kể về sản phẩm/dịch vụ của họ hoặc về các cơ hội, sau đó sử dụng các thông tin trên để giới thiệu những sản phẩm mới mang tính sáng tạo. Vài năm trước, một nhóm kỹ sư của hãng IBM phát hiện ra rằng có thể làm giảm tác động của chứng tiểu đường bằng một số liệu pháp âm nhạc nhất định, và chưa có bất kỳ ai trên thế giới chính thức hỏi đến, nói đến, hay thu thập dữ liệu về thông tin này – nhưng IBM đã xác định được một xu hướng như vậy trong một số cuộc trao đổi trên Twitter, và sau đó tiến hành các nghiên cứu sâu thêm để rút ra được tất cả những kiến thức liên quan tới thông tin này. Ngày càng nhiều thông tin hơn từ khách hàng – về các nhu cầu hay ý kiến phản hồi của họ – giúp doanh nghiệp đưa ra được nhiều sản phẩm và dịch vụ sáng tạo hơn.

Với khả năng đo lường nhu cầu và sự hài lòng của khách hàng thông qua phân tích dữ liệu, doanh nghiệp sẽ có thể cung cấp cho khách hàng thứ họ muốn.

Davenport đã chỉ ra rằng với sự xuất hiện của kỹ thuật phân tích dữ liệu lớn, nhiều công ty hiện đang sáng chế ra những sản phẩm mới để thỏa mãn các nhu cầu của khách hàng.

Như đã đề cập từ trước, phân tích dữ liệu lớn không phải ý tưởng mới trong thế giới kinh doanh, nhưng ngày nay, nó đang từ vị trí thấp kém đàng hoàng bước vào phòng họp của ban lãnh đạo các công ty.

CÁC KỸ THUẬT PHÂN TÍCH DỮ LIỆU

Giá của ánh sáng thấp hơn chi phí của bóng tối.
– Arthur C. Nielsen (nhà sáng lập công ty nghiên cứu thị trường AC Nielsen)

Ở các chương trước, chúng ta đã cùng bàn đến dữ liệu – dữ liệu có cấu trúc, dữ liệu phi cấu trúc, và dữ liệu lớn. Bây giờ, chúng ta có thể bàn đến công việc xử lý số dữ liệu mà các tổ chức đang thu thập đều đặn theo khối lượng lớn từ các nguồn bên trong và bên ngoài – bao gồm dữ liệu công khai hoặc dữ liệu từ các đối tác kinh doanh của họ. Hiện tượng sử dụng lẫn lộn hai khái niệm "xử lý" và "phân tích" rất phổ biến – bạn có thể sử dụng nhiều công cụ và kỹ thuật khác nhau cho mục đích này, và mỗi công ty lại có cách dùng thuật ngữ khác nhau. Tuy nhiên có một điều mà ai cũng thấy rõ ràng: có thể thu được giá trị

của dữ liệu lớn theo nhiều cách khác nhau. Có nhiều cách lưu trữ và phân tích dữ liệu lớn, và bạn sẽ thấy rằng một số kỹ thuật và phương pháp không hoàn toàn mới hoặc chỉ dành riêng cho dữ liệu lớn – các kỹ thuật phân tích cơ bản đã được xác định trước trong lĩnh vực thống kê, và ứng dụng của chúng có sự thay đổi theo những trường hợp khác nhau. Chúng ta đã chứng kiến sự ra đời của nhiều giải pháp công nghệ mới (chúng tôi sẽ bàn đến vấn đề này sau) và nhiều cách thức mới để định dạng trước các loại hình phân tích khác nhau, và chúng ta sẽ thấy nhiều kỹ thuật và phương pháp mới mẻ khác được bổ sung vào danh sách này trong tương lai.

Thực ra, chưa có ai thực hiện phân nhóm các kỹ thuật phân tích này; ở đây chúng tôi cố gắng sắp xếp chúng theo trật tự logic, nhưng một số kỹ thuật có thể cùng thuộc về vài phân nhóm khác nhau. Cũng có thể bạn sẽ thấy rằng không phải mọi kỹ thuật hay công cụ đều áp dụng được cho công ty của mình. Bên cạnh đó, chúng tôi cũng sẽ trình bày vắn tắt thông tin giới thiệu về các kỹ thuật này, vì chúng tôi cho rằng các nhà quản lý nên có kiến thức ở tầm tổng quan về các kỹ thuật phân tích, nhưng họ có thể không cần tham gia vào quá trình thực hiện – đội ngũ chuyên trách mảng công nghệ trong tổ chức có thể đảm nhiệm phần việc này. Ở mức độ tổng quan, có thể chia các

kỹ thuật phân tích dữ liệu lớn thành các phân nhóm như sau:

- Phân tích thống kê
- Phân tích mạng xã hội
- Phân tích ngữ nghĩa
- Trực quan hóa dữ liệu
- Phân tích dự đoán
- Các loại hình phân tích khác

Phân tích thống kê

Trong kỷ nguyên thông tin này, dữ liệu không còn là một loại hàng hóa hiếm nữa – nó đã trở nên quá mạnh mẽ. Giá trị chính nằm ở việc triển khai các biện pháp lưu trữ dữ liệu, và quan trọng hơn, ở việc xử lý và phân tích nhiều khối dữ liệu lớn mà doanh nghiệp thu thập được và diễn giải chính xác những ý nghĩa của chúng. Để tạo ra được những thông tin chính, có thể bạn sẽ cần đến dữ liệu thống kê và các công cụ phân tích. Loại hình phân tích này không phải khái niệm mới trong thế giới kinh doanh, và là một phần hữu cơ trong các hệ thống tình báo doanh nghiệp cổ điển, nhưng với sự trỗi dậy của dữ liệu lớn, bây giờ chúng ta có thể tạo ra những giá trị tốt hơn vẫn với các phương pháp thống kê truyền thống.

Mối tương quan và hồi quy

Đây là hai trong số các phương pháp thống kê phổ biến nhất được các doanh nghiệp sử dụng từ nhiều năm nay để thiết lập mối quan hệ giữa hai biến độc lập hoặc phụ thuộc lẫn nhau, là nguyên nhân dẫn đến hoặc tác động đến sự thay đổi ở nhau. Ngày nay, với dữ liệu lớn, độ chính xác của kết quả phân tích có thể được củng cố thêm bởi các doanh nghiệp có thêm thông tin đến từ nhiều nguồn bên trong và bên ngoài tổ chức – nghĩa là có thêm nhiều thông tin đổ về và các doanh nghiệp vẫn có năng lực lưu trữ và xử lý các dữ liệu này. Khi sử dụng kỹ thuật phân tích mối tương quan, người ta có thể hiểu được mối quan hệ giữa hai biến, nhưng không thể điều chỉnh các biến này để thử nghiệm. Ví dụ, sử dụng kỹ thuật phân tích mối tương quan, người ta có thể thiết lập mối liên hệ sau: học sinh có điểm số cao hơn ở trường học thường có xác suất đạt điểm cao hơn trong các kỳ thi SAT, nhưng mối liên hệ này không thể giúp bạn xác định được rằng học sinh cần phải đạt được điểm số bao nhiêu ở trường để có thể đạt một mức điểm nhất định trong kỳ thi SAT. Ngoài ra, bạn cũng không thể xác định được quan hệ nhân quả khi sử dụng kỹ thuật phân tích mối tương quan.

Mối tương quan thuận tồn tại khi hai biến có xu hướng di chuyển theo cùng một chiều. Đôi khi, một trong hai biến này có tác động đến biến còn lại.

Một nhân tố hoặc một biến thứ ba cũng có thể tác động đến hai biến độc lập. Có thể tìm kiếm các ví dụ về mối tương quan thuận trong các lý thuyết kinh tế hiện đại. Chẳng hạn, luật cung cầu xác định mối tương quan thuận giữa hai biến. Đây là một trong những bài học đầu tiên mà bất kỳ sinh viên nào thuộc chuyên ngành kinh tế học hay xác suất học đều phải biết.

Luật cung cầu phát biểu rằng, khi tất cả các điều kiện khác giữ nguyên, trong một thị trường cạnh tranh, khi nhu cầu đối với một mặt hàng hoặc dịch vụ gia tăng, giá cả cũng gia tăng đến mức nguồn cung cũng gia tăng theo và như vậy, số lượng hàng hóa được yêu cầu và số lượng hàng hóa đáp ứng trở nên ngang bằng nhau, từ đó hình thành trạng thái cân bằng. Lý do là vì khi nhu cầu gia tăng và khách hàng sẵn sàng chi trả ở mức giá cao hơn, các nhà sản xuất và nhà cung cấp cũng bắt đầu cung cấp thêm nhiều hàng hóa ở mức giá cao hơn. Ngược lại, khi nhu cầu đối với một sản phẩm hay dịch vụ giảm đi, giá cả giảm xuống, nguồn cung cũng sụt giảm và thế cân bằng lại được thiết lập một lần nữa. Điều này tạo nên mối tương quan thuận giữa cung và cầu, đồng thời tạo nên mối quan hệ nhân quả. Nhu cầu gia tăng kéo theo nguồn cung gia tăng, và ngược lại. Tuy nhiên, giữa nguồn cung và giá cả lại có mối tương quan nghịch. Khi nguồn cung giảm, giá cả sẽ tăng lên vì nhiều người cùng tranh giành nhau để mua một lượng hàng hóa không đổi.

Giữa chi tiêu tiêu dùng và GDP (tổng sản phẩm quốc nội) cũng có mối tương quan thuận. Khi người tiêu dùng tiêu thụ nhiều hơn, nghĩa là chi tiêu nhiều hơn, các nhà sản xuất hàng hóa và dịch vụ sẽ sản xuất với số lượng lớn hơn, khiến GDP tăng lên. Khi hoạt động chi tiêu của người tiêu dùng giảm đi, các nhà sản xuất sẽ sản xuất ít hơn so với giai đoạn trước đó để kiểm soát chi phí, bởi lúc này người tiêu dùng có ý muốn bỏ ra một số tiền ít hơn để mua cùng một loại sản phẩm/dịch vụ.

Hồi quy là một phương pháp thống kê được sử dụng trong lĩnh vực tài chính, đầu tư, và các lĩnh vực khác nhằm xác định độ mạnh của mối quan hệ giữa một biến phụ thuộc (thường được quy định gọi là Y) với một loạt các biến số thay đổi khác (gọi là biến độc lập). Phương pháp hồi quy giúp các nhà quản lý đầu tư và tài chính đánh giá giá trị sản phẩm và tìm hiểu mối quan hệ giữa các biến số, chẳng hạn giữa giá hàng hóa với cổ phiếu của các doanh nghiệp cung cấp những loại hàng hóa đó.

Hai loại hồi quy cơ bản là hồi quy tuyến tính và hồi quy tuyến tính đa biến, và có các phương pháp hồi quy phi tuyến tính phục vụ cho những dữ liệu và phân tích phức tạp hơn. Hồi quy tuyến tính sử dụng một biến độc lập để giải thích hoặc dự đoán về kết quả của biến phụ thuộc Y, trong khi hồi quy đa biến sử dụng hai hoặc nhiều hơn hai biến độc lập để dự đoán kết quả.

Hồi quy có thể hỗ trợ cho các chuyên gia hoạt động trong lĩnh vực tài chính và đầu tư cũng như các chuyên gia trong nhiều lĩnh vực ngành nghề khác. Hồi quy có thể giúp một doanh nghiệp dự đoán doanh số dựa trên điều kiện thời tiết, kết quả kinh doanh trong giai đoạn trước, tăng trưởng GDP, hoặc các điều kiện khác. Hình thức chung của từng loại hồi quy là:

Hồi quy tuyến tính: $Y = a + bX + u$

Hồi quy đa biến: $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_tX_t + u$

Trong đó:

Y = biến số mà bạn đang tìm cách dự đoán (biến phụ thuộc)

X = biến số mà bạn sử dụng để dự đoán Y (biến độc lập)

a = hệ số chặn

b = độ dốc

u = phần dư hồi quy

Khi thực hiện phân tích hồi quy, người ta lấy một nhóm biến số ngẫu nhiên được cho là dự đoán được giá trị Y, và cố gắng tìm ra một mối quan hệ toán học giữa chúng. Sự liên hệ giữa các biến thường được biểu thị trong hình thức một đường thẳng.

Phân tích hồi quy thiết lập mối quan hệ giữa các biến độc lập và phụ thuộc để phân tích và dự đoán tác động đến biến phụ thuộc bằng cách điều chỉnh biến độc lập. Phân tích hồi quy có nhiều ứng dụng trong kinh doanh. Giả sử bạn phụ trách đội ngũ bán hàng trong công ty; bạn có thể dự đoán khối lượng hàng hóa bán ra và doanh thu dựa trên số lượng tiền dành cho các hoạt động quảng cáo và quy mô của đội ngũ bán hàng. Dĩ nhiên, trong đời thực, bạn sẽ có nhiều biến độc lập khác cần cân nhắc cho phân tích loại này, và giờ đây, với sự hỗ trợ của dữ liệu lớn, bạn có thể dung nạp và phân tích nhiều biến độc lập hơn để có thể dự đoán chính xác hơn về doanh số bán hàng và qua đó điều chỉnh chiến lược kinh doanh. Bây giờ bạn có thể dễ dàng tiếp cận với dữ liệu, chẳng hạn thông tin về thiết bị mà khách hàng đang sử dụng, và họ có đặt hàng trong khi kết nối Wi-Fi hoặc sử dụng gói dữ liệu trên điện thoại di động không. Những thông tin này có thể mang lại giá trị rất lớn giúp bạn đưa ra những quyết định mang tính chiến lược về việc ra mắt một ứng dụng di động có các tính năng với mức độ sử dụng dữ liệu ít.

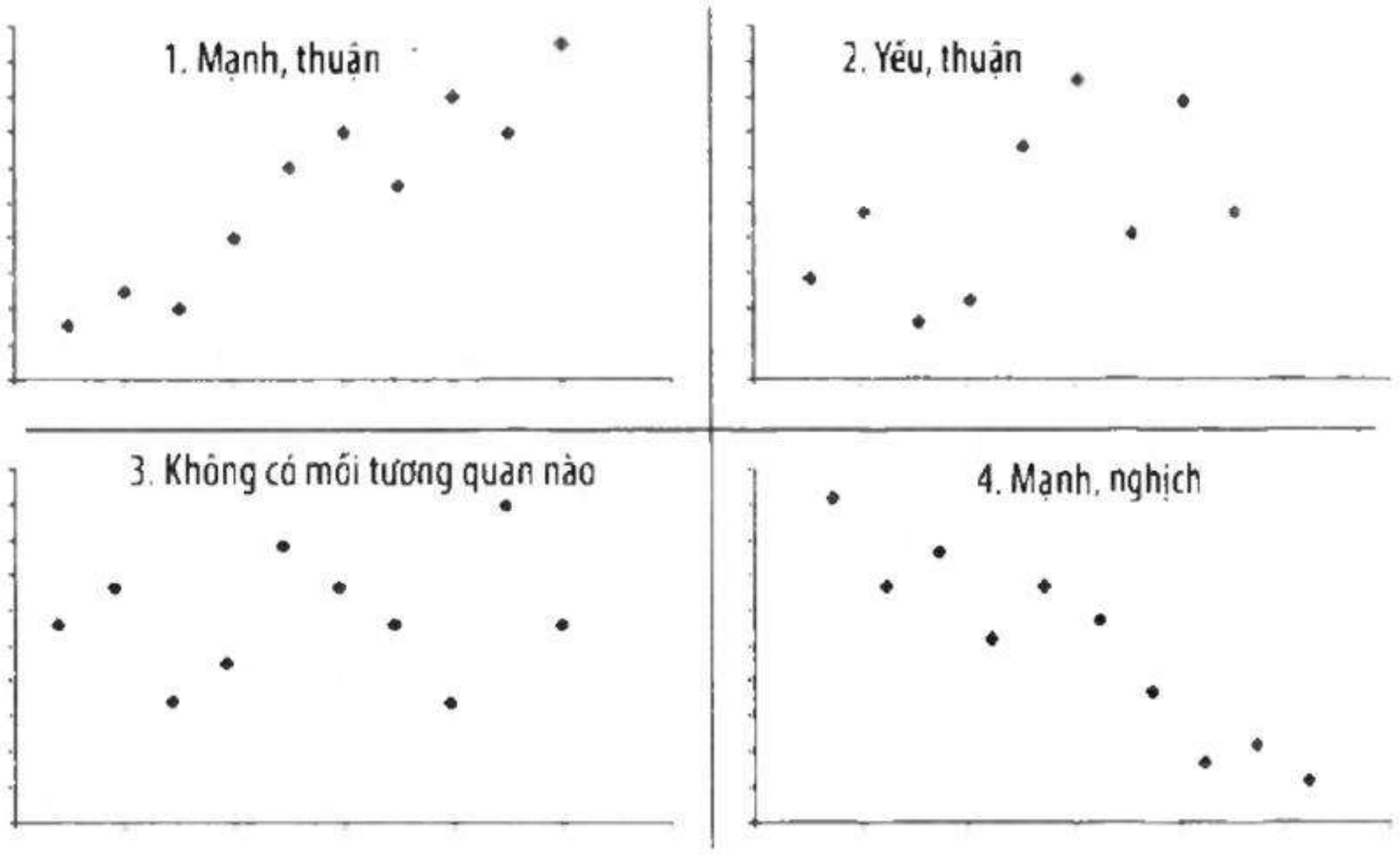
Hình 4.1 minh họa bốn kịch bản giả tưởng, trong đó một biến liên tục được biểu thị dọc trục X, và biến còn lại được biểu thị trên trục Y.

- Kịch bản 1 mô tả một mối quan hệ thuận mạnh mẽ, tương tự như mối tương quan giữa

cân nặng và chiều cao của trẻ sơ sinh khi mới chào đời.

- Kịch bản 2 mô tả một mối quan hệ yếu, giống như mối tương quan giữa chỉ số về trọng lượng cơ thể với chỉ số về tuổi tác (trọng lượng cơ thể thường gia tăng theo tuổi tác).
- Kịch bản 3 có thể mô tả sự thiếu vắng mối quan hệ giữa mức độ tiếp cận truyền thông ở độ tuổi vị thành niên với độ tuổi mà người ở tuổi vị thành niên bắt đầu có quan hệ tình dục.
- Kịch bản 4 có thể mô tả mối quan hệ nghịch mạnh mẽ thường thấy giữa số giờ tập aerobic mỗi tuần và tỉ lệ chất béo trong cơ thể.

HÌNH 4.1 Biến hồi quy liên tục được biểu diễn trên trục X và Y



Một ứng dụng lớn khác của phân tích hồi quy là nhằm tối ưu hóa các quy trình nghiệp vụ. Chẳng hạn, nhà quản lý tổng đài có thể thiết lập mối quan hệ giữa thời gian chờ của người gọi với số lượng các lời phàn nàn được gửi đến, từ đó tìm cách giảm thiểu thời gian chờ bằng cách đưa ra những quy trình hoặc nguồn lực mới. Người thợ làm bánh có thể thiết lập mối quan hệ giữa thời hạn sử dụng bánh với nhiệt độ trong lò nướng, từ đó có thể điều chỉnh nhiệt độ để những chiếc bánh của anh ta có thời hạn sử dụng dài hơn.

Dữ liệu lớn mang lại rất nhiều điểm dữ liệu, mở ra những cơ hội phân tích và cải thiện lớn cho các doanh nghiệp.

Kiểm định A/B (kiểm định giả thiết thống kê)

Kiểm định A/B là một kỹ thuật đã có từ lâu được sử dụng cho thử nghiệm ngẫu nhiên mà các doanh nghiệp và chuyên gia thống kê hay dùng. Trong một môi trường có kiểm soát, doanh nghiệp chạy các chương trình thử nghiệm trên nhiều hơn một biến với một chuẩn đo để xác định thành công hay thất bại của thử nghiệm, và sau đó phân tích kết quả.

Nói cách khác, trong kiểm định A/B, người ta lấy hai nhóm người dùng để so sánh và đưa ra cho họ hai phiên bản khác nhau của một sản phẩm (phiên bản đối chứng và phiên bản biến thể). Sau đó, khi đo lường

phản ứng và thái độ của từng nhóm người dùng với sản phẩm, người ta hy vọng suy ra được phiên bản nào có thể đáp ứng mục đích của doanh nghiệp tốt hơn.

Nhiều ngân hàng bán lẻ sử dụng kỹ thuật này để tối ưu hóa các sản phẩm và dịch vụ của mình. Giả sử bạn là người đứng đầu một bộ phận trong một ngân hàng bán lẻ, và bộ phận này sắp ra mắt một sản phẩm thẻ tín dụng mới nên muốn kiểm tra xem khách hàng sẽ phản ứng ra sao trước sự điều chỉnh nhỏ về lãi suất và phí sử dụng thẻ thường niên. Giả sử bạn gửi lời mời dùng thử thẻ tín dụng mới này cho 5.000 khách hàng tiềm năng theo hai phiên bản khác nhau – một phiên bản thẻ có lãi suất 10,7% và phí thường niên 54 đô-la, và một phiên bản có lãi suất 13,7% nhưng không mất phí thường niên. Dựa trên thông tin phản hồi từ khách hàng, bạn có thể tối ưu hóa cách chào bán dòng thẻ mới này và cuối cùng đưa ra được sản phẩm khả dĩ cho tất cả các khách hàng mục tiêu (chính là phiên bản được ưa chuộng hơn trong bài kiểm định hai phiên bản vừa nêu). Các ngân hàng như Capital One đang sử dụng rộng rãi kỹ thuật này, và theo một báo cáo công bố vào năm 2017, ngân hàng này đã chạy 28.000 bài kiểm định riêng rẽ đối với các khách hàng hiện tại và tiềm năng của họ.

Các bài kiểm định hai phiên bản A/B rất phổ biến trong những công ty công nghệ và trong nhiều

trường hợp, người ta nhầm lẫn cho rằng kiểm định A/B chỉ áp dụng được trên các website, email hoặc trên thang đo tính ứng dụng của hệ thống. Nhiều công ty sử dụng các biến thể A/B trên website của mình để xác định phiên bản phù hợp, giúp họ đạt được các mục tiêu của mình.

Kiva.org là một tổ chức từ thiện muốn gia tăng số tiền tài trợ trên website của mình, nên họ quyết định sẽ bổ sung thêm nội dung vào trang chủ nhằm giúp khách truy cập trang cảm thấy yên tâm hơn. Cuộc kiểm định này được chạy trên hai phiên bản trang chủ khác nhau – trang hiện tại và trang mới được bổ sung thêm một hộp văn bản ở chân trang. Hộp văn bản có một số nội dung chính, chẳng hạn liệt kê các câu hỏi thường gặp, thông tin chứng thực về tổ chức, và các số liệu thống kê về hoạt động sử dụng website này. Phiên bản mới đã giúp Kiva gia tăng 11,5% tỉ lệ chuyển đổi thành nhà tài trợ.

Bạn có thể thực hiện kiểm định A/B đối với hầu như bất kỳ yếu tố nào trên website của bạn có tác động tới hành vi của khách truy cập.

1. Tiêu đề
2. Tiêu đề phụ
3. Văn bản trong đoạn văn

4. Thông tin chứng thực cho năng lực của công ty
5. Văn bản kêu gọi hành động
6. Nút kêu gọi hành động
7. Các đường dẫn
8. Hình ảnh
9. Nội dung gần phần sang trang mới
10. Bằng chứng về sự hòa đồng với xã hội
11. Những trường hợp được giới truyền thông nhắc đến
12. Các giải thưởng và huy hiệu đạt được

Ngoài ra, còn có các cuộc kiểm định cao cấp hơn, chẳng hạn kiểm định về cấu trúc định giá, chương trình khuyến mãi bán hàng, thời hạn dùng thử sản phẩm miễn phí, trải nghiệm giao diện người dùng, giao hàng miễn phí hay có phí...

Có thể thực hiện chạy các cuộc kiểm định A/B trên những nền tảng giao tiếp marketing khác như email. Một đặc điểm của phương pháp kiểm định A/B qua email là cho phép bạn gửi đi hai phiên bản email khác biệt nhau một chút, như vậy bạn thu thập được thông tin để tìm hiểu xem yếu tố nào khuyến khích khách hàng tham gia nhiều hơn, sau đó sử dụng

phiên bản email có hiệu quả nhất để gửi cho toàn bộ khách hàng.

Công ty tư vấn về tỉ lệ chuyển đổi khách hàng trực tuyến Conversion Vodoo đã thực hiện một cuộc kiểm định cho một khách hàng của họ, trong đó họ thay đổi tiêu đề email để xem người nhận phản ứng ra sao với nội dung của email. Họ đưa vào cuộc kiểm định này ba phiên bản email quảng bá với ba dòng tiêu đề email khác nhau, và kết quả được minh họa ở Hình 4.2.

Rõ ràng, các email có dòng tiêu đề "Chúc mừng Giáng sinh %TÊN KHÁCH HÀNG%" có tỉ lệ mở email và nhấp chuột vào đường dẫn trong nội dung email cao gấp đôi hai phiên bản còn lại. Vì vậy, giả sử bạn được giao nhiệm vụ gửi email thông báo chương trình khuyến mãi của công ty bạn trong dịp Giáng sinh cho 100.000 khách hàng. Dựa trên kết quả của cuộc kiểm định A/B, bạn biết nên đặt tiêu đề email như thế nào, khi đó bạn sẽ có cơ hội thu về nhiều câu trả lời hơn và gia tăng tỉ lệ chuyển đổi khách hàng của mình.

Ngoài những trường hợp trên, còn một số phương pháp phân tích thống kê đơn giản khác, bao gồm số trung bình, số trung vị, yếu vị, độ lệch chuẩn, và xác định số mẫu...

HÌNH 4.2 Sự thay đổi về tỉ lệ mở email và tỉ lệ nhấp chuột dựa trên sự thay đổi về tiêu đề email

Tiêu đề email	Gửi	Mở	Tỉ lệ mở	Nhấp chuột	Tỉ lệ nhấp chuột
Chúc mừng Giáng sinh %TÊN KHÁCH HÀNG%	33.500	5.432	16,21%	1.832	5,47%
Kỳ nghỉ vui vẻ %TÊN KHÁCH HÀNG%	33.500	3.724	11,12%	971	2,90%
Chúc mừng Giáng sinh & Kỳ nghỉ vui vẻ %TÊN KHÁCH HÀNG%	33.500	2.973	8,87%	943	2,81%

Số trung bình

Số trung bình là tổng các số trong một danh sách chia cho số lượng các số có trong danh sách đó. Số trung bình xác định xu hướng chung của một tập dữ liệu hoặc cung cấp một cái nhìn nhanh về dữ liệu của bạn. Một lợi thế khác của số trung bình là rất dễ tính toán. Trong một số tập dữ liệu, số trung bình còn có mối liên hệ mật thiết với số yếu vị và số trung vị (hai đơn vị đo lường khác gần số trung bình). Tuy nhiên, trong một tập dữ liệu có nhiều giá trị ngoại lệ hoặc có dải phân bổ lệch, thì số trung bình sẽ không mang lại thông tin chính xác mà bạn cần để đưa ra một quyết định cụ thể.

Điểm mạnh:

- Phát huy tác dụng tốt với các danh sách được kết hợp (bổ sung) theo cách đơn giản.

- Dễ tính toán: chỉ cần thực hiện phép cộng rồi sau đó thực hiện phép chia.
- Dễ hiểu – là con số "ở giữa", có giá trị lớn khi dãy số có giá trị lớn, và có giá trị nhỏ nếu dãy số có giá trị nhỏ.

Điểm yếu:

- Số trung bình có thể bị các giá trị ngoại lệ làm cho thiên lệch – nó không phát huy tác dụng với những số mẫu có độ biến thiên quá lớn. Ví dụ, số trung bình của ba số 100, 200, và -300 là 0 – đây là kết quả gây hiểu lầm.

Số trung vị

Trong một danh sách cho trước bao gồm các phần tử, số trung vị là phần tử xuất hiện ở giữa danh sách được lọc. Những dữ liệu như giá nhà đất luôn luôn được biểu thị bằng giá trị trung vị vì nó giúp bạn hình dung được điểm ở giữa trong tập dữ liệu tổng thể.

Điểm mạnh:

- Xử lý tốt các giá trị ngoại lệ – thường thì trung vị là đại diện chính xác nhất trong một nhóm.
- Chia dữ liệu thành hai nhóm, mỗi nhóm có một số lượng phần tử giống nhau.

Điểm yếu:

- Có thể khó tính ra số trung vị hơn: Bạn cần phải lọc danh sách trước.

Số yếu vị

Trong một danh sách cho trước, yếu vị là phần tử có tần suất xuất hiện thường xuyên nhất hoặc là phần tử lớn nhất. Yếu vị có thể dùng để đưa một đặc điểm cụ thể vào sản phẩm dựa trên ý kiến phản hồi của khách hàng. Bạn sẽ tính đến tất cả các thông tin phản hồi của khách hàng, nhưng sẽ hợp lý hơn nếu lựa chọn ra đặc điểm được họ nhắc đến nhiều nhất để đưa vào sản phẩm.

Điểm mạnh:

- Phát huy tác dụng tốt đối với các tình huống biểu quyết độc quyền (phương án này hoặc phương án khác; không thỏa hiệp).
- Đưa ra phương án lựa chọn được nhiều người mong muốn nhất (trong khi số trung bình có thể là phương án không ai mong muốn).
- Đơn giản, dễ hiểu.

Điểm yếu:

- Đòi hỏi nhiều nỗ lực để tính toán (phải cộng tổng các ý kiến biểu quyết).
- "Được ăn cả" – không có con đường ở giữa.

Độ lệch chuẩn

Độ lệch chuẩn, thường được biểu thị bằng chữ cái sigma trong tiếng Hy Lạp, là đại lượng thống kê dùng để đo độ phân tán của dữ liệu xung quanh giá trị trung bình. Độ lệch chuẩn cao cho thấy dữ liệu phân tán rộng khỏi giá trị trung bình, còn độ lệch chuẩn thấp nghĩa là dữ liệu phù hợp với giá trị trung bình. Trong danh sách các phương pháp phân tích dữ liệu, độ lệch chuẩn là một phương pháp hữu ích giúp bạn nhanh chóng xác định được độ phân tán của các điểm dữ liệu. Cũng giống như số trung bình, độ lệch chuẩn có thể sẽ cung cấp thông tin thiếu chính xác nếu chỉ nhìn vào nó. Chẳng hạn, nếu dữ liệu có một xu hướng rất lạ, ví dụ như một đường cong bất thường hoặc một số lượng lớn các giá trị ngoại lệ, khi đó độ lệch chuẩn sẽ không mang lại cho bạn tất cả những thông tin bạn cần.

Hãy giả sử có hai công ty nhỏ, mỗi công ty có bốn nhân viên (xem Hình 4.3). Trong một công ty, hai nhân viên kiếm được 19 đô-la/giờ, và hai người còn lại kiếm được 21 đô-la/giờ. Trong công ty thứ hai, hai nhân viên kiếm được 15 đô-la/giờ, một người kiếm được 24 đô-la/giờ, và người thứ tư kiếm được 26 đô-la/giờ.

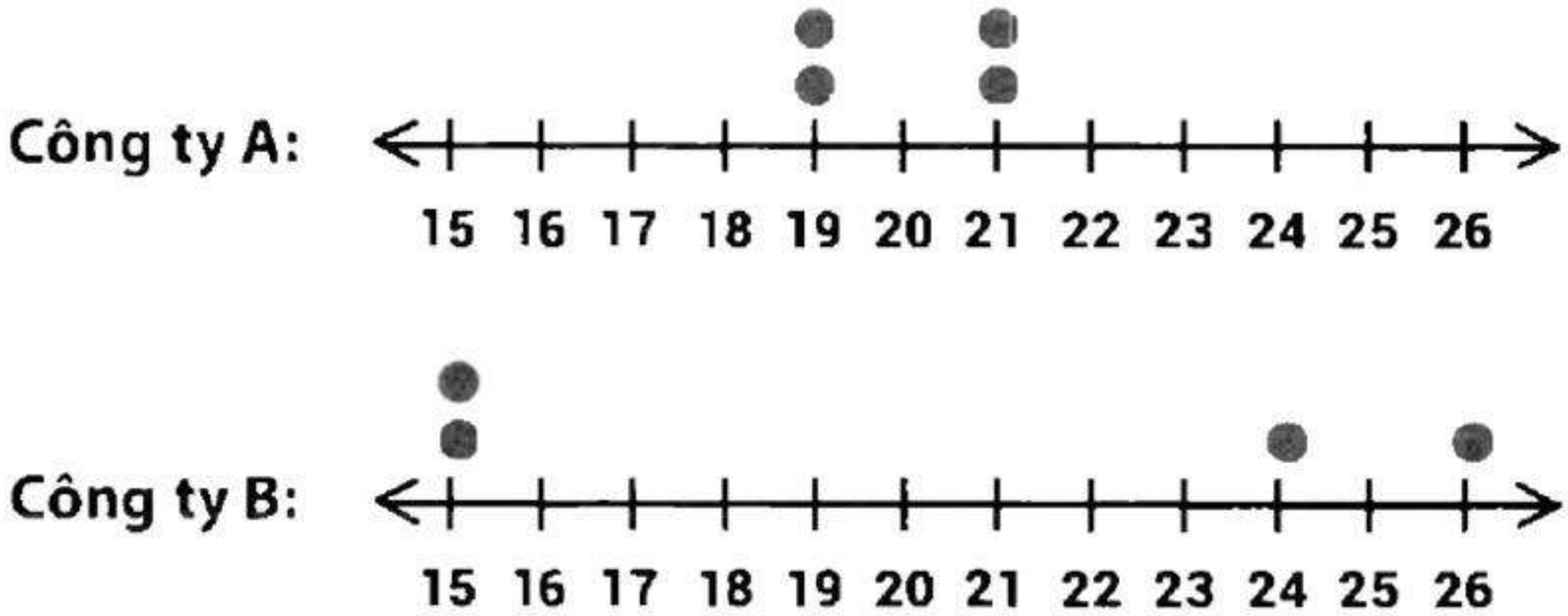
Ở cả hai công ty, mức lương trung bình đều là 20 đô-la/giờ, nhưng rõ ràng, mức độ phân tán mức lương tính theo giờ mỗi nơi mỗi khác. Ở công ty A,

mức lương của cả bốn nhân viên đều tập hợp chặt chẽ xung quanh giá trị trung bình, trong khi ở công ty B, có sự khác biệt lớn giữa hai nhân viên kiếm được 15 đô-la/giờ và hai nhân viên còn lại.

Xác định cỡ mẫu

Khi đo lường một tập dữ liệu lớn, hoặc một nhóm dân số, chẳng hạn một lực lượng lao động, không phải lúc nào bạn cũng cần thu thập thông tin từ mọi thành viên trong nhóm dân số đó, vì chỉ cần một số mẫu mang tính đại diện là đủ. Cái khó là làm sao xác định được đúng số mẫu để tính toán chính xác. Áp dụng các phương pháp tính toán tỉ lệ và độ lệch chuẩn, bạn sẽ xác định được đúng cỡ mẫu cần thiết để phần dữ liệu thu thập được có ý nghĩa về mặt thống kê. Khi nghiên cứu một biến mới, chưa qua kiểm định trong một nhóm dân số, các phương trình tỉ lệ có thể sẽ phải dựa vào một số giả định nhất định. Tuy nhiên, những giả định này có thể không hoàn toàn sai lầm. Khi đó, phần sai số sẽ được đưa vào trong quá trình xác định cỡ mẫu và tiếp đến là vào toàn bộ phần phân tích dữ liệu thống kê sau này.

HÌNH 4.3 Lập đồ thị thu nhập của nhân viên ở Công ty A và Công ty B



Phân tích mạng xã hội

Phân tích mạng xã hội, còn được viết tắt là SNA – Social Network Analysis, là một phương pháp rất phổ biến trong công nghệ dữ liệu lớn vì đây là phương pháp giúp bạn rút ra được những thông tin có ý nghĩa từ các dữ liệu phi cấu trúc. Với sự xuất hiện và phổ biến của các ứng dụng mạng xã hội như Facebook, Twitter, Instagram, Snapchat... các doanh nghiệp đang đứng trước nhiều cơ hội mới thú vị.

Phân tích mạng xã hội (SNA) chỉ hoạt động lập bản đồ và đo lường các mối quan hệ cũng như luồng chảy thông tin giữa con người, nhóm, tổ chức, máy tính, URL[1], và các thực thể thông tin/kiến thức có kết

[1]URL (Uniform Resource Locator – định vị tài nguyên thống nhất) là địa chỉ của một trang web trên mạng diện rộng toàn cầu World Wide Web, được dùng để tham chiếu tới tài nguyên trên mạng internet (ND).

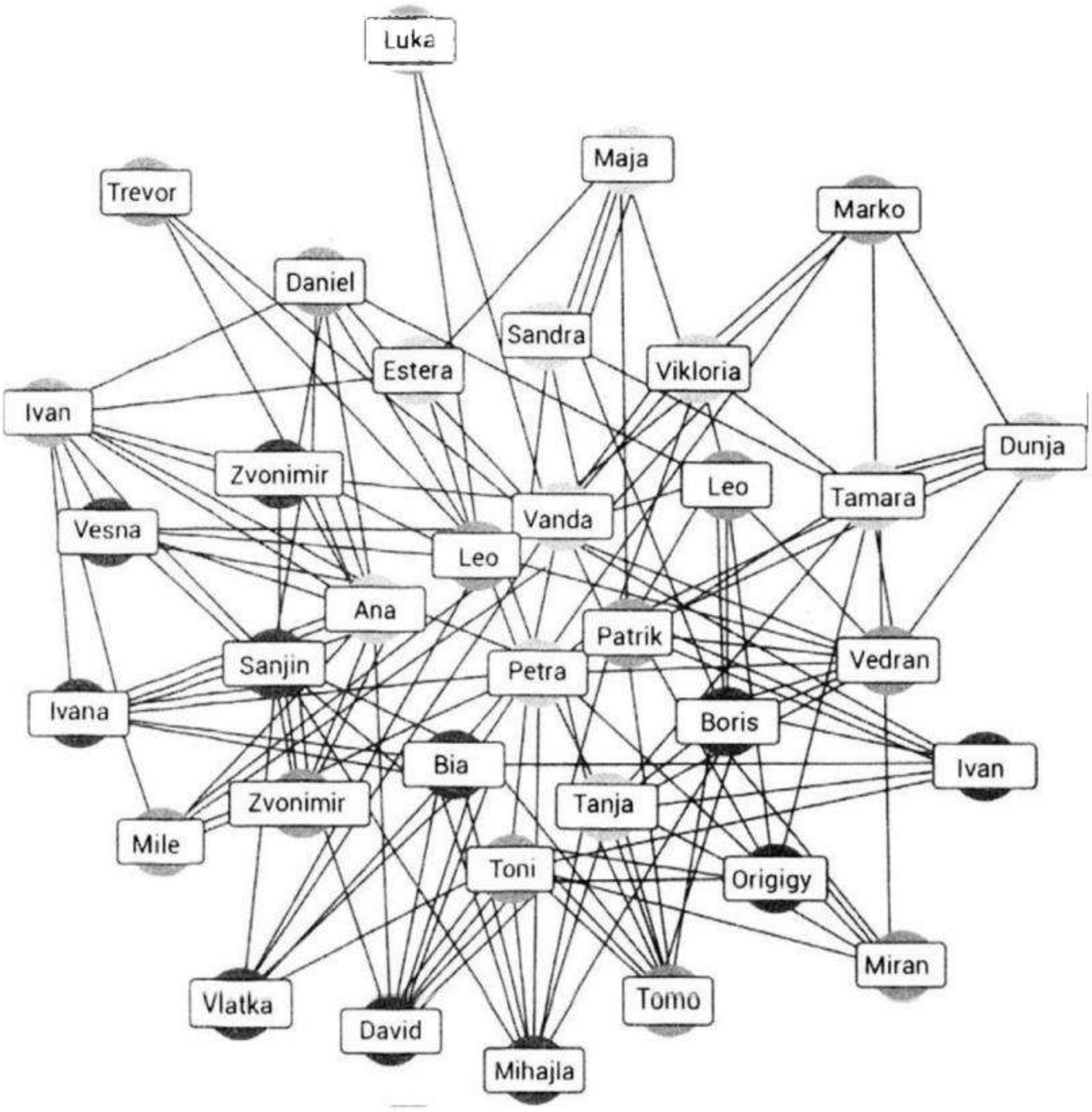
nối mạng khác. Các nút trong mạng lưới là con người và các nhóm, còn đường dẫn liên kết cho thấy mối quan hệ hay luồng chảy thông tin giữa các nút mạng. SNA mang đến cả phân tích trực quan lẫn phân tích toán học về mối quan hệ của con người. Trong lĩnh vực tư vấn quản trị, phương pháp này được gọi là Phân tích Mạng lưới Tổ chức (Organizational Network Analysis – ONA). ONA giúp bạn chụp X-quang cho tổ chức của mình và hiển thị hệ thần kinh quản lý, vốn là nơi làm nhiệm vụ kết nối tất cả mọi thứ trong tổ chức.

Phân tích mạng xã hội được sử dụng rộng rãi trong các lĩnh vực khoa học xã hội, khoa học hành vi, kinh tế học, marketing, và kỹ thuật công nghiệp. Góc nhìn của mạng xã hội tập trung vào mối quan hệ giữa các thực thể trong xã hội, và đây là một sự bổ sung quan trọng vào lĩnh vực nghiên cứu xã hội và nghiên cứu hành vi, vốn chủ yếu quan tâm đến những thuộc tính của các đơn vị xã hội.

Phân tích mạng xã hội là phân tích về mối quan hệ giữa hai hoặc nhiều hơn hai người/người dùng mạng xã hội, và những hoạt động họ tham gia vào. Xét từ góc độ SNA, mỗi người dùng có kết nối đều được gọi là một "nút" trong một mạng xã hội hoặc đồ thị minh họa mạng xã hội – và các nhà khoa học dữ liệu có thể thiết lập mối quan hệ có logic giữa các

nút này. Hình 4.4 mô tả một mạng lưới các nút trên Facebook – ở đây bạn có thể tìm thấy các thông tin như nút phổ biến, nút có ảnh hưởng, và nút/người dùng năng động... Hình 4.4 mô tả đồ thị về những người bạn chung.

HÌNH 4.4 Đồ thị về những người bạn chung trên một mạng xã hội



Loại thông tin này có thể rất hữu ích trong nhiều khía cạnh khác nhau của đời sống kinh doanh. Năm 2013, Qatar Telecom, một hãng viễn thông có trụ sở tại Trung Đông, thay đổi bộ nhận diện thương hiệu, và họ muốn kêu gọi sự tham gia của các khách hàng hiện tại, tranh thủ mạng lưới xã hội của họ để lan tỏa thông tin về thương hiệu mới của doanh nghiệp. Họ thuê một công ty phân tích mạng xã hội có trụ sở tại London để giúp họ xác định những người có ảnh hưởng lớn trên mạng xã hội trong thị trường của họ và quyết định tiếp cận những người này nhằm mời họ tham gia vào việc lan tỏa thông điệp của thương hiệu mới. Họ chuẩn bị một chiến lược trước khi ra mắt thương hiệu mới và những người có ảnh hưởng trên mạng xã hội làm việc cùng họ để quảng bá về thương hiệu mới trước khi công ty này đưa ra thông báo chính thức. Nhờ những kết quả phân tích tình hình thị trường trước thời điểm công bố và sự phối hợp chuẩn bị cho sự kiện ra mắt thương hiệu mới, công ty này đã hoàn thành việc chuyển đổi thương hiệu một cách thành công.

Vài năm trước, LinkedIn giới thiệu một đặc điểm rất hiệu quả gọi là "Những người bạn có thể biết" – trên thực tế đây là một công cụ gợi ý vận hành dựa trên đồ thị mạng xã hội của người dùng, các mạng lưới quan hệ trực tiếp và gián tiếp của anh ta, và những nơi anh ta đã theo học hoặc làm việc trong quá khứ...

Dựa tên tính năng SNA này, LinkedIn có thể đề xuất chính xác cho người dùng của họ những người mà họ có thể quen biết, để từ đó mạng này có thể thu thập thêm người dùng và gắn kết với người dùng hơn nữa.

Điều thú vị nằm ở chỗ, Facebook cũng triển khai một tính năng tương tự dựa trên kỹ thuật phân tích mạng xã hội. Họ bắt đầu xác định và gợi ý bạn bè cho người dùng dựa trên tính năng "Những người bạn có thể biết", nhưng khác với LinkedIn, Facebook lại sử dụng dữ liệu về điện thoại của người dùng, do đó kết nối sai người dùng với nhau khi dựa trên cơ sở rằng cả hai cùng có chung một số điện thoại của bất kỳ nhà mạng nào trong danh sách liên lạc của mình. Vì thế, nhiều bệnh nhân của một bác sĩ trị liệu có thể xuất hiện trong phần gợi ý kết bạn trên tài khoản Facebook của nhau. Chẳng hạn, nếu tìm kiếm cụm từ "những người bạn có thể biết bác sĩ trị liệu" trên Twitter, bạn sẽ thấy rằng rất nhiều người than phiền về cách gợi ý kết bạn sai của Facebook.

Phân tích hoạt động và nội dung

Một khía cạnh quan trọng khác của SNA là theo dõi dữ liệu về các hoạt động và nội dung do người dùng tạo ra trên mạng xã hội. Doanh nghiệp sử dụng thông tin về hoạt động và hành vi của người dùng trên mạng xã hội nhằm tìm hiểu rõ hơn về khách hàng của họ, đồng thời nhằm dự đoán các xu hướng mua sắm và

hoạt động của họ trong tương lai. Các quảng cáo của Facebook là một trong những ví dụ điển hình nhất. Nếu một người dùng chia sẻ nội dung nói về trẻ em mới sinh và có quan tâm đến các sản phẩm dành cho trẻ nhỏ, khi đó khả năng cao là họ vừa mới sinh con hoặc có quan hệ với ai đó sắp sinh con, và thông tin này có thể mang lại nhiều cơ hội tiềm năng cho các doanh nghiệp kinh doanh sản phẩm phục vụ những người mới làm cha mẹ. Gần đây Facebook bổ sung thêm một mục mới ở bộ lọc nhân khẩu học trong các module quảng cáo và đặt tên là "Sắp sinh con" – dĩ nhiên, không phải người dùng nào cũng cập nhật hồ sơ cá nhân về tình trạng sinh con của mình, nhưng Facebook sử dụng dữ liệu về các hoạt động và nội dung bài đăng của họ để đoán ra tình trạng của họ rồi tự động đưa họ vào nhóm Sắp có em bé (xem Hình 4.5).

Các doanh nghiệp không chỉ vận dụng những kiến thức về hành vi người dùng này để quảng bá và bán các sản phẩm hiện có, mà họ còn triển khai những kiến thức đã đúc rút được để phát minh hoặc đổi mới sản phẩm và các đặc điểm sản phẩm để cung cấp cho thị trường. Mạng xã hội đã mở ra một kênh mới mẻ để mọi người đều có thể giao tiếp công khai với thế giới, và điều này mở ra nhiều cơ hội lớn cho các doanh nghiệp.

HÌNH 4.5 Ví dụ về việc khoanh vùng mục tiêu quảng cáo trên Facebook



Các doanh nghiệp cũng phân tích những hoạt động bên ngoài mạng xã hội của người dùng và sau đó sử dụng kết quả phân tích để lập kế hoạch tiếp cận khách hàng trên mạng xã hội. Nếu bạn đang có kế hoạch đi nghỉ và muốn tìm mua một số sản phẩm phục vụ cho chuyến đi, chẳng hạn vé máy bay hay khách sạn... khi đó, sau khi bạn truy cập vào một số website liên quan,

có thể bạn cũng sẽ bắt đầu nhìn thấy nhiều nội dung quảng cáo tương tự xuất hiện trên màn hình của mình. Những quảng cáo này xuất hiện dựa trên hành động truy cập website của bạn, qua đó để lại một đoạn mã lập trình nhỏ trên máy tính – đoạn mã này được gọi là "cookie" – và sau đó các nhà cung cấp bắt đầu hiển thị cho bạn các nội dung quảng cáo theo thời gian thực trên các website như Facebook, vốn là một trong những cổng thông tin phổ biến nhất hiện nay.

Việc phân tích nội dung do các hoạt động sử dụng mạng xã hội tạo ra cũng mang lại rất nhiều lợi ích – đề tài này sẽ được trình bày chi tiết ở phần sau nói về bối cảnh ngữ nghĩa.

Phân tích ngữ nghĩa

Ngữ nghĩa là một nhánh của ngôn ngữ học và logic học liên quan đến ý nghĩa. Ngữ nghĩa có hai khía cạnh quan trọng là (1) ngữ nghĩa logic, liên quan đến các vấn đề như tri giác, tham chiếu, giả định, và ngụ ý, và (2) ngữ nghĩa từ vựng, liên quan đến việc phân tích ý nghĩa của từ ngữ được sử dụng và mối quan hệ giữa chúng. Trong cuốn sách này, chúng tôi chủ yếu nói đến ngữ nghĩa từ vựng.

Mạng xã hội và sự dân chủ hóa mạng internet đã cho ra đời nhiều nền tảng khác nhau để mọi người cùng có cơ hội tham gia vào các cuộc thảo luận hoặc

chủ động đơn phương bày tỏ chính kiến của mình – từ các diễn đàn mạng đến blog, Facebook, và Twitter, người dùng đang tích cực tham gia hoạt động và qua đó tạo ra những luồng dữ liệu lớn bằng văn bán hoặc các loại hình dữ liệu khác. Những diễn biến này đã thu hút nhiều sự quan tâm đối với các lĩnh vực như khai thác và phân tích văn bản để rút ra ý nghĩa và thông tin từ những lượng lớn văn bản mà không cần đọc chúng. Cụ thể, Real Time Intention (RTI – ý định theo thời gian thực) đã bắt đầu nhận được nhiều sự chú ý – có thể định nghĩa RTI là "một sự diễn đạt bằng văn bản cho thấy ý định thực hiện một hoạt động trong tương lai gần". Khi xác định được ý định đó, doanh nghiệp có thể cân nhắc giới thiệu sản phẩm, chương trình giảm giá, và các lời đề nghị bán hàng khác cho khách hàng mục tiêu. Chúng ta có rất nhiều ví dụ như vậy, chẳng hạn ứng dụng thẻ thành viên Clubcard của Tesco và công cụ gợi ý sản phẩm của Amazon – các công cụ này sẽ tìm hiểu và xác định ra các ý định của khách hàng dựa trên xu hướng mua sắm trong quá khứ của họ; không chỉ có thế, giờ đây, dựa trên những gì bạn viết trên mạng xã hội, chúng cũng có thể hiển thị cho bạn những đề nghị mua sản phẩm phù hợp với bạn.

Văn bản phi cấu trúc trực tuyến có thể được xử lý và xếp vào một trong số nhiều nhóm ý định của con người đã được xác định từ trước – sở thích, câu hỏi,

thắc mắc, phản đối, than phiền... Từ lâu các nhà nghiên cứu đã chỉ ra "cơ hội lớn đối với việc cá nhân hóa web" và "bối cảnh người dùng năng động" từ các kỹ thuật phân tích văn bản, song việc tìm hiểu RTI thậm chí còn mang lại nhiều lợi ích hơn nữa cho các doanh nghiệp.

Phương pháp này cũng giúp ích cho các ngân hàng và nhiều tổ chức tài chính khác, nơi hầu hết những nội dung trao đổi qua email giữa nhân viên và nhà cung cấp đều chứa những thông tin bí mật và nhạy cảm. Từ góc độ an ninh, nhiều tổ chức tài chính đã triển khai cài đặt một tầng phân tích ý định ở giữa người gửi và người nhận email nhằm bảo đảm rằng thông tin được an toàn và được mã hóa trước khi rời khỏi ranh giới của tổ chức.

Các tổ chức tài chính như HSBC cũng đã sử dụng RTI để thay thế một phần lớn đội ngũ nhân viên hỗ trợ bằng các chatbot[1] có khả năng trò chuyện tự động. Trên website của HSBC có hiển thị một ô trò chuyện để bạn gõ bất kỳ câu hỏi nào, và họ sẽ sử dụng công nghệ phân tích ý định để tìm ra câu trả lời đã soạn sẵn cho bạn. Bằng cách này, hầu hết người dùng sẽ nhận được lời giải đáp cho các thắc mắc của họ mà không

[1]Chatbot: Một chương trình máy tính được thiết kế để thực hiện các cuộc trao đổi với người dùng, đặc biệt là qua mạng internet. (ND)

cần gọi điện cho tổng đài, từ đó tiết kiệm được tiền cho phía ngân hàng. Nhiều tổ chức hiện cũng đang sử dụng tính năng trò chuyện dựa trên công nghệ phân tích ý định người dùng như vậy trên các cổng thông tin của họ.

Công nghệ phân tích ngữ nghĩa giúp doanh nghiệp rút được giá trị thực từ biển văn bản phi cấu trúc ồ ạt chảy về từ các nền tảng khác nhau, bao gồm email, blog, diễn đàn, mạng xã hội, và nhiều nền tảng khác. Việc tìm ra ý định và các thông tin chính khác từ những cuộc trao đổi trên mạng xã hội có thể giúp doanh nghiệp đổi mới và cải thiện các quy trình vận hành của mình, gia tăng doanh thu, và xây dựng các chiến lược kinh doanh.

Phân tích văn bản

Mỗi ngày, nhiều luồng dữ liệu phi cấu trúc khổng lồ lại được tạo ra từ nhiều nơi khác nhau như website, cổng thông tin mạng xã hội, và các file nhật ký. Phần lớn những dữ liệu này đều có định dạng văn bản phi cấu trúc, và các doanh nghiệp hiện đang sử dụng nhiều kỹ thuật vô cùng đa dạng, bao gồm kỹ thuật khớp nối xu hướng, kỹ thuật gom cụm, và kỹ thuật phân nhóm để tìm hiểu những dữ liệu văn bản khổng lồ này.

Phân tích văn bản nghĩa là rút ra được dữ liệu có cấu trúc có chất lượng cao từ dữ liệu văn bản phi

cấu trúc. Phân tích văn bản còn được gọi là khai thác văn bản. Một trong những lý do doanh nghiệp sử dụng kỹ thuật phân tích văn bản là để thu thập thêm dữ liệu về khách hàng từ các nguồn dữ liệu phi cấu trúc, qua đó làm phong phú thêm cho cơ sở khách hàng mà họ đang gây dựng, đồng thời rút ra được những kiến giải mới về khách hàng hoặc xác định quan điểm chung của khách hàng về các sản phẩm và dịch vụ. Phân tích văn bản có rất nhiều ứng dụng khác nhau, dưới đây là một số ứng dụng tiêu biểu:

- Quản lý hồ sơ: Từ bản đánh giá về các yêu cầu đòi bồi thường bảo hiểm, hồ sơ bệnh án, đến nội dung các bài phỏng vấn và báo cáo liên quan đến hoạt động điều tra tội phạm, có rất nhiều dữ liệu văn bản đã và đang được thu thập và tích lũy lại – và việc đọc qua tất cả những văn bản này khi phát sinh nhu cầu quả thực là một nhiệm vụ vô cùng gian nan. Phân tích văn bản có thể hỗ trợ bạn bằng cách tóm tắt và xác định những điểm quan trọng và những điểm chính.
- Phân tích đối thủ và nghiên cứu thị trường: Phân tích thị trường, khảo sát, và phân tích đối thủ cạnh tranh... không chỉ giới hạn ở những yếu tố được xác định từ trước có liên quan đến doanh nghiệp. Thông tin về thị

trường và ngành nghề tồn tại trong nhiều hình thức khác nhau ở nhiều nơi khác nhau, kể cả dưới dạng kỹ thuật số hay trên các tài liệu giấy. Nếu toàn bộ dữ liệu đều có thể được chuyển thành dữ liệu số và liên hệ với các tập dữ liệu có định dạng khác, khi đó việc phân tích văn bản kết hợp có thể trở nên vô cùng hữu ích vì nó sẽ hé lộ những kiến giải ẩn giấu bên dưới biển dữ liệu đó.

- Phân tích độ phủ sóng truyền thông: Hầu hết các công ty quan hệ công chúng đều sử dụng những quy trình khai thác và phân tích văn bản để tìm hiểu xem giới truyền thông nhắc đến hoặc đưa tin về khách hàng của họ như thế nào, để từ đó đưa ra các hành động phản ứng cần thiết. Google Alert là một công cụ như vậy, nó thường xuyên khai thác dữ liệu trên mạng và đưa ra kết quả về các trường hợp tên thương hiệu hoặc những từ khóa liên quan đến doanh nghiệp được nhắc đến trên các phương tiện truyền thông.

- Ý kiến khách hàng: Rút trích thực thể (tức kỹ thuật phân tích cú pháp và rút trích thực thể từ văn bản thô) là một phần quan trọng trong phân tích văn bản. Dưới đây là một số ví dụ về kỹ thuật rút trích thực thể từ văn bản phi cấu trúc thông thường:

- Tên công ty
- Ngày tháng và thời gian
- Tên sản phẩm...

Trong nhiều trường hợp, rút trích thực thể từ văn bản thô có thể được chuyển thành nhận dạng thực thể tự động, trong đó văn bản sẽ được tiến hành phân tích cú pháp và phân loại. Đây là một yêu cầu có thể thấy ở bất kỳ doanh nghiệp nào nhìn nhận nghiêm túc về dữ liệu phi cấu trúc và đã bắt tay vào thu thập cũng như xử lý các dữ liệu thô như thông tin trên mạng xã hội hay dữ liệu trong các file nhật ký của máy móc. Trong danh sách các ứng dụng nêu trên, việc nhận dạng ngày tháng, giờ, số lượng tiền... là các nhiệm vụ mà phần mềm phân tích văn bản có thể thực hiện ngay lập tức, không cần bạn phải hỗ trợ xác định những thực thể cần rút trích. Cách tiếp cận này mang lại một số lợi ích rõ ràng. Trong giai đoạn gần đây, với sự hậu thuẫn của các công cụ lưu trữ và phân tích dữ liệu lớn, thời gian cần thiết để thực hiện rút trích thực thể dữ liệu đã giảm xuống đáng kể, và có thể nhanh chóng gia tăng quy mô khối lượng văn bản cần thực hiện rút trích thực thể. Sau quá trình rút trích thực thể bằng kỹ thuật phân tích văn bản, kết quả thu về sẽ là một nhóm dữ liệu có cấu trúc, có thể kết hợp với dữ liệu đã có của doanh nghiệp để tiếp tục thực hiện các phân tích sâu thêm.

Nếu trong tài liệu văn bản xuất hiện các xu hướng đã biết như địa chỉ email hay số điện thoại, khi đó công cụ phân tích văn bản sẽ có thể dễ dàng xác định và phân loại nội dung của tài liệu. Cũng có thể sử dụng kỹ thuật phân tích văn bản để đo lường mật độ các từ dùng trong văn bản (thường được biểu diễn bằng đám mây chữ[1]) cũng như xác định mối quan hệ và sự liên hệ giữa các từ với nhau trong một tài liệu. Cũng có thể xác định các xu hướng và thực thể cụ thể liên quan đến doanh nghiệp để các công cụ phân tích văn bản có thể phân loại và rút trích từ ngữ và các câu rồi đưa chúng vào các nhóm đã xác định trước.

Thông thường, một công cụ phân tích văn bản trên mạng xã hội hoặc vận hành trên nền tảng web nói chung sẽ đọc và thu thập dữ liệu về nội dung các cuộc trao đổi trên mạng xã hội và trên web trong dạng thức văn bản, hình ảnh, video... thông qua các kỹ thuật như web scraping[2] hoặc truy cập vào các giao diện lập trình ứng dụng (API) về dữ liệu mạng xã hội như GNIP. Sau đó, công cụ này sẽ thực hiện phân loại dựa trên những dữ liệu và nguyên tắc ban đầu theo định nghĩa của ngành nghề hoặc lĩnh vực chuyên

---

[1] Đám mây chữ (word cloud): Một hình ảnh biểu diễn các từ được sử dụng trong một hoặc một vài đoạn văn bản điện tử. Kích cỡ các từ được thể hiện khác nhau tùy theo tần suất xuất hiện của chúng trong văn bản được chọn. (ND)

[2] Web scraping: Một quá trình tự động thu thập thông tin trên mạng. (ND)

môn đã chọn. Các thuật toán phân loại ban đầu sẽ sử dụng các tập dữ liệu huấn luyện[1] để nhận dạng và phân loại văn bản vào các nhóm, sau đó tiến hành tinh chỉnh các phân loại này dựa trên cơ sở nguồn dữ liệu mới được đưa vào và khối lượng lớn các dữ liệu mà chúng đang xử lý.

Một điều quan trọng cần lưu ý là phân tích ngữ nghĩa không chỉ liên quan đến việc nhận dạng và phân loại từ ngữ, mà bối cảnh trước khi diễn ra nội dung của dữ liệu cũng đóng vai trò hết sức quan trọng trong quá trình phân tích. Bối cảnh được xác định thông qua việc sử dụng các nguyên tắc phân loại khác nhau và các ontology[2] có kết nối mạng. Như vậy, một từ khi xuất hiện trong hai câu khác nhau sẽ hàm chứa những ý định và trạng thái cảm xúc khác nhau hoàn toàn theo cách phân loại của hệ thống, chẳng hạn:

- Chiếc ti-vi này thật *lớn* => hàm chứa cảm xúc Tích cực, bởi mọi người thường thích ti-vi có kích thước lớn.

---

[1] Dữ liệu huấn luyện (training data): Là tập dữ liệu ban đầu dùng để giúp một chương trình máy tính học cách xử lý thông tin theo các quy định cho trước. (ND)

[2] Ontology (bản thể học): Trong khoa học máy tính và khoa học thông tin, ontology là một phương thức thể hiện các thuộc tính của một nhóm đối tượng và mối quan hệ giữa chúng với nhau thông qua việc xác định ra một tập hợp các khái niệm và phân nhóm đại diện cho đối tượng đó. (ND)

- Chiếc điện thoại di động này thật *lớn* => hàm chứa cảm xúc Tiêu cực, bởi mọi người thường không thích điện thoại di động có kích thước lớn.

Như vậy, khi từ khóa "lớn" xuất hiện trong nhóm Ti-vi, nó sẽ được phân loại là Tích cực, và khi xuất hiện trong nhóm Điện thoại di động, nó sẽ được phân loại là Tiêu cực – cách phân loại này dựa trên tập dữ liệu huấn luyện đã được xác định trước, và quá trình phân loại tự động diễn ra theo hướng dẫn đó.

Việc xác định bối cảnh không chỉ dựa trên các cuộc trao đổi liên quan và cuộc trao đổi diễn ra ngay trước đó; nó còn có thể được xác định thông qua việc sử dụng các tín hiệu khác như thông tin về hồ sơ cá nhân của đối tượng cần phân tích, vị trí của người đó, thiết bị và thời gian diễn ra cuộc trao đổi. Có thể xác định một nguyên tắc phân loại riêng cho các tín hiệu phân loại này.

Phân tích cảm xúc

Phân tích cảm xúc là quá trình xác định xem liệu ý nghĩa của câu được viết bằng một ngôn ngữ nhất định có thể được coi là tích cực, tiêu cực, hay trung lập. Phân tích cảm xúc còn được gọi là khai thác quan điểm, nghĩa là tìm hiểu quan điểm hoặc thái độ của

người nói. Một ứng dụng phổ biến của công nghệ này là nhằm tìm hiểu xem người khác nghĩ gì về một đề tài cụ thể.

Công cụ phân tích sẽ phân tích cú pháp câu và các cụm từ được sử dụng dựa trên các nguyên tắc phân loại và các ontology đã được xác định từ trước, qua đó xác định bối cảnh và ý nghĩa của câu cần phân tích. Việc hiểu được bối cảnh của cuộc trao đổi là rất quan trọng để có thể hiểu, phân loại, và hành động dựa trên ý nghĩa của câu nói được trình bày trong một đoạn văn bản đã cho và liên hệ các trạng thái cảm xúc với đoạn văn bản đó.

Mỗi ngày có hàng triệu cuộc trao đổi diễn ra sôi động trên mạng xã hội, và các doanh nghiệp phải biết người tiêu dùng đang nghĩ gì về mình, để qua đó họ có thể cải thiện các sản phẩm/dịch vụ, giải quyết các vấn đề người tiêu dùng gặp phải, và giành được sự tín nhiệm trên thị trường. Phân tích cảm xúc xác định các cảm xúc được liên hệ đến trong một cuộc trao đổi, từ đó cho phép các thương hiệu rút ra được những hiểu biết sâu sắc hơn.

Giả sử bạn muốn biết liệu mọi người trên Twitter cho rằng đồ ăn Trung Quốc ở San Francisco là ngon hay dở. Phân tích cảm xúc trong các cuộc trao đổi trên Twitter có thể sẽ giúp bạn trả lời câu hỏi này. Bạn thậm

chỉ còn có thể biết vì sao người ta lại cho rằng đồ ăn Trung Quốc là ngon hay dở, bằng cách rút trích những từ ngữ chính xác được dùng để cho thấy vì sao người ta lại thích hoặc không thích chúng. Chẳng hạn, nếu nhiều cuộc trao đổi cùng nhắc đến việc đồ ăn Trung Quốc quá mặn, bạn sẽ ngay lập tức hình dung rõ hơn về lý do người tiêu dùng không hài lòng.

Phân tích cảm xúc cũng có thể được sử dụng cho mục đích tách riêng, phân loại, và tóm tắt các bình luận trực tuyến về sản phẩm/dịch vụ của doanh nghiệp, đồng thời xác định được đâu là những thông tin bình luận giả mạo trên các website. Với phân tích cảm xúc, tất cả các thông tin phản hồi đều có thể được tóm tắt lại thành một bản báo cáo ngắn gọn, dễ đọc mà khách hàng có thể tham khảo trước khi ra quyết định mua sắm. Còn có một thách thức nữa: Danh tiếng và công việc kinh doanh của doanh nghiệp có thể sẽ chịu ảnh hưởng tiêu cực nếu trên mạng internet xuất hiện các thông tin bình luận và đánh giá giả mạo về họ – một công ty đối thủ hoặc một cá nhân có ý định trêu đùa hoàn toàn có thể thực hiện điều này. Tuy nhiên bây giờ, doanh nghiệp đã có thể dùng công nghệ phân tích cảm xúc để xác định ra những bình luận giả mạo đó để yêu cầu gỡ bỏ chúng và bảo vệ vị thế của mình trên thị trường.

Tình báo marketing và bán hàng

Nhiều doanh nghiệp và công ty nghiên cứu thị trường hiện nay vẫn sử dụng các công cụ, kỹ thuật, và quy trình của mười năm trước để thực hiện thu thập và phân tích dữ liệu. Các cuộc phỏng vấn nhóm, khảo sát, và dự án theo dõi thụ động – tất cả đều phải thay đổi trước sự xuất hiện của các phát minh công nghệ mới và những thay đổi về hành vi của người tiêu dùng ngày nay. Các phương pháp như khảo sát nên được rút ngắn lại và cần thiết kế sao cho có thể dễ dàng thực hiện trên thiết bị di động...; đây là một vài vấn đề về mặt chiến thuật rốt cuộc cũng đang có sự thay đổi. Nhưng ngoài ra, trong bản thân nội bộ các doanh nghiệp cũng đang chứa nhiều khối lượng lớn dữ liệu giao dịch, và những dữ liệu này lại hàm chứa những thông tin vô cùng giá trị về hành vi của khách hàng. Đây là điều đặc biệt quan trọng giúp doanh nghiệp hiểu được về thị trường cũng như về những thay đổi đang diễn ra trong hệ sinh thái mà họ hoạt động. Câu trả lời cho nhiều câu hỏi mà các hoạt động nghiên cứu thị trường đang tìm hiểu có thể đã tồn tại ngay trong chính kho dữ liệu nội bộ đó, và điều quan trọng nữa là doanh nghiệp hoàn toàn có khả năng phân tích và sử dụng kiến thức từ kho dữ liệu đó.

Doanh nghiệp có thể sử dụng không chỉ các kết quả nghiên cứu thị trường mà cả những dữ liệu phi cấu trúc có sẵn trên mạng xã hội và các cuộc trao đổi

khác diễn ra trên các website, sau đó đưa luồng dữ liệu đó qua công cụ phân tích nhằm xác định nhóm khách hàng mục tiêu tiềm năng (cũng là các đầu mối liên hệ cho đội ngũ bán hàng) cho các sản phẩm và dịch vụ của mình. Hiện nay đã có nhiều lĩnh vực ngành nghề, từ phát triển web đến kinh doanh phần mềm, đã bắt đầu áp dụng các cách tiếp cận thiên về phân tích như vậy để tạo ra các đầu mối liên hệ cho đội ngũ bán hàng. Mọi người nói về nhiều sản phẩm và dịch vụ trên mạng xã hội – những cuộc trao đổi này có thể được phân loại theo thời gian thực để sau đó giới thiệu cho những người đang có ý định mua một sản phẩm hoặc dịch vụ cụ thể giúp họ có được đầy đủ thông tin cần thiết để ra quyết định. Cũng giống như một số ứng dụng khác, các công cụ phân tích dữ liệu lớn cũng hỗ trợ vận hành các cổng thông tin bình luận công khai về sản phẩm/dịch vụ, trong đó những dữ liệu về thông tin bình luận về sản phẩm/dịch vụ đến từ nguồn công cộng và kỹ thuật phân tích văn bản sẽ tự động hình thành quan điểm của thị trường. Đây là một cách sử dụng các thông tin bình luận về sản phẩm rất minh bạch và thông minh cho người tiêu dùng ở bất kỳ lĩnh vực ngành nghề nào.

Có nhiều ứng dụng được xây dựng dựa trên kỹ thuật phân tích và phân loại dữ liệu lớn nhằm khám phá những thông tin giá trị ẩn giấu trong dòng chảy

không ngừng của những cuộc trao đổi trên mạng xã hội, và cũng có nhiều ứng dụng được xây dựng dựa trên kỹ thuật phân tích văn bản phi cấu trúc để thực hiện các phân tích về đối thủ cạnh tranh, nghiên cứu về chiến lược tiếp cận thị trường, phân tích về tình hình cung-cầu...

Trực quan hóa dữ liệu

Một bức ảnh đáng giá cả nghìn lời nói

Trực quan hóa dữ liệu nghĩa là trình bày dữ liệu ở dạng thức đồ họa. Điều này khiến quá trình ra quyết định dễ dàng hơn, bởi bạn có thể nhìn thấy các kết quả phân tích/phát hiện được hiển thị một cách sống động và trực quan ngay trước mắt mình. Thay vì phải vật lộn với các con số, với kỹ thuật trực quan hóa dữ liệu, giờ đây bạn có thể tập trung vào những thông tin rõ ràng, nhanh chóng nắm được những khái niệm khó hoặc xác định được những xu hướng mới. Bây giờ, người ta cũng có thể thực hiện trực quan hóa dữ liệu tương tác, theo đó các nhà quản lý có thể tìm hiểu sâu về từng biểu đồ và đồ thị để hiểu các thông tin chi tiết, đồng thời thực hiện những thay đổi nhằm hiểu rõ về các xu hướng trong dữ liệu.

Trực quan hóa dữ liệu là một phương pháp hết sức hiệu quả để tìm hiểu và sử dụng kỹ thuật phân tích dữ liệu – và bạn có thể triển khai phân tích dữ

liệu ở bất kỳ khía cạnh nào trong doanh nghiệp của mình, từ việc cải thiện hoạt động bán hàng ở phía trên đến hiệu quả vận hành ở phía dưới.

Các nhà quản lý đã sử dụng báng biểu dựng trên Excel suốt một thời gian dài. Nhưng với sự ra đời của các loại bảng biểu đầy tính sáng tạo của các công ty như Qlik và Tableau, việc trực quan hóa dữ liệu đã được nâng lên một cấp độ mới. Phương pháp kể chuyện với sự hỗ trợ của các thiết kế sáng tạo và kỹ thuật trực quan hóa dữ liệu còn được gọi là Infographics (đồ họa thông tin) – phương pháp này được nhiều doanh nghiệp phục vụ người tiêu dùng sử dụng để trình bày thông tin và chia sẻ rộng rãi với mọi người. Có nhiều loại đồ thị khác nhau có thể giúp bạn trình bày những thông tin ẩn đằng sau các tập dữ liệu phức tạp một cách hết sức hiệu quả.

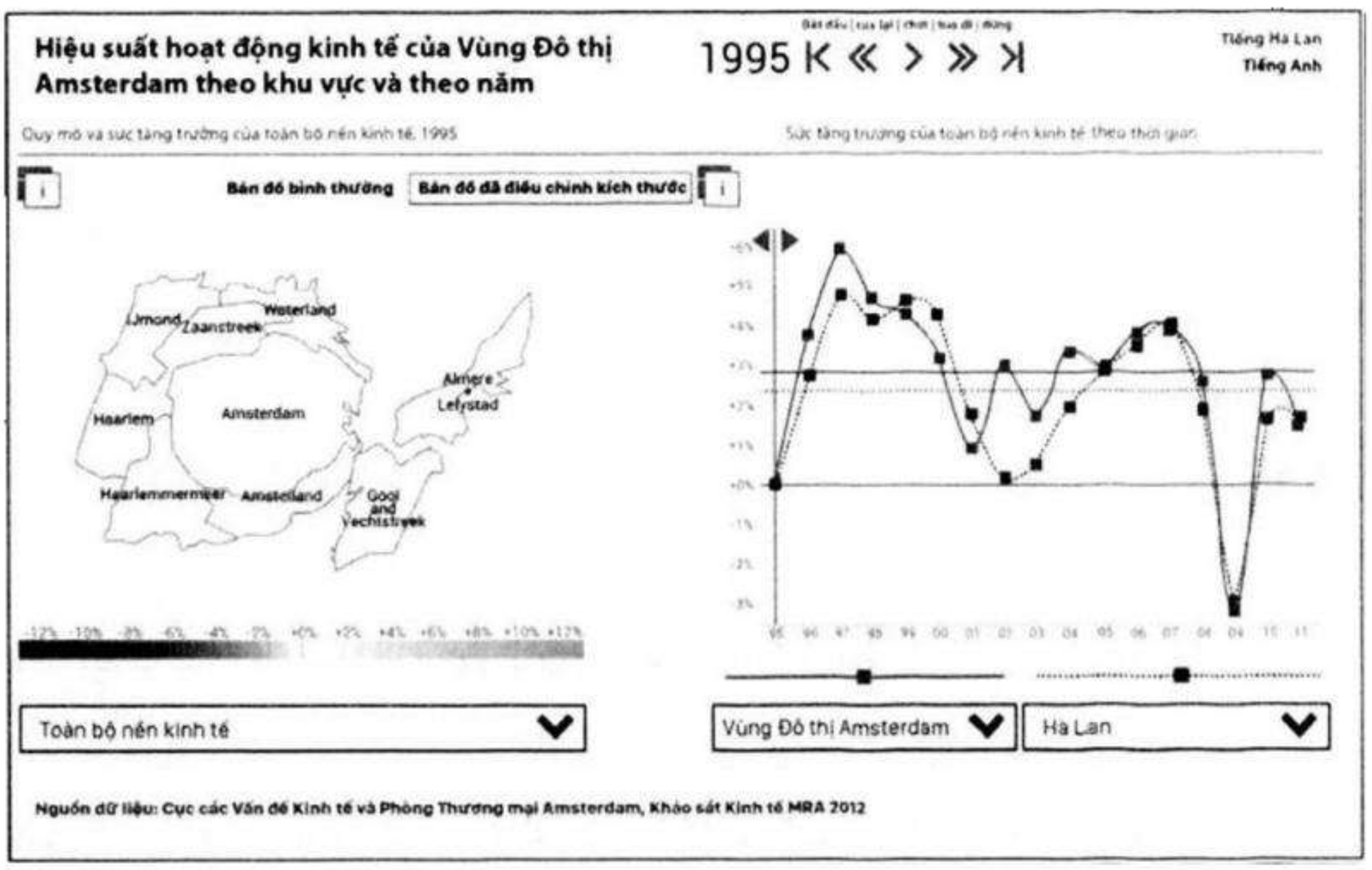
Trong hầu hết các doanh nghiệp, trực quan hóa dữ liệu có thể mang lại rất nhiều giá trị theo nhiều cách khác nhau, dưới đây là một số khía cạnh chính của kỹ thuật này mà phần lớn các doanh nghiệp đều có thể sử dụng:

- Việc trực quan hóa dữ liệu rõ ràng có thể giúp doanh nghiệp dễ dàng xác định các khía cạnh cần cải thiện dựa trên hiệu suất hoạt động và lợi nhuận kỳ vọng của họ. Kỹ thuật này không những có thể cải thiện hoặc hướng sự chú ý

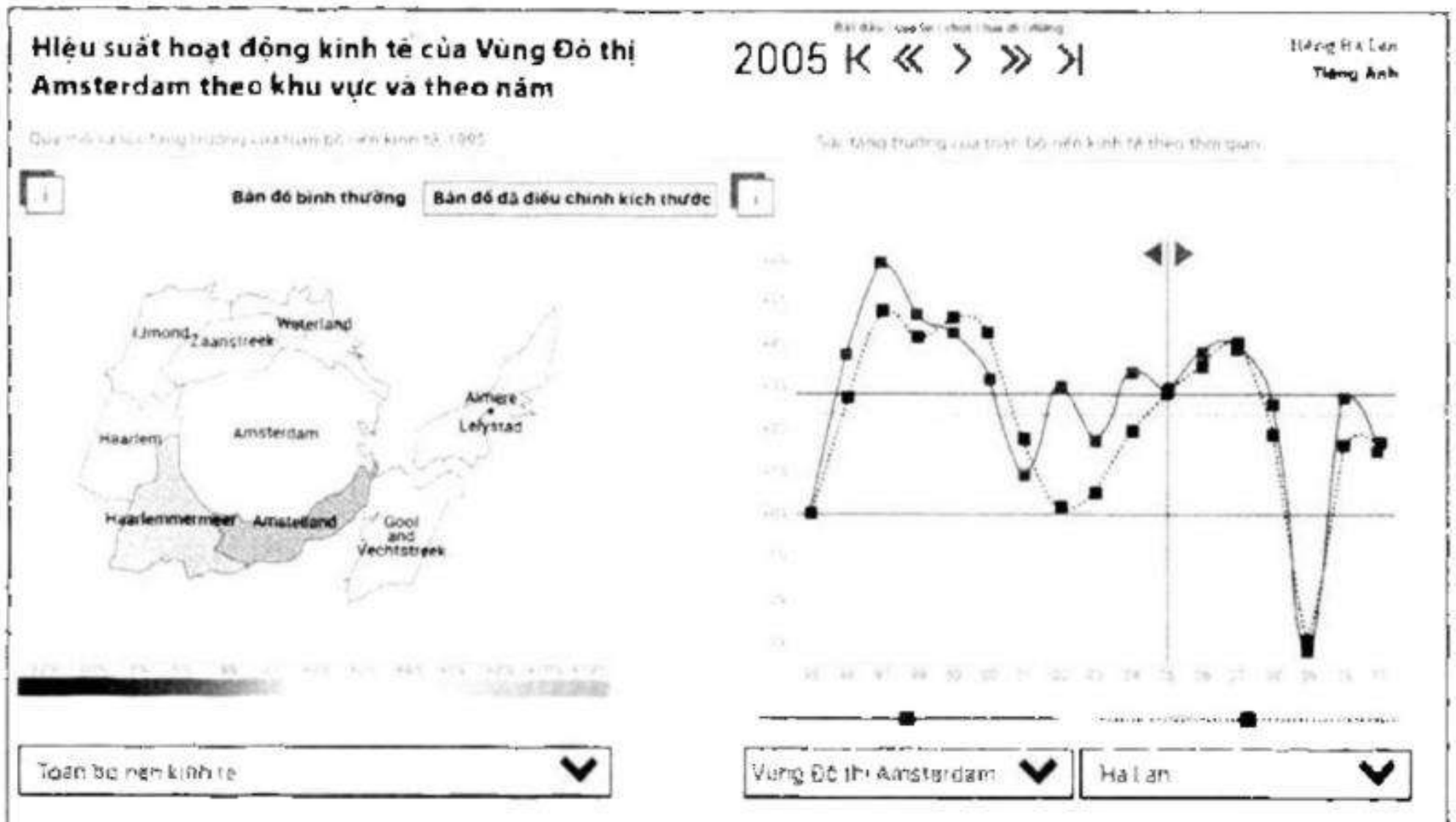
đến những khía cạnh hoạt động kém, mà còn giúp doanh nghiệp xác định được những khía cạnh hoạt động tốt hơn, để có thể triển khai thêm nhiều nguồn lực vào đó nhằm đạt được những kết quả khả quan hơn nữa.

Cũng có thể trực quan hóa hiệu suất hoạt động theo dòng thời gian để làm nghiên cứu so sánh. Hãy xem hình 4.6-4.8, trong đó hiệu suất hoạt động kinh tế của Vùng đô thị Amsterdam được mô tả theo khu vực và theo năm. Bạn có thể thấy các đồ thị và thay đổi trong ba năm – 1995, 2005, và 2009.

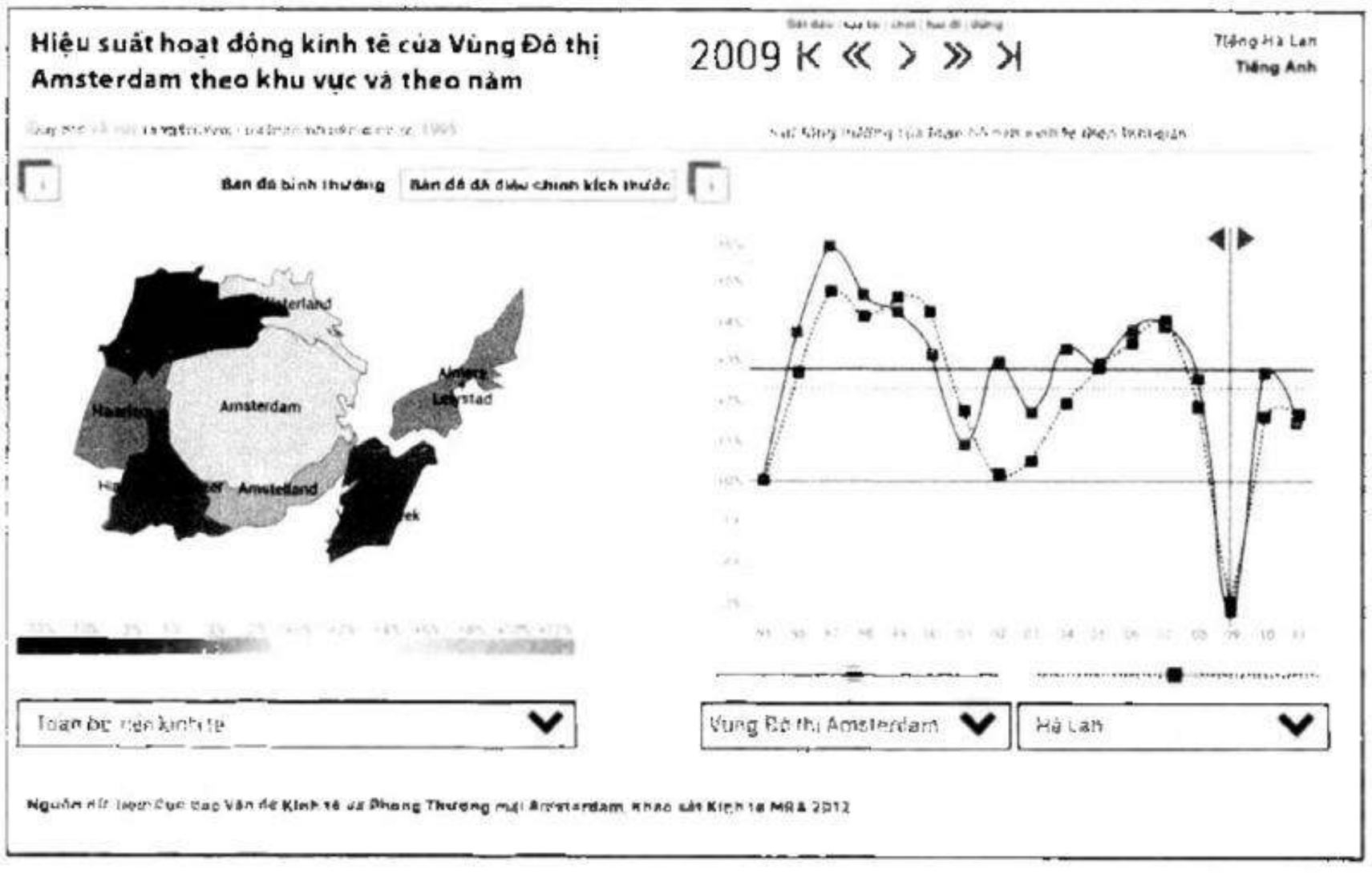
HÌNH 4.6 Hiệu suất hoạt động kinh tế của Vùng đô thị Amsterdam năm 1995



HÌNH 4.7 Hiệu suất hoạt động kinh tế của Vùng đô thị Amsterdam, 2005

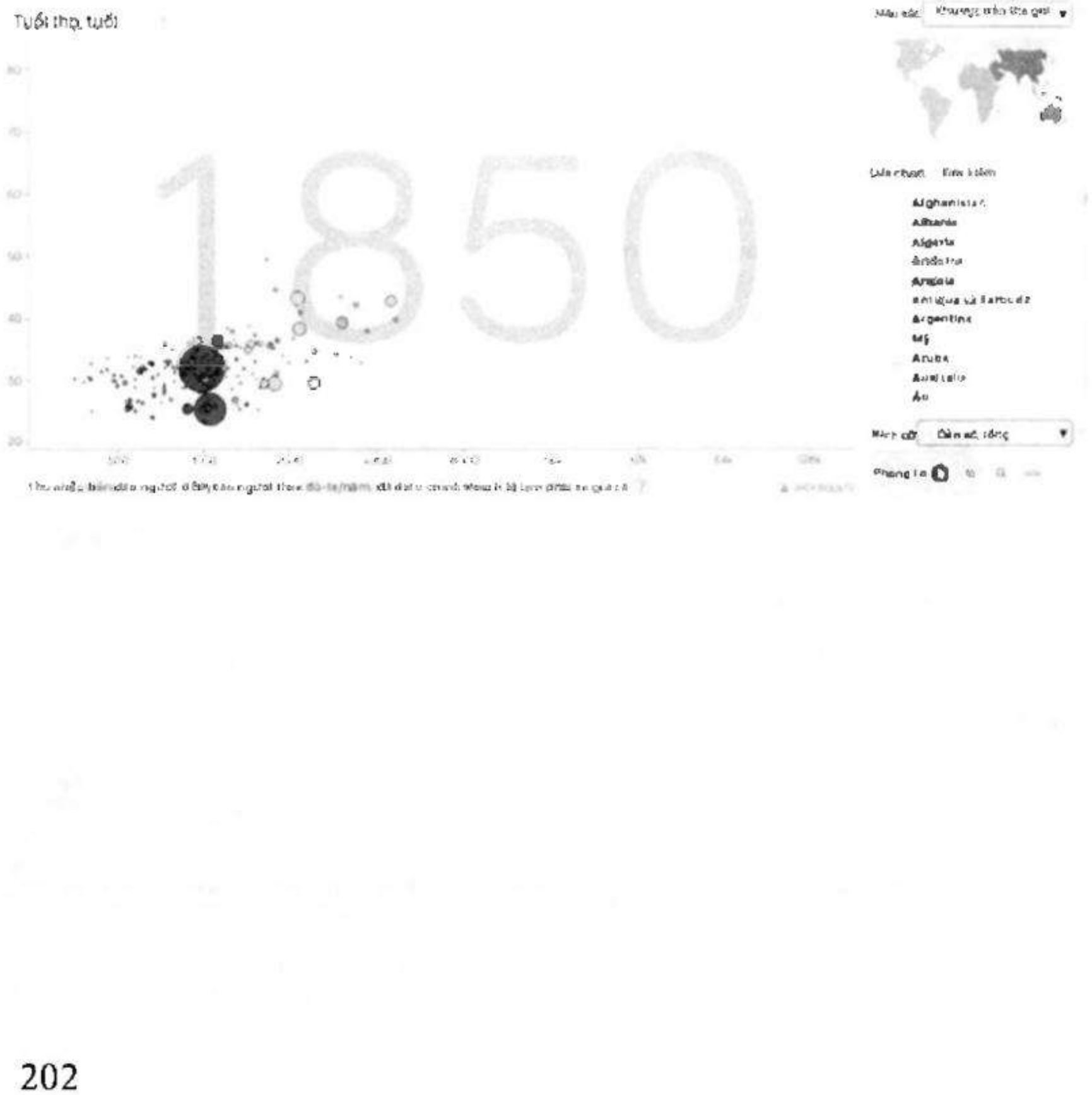


HÌNH 4.8 Hiệu suất hoạt động kinh tế của Vùng đô thị Amsterdam, 2009



- Với sự hỗ trợ của kỹ thuật trực quan hóa dữ liệu năng động, có thể xác định được các yếu tố tác động đến hành vi và xu hướng mua sắm của khách hàng. Cũng có thể thực hiện điều này với kỹ thuật phân tích hồi quy mà chúng ta đã bàn đến ở phần trước của chương này, nhưng với sự hỗ trợ của kỹ thuật trực quan hóa năng động, nhà quản lý có thể chỉ cần kéo giá trị của một yếu tố để xem sự thay đổi ở yếu tố khác theo thời gian thực.

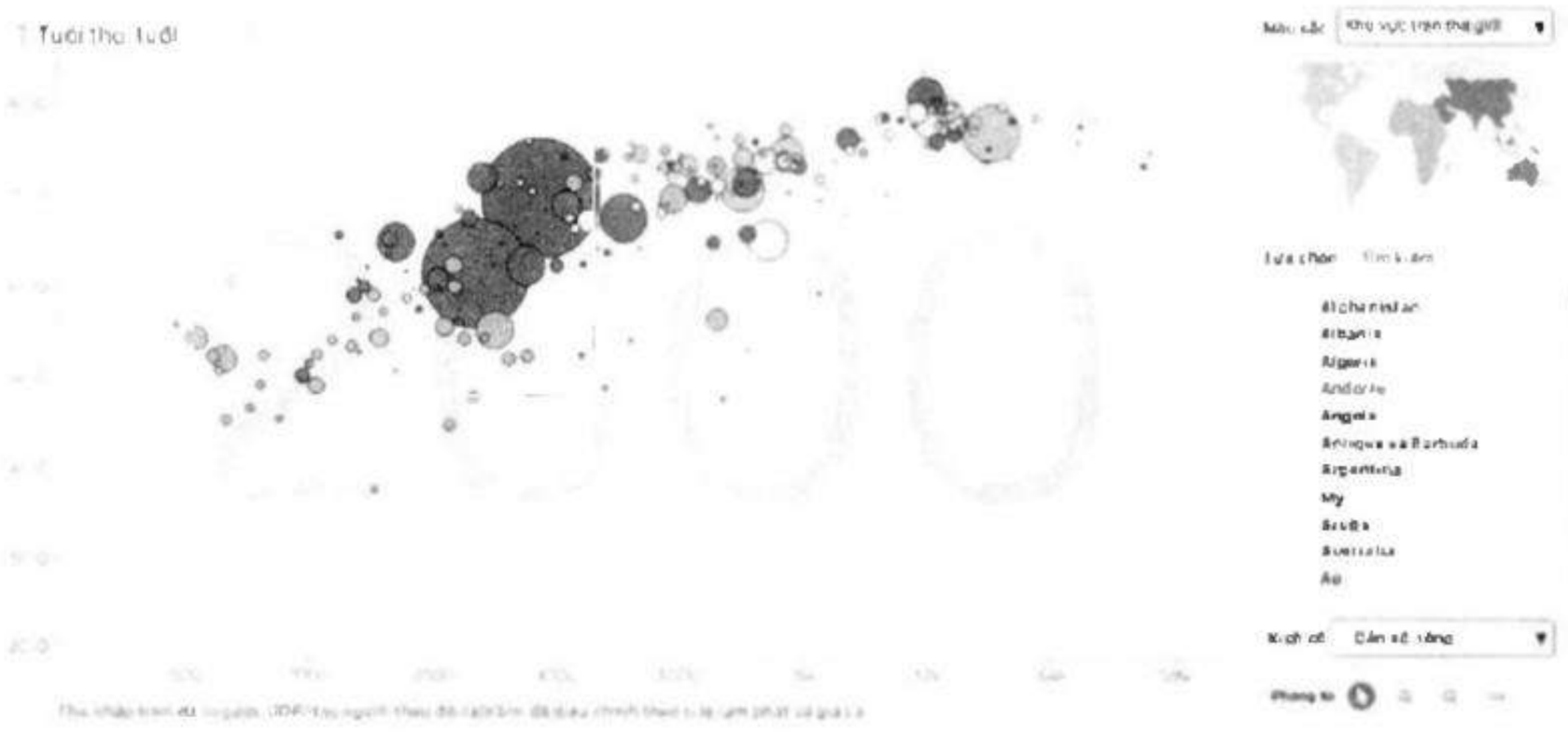
HÌNH 4.9 Tuổi thọ ở các quốc gia năm 1850



HÌNH 4.10 Tuổi thọ ở các quốc gia năm 1950



HÌNH 4.11 Tuổi thọ ở các quốc gia năm 2000



Hãy xem các hình từ 4.9 đến 4.11, trong đó sự tthay đổi về mức thu nhập trên đầu người được phản ánh trong tuổi thọ của người dân.

- Doanh số dự đoán là một ứng dụng thú vị khác của trực quan hóa dữ liệu, bởi nó có thể hiển thị một cái nhìn toàn thể về doanh số bán hàng và tất cả các yếu tố ảnh hưởng đến doanh số trên cùng một màn hình. Nhà quản lý có

thể ra quyết định nhanh chóng và thông minh khi thấy được bao quát vấn đề trên một màn hình duy nhất – xem Hình 4.12.

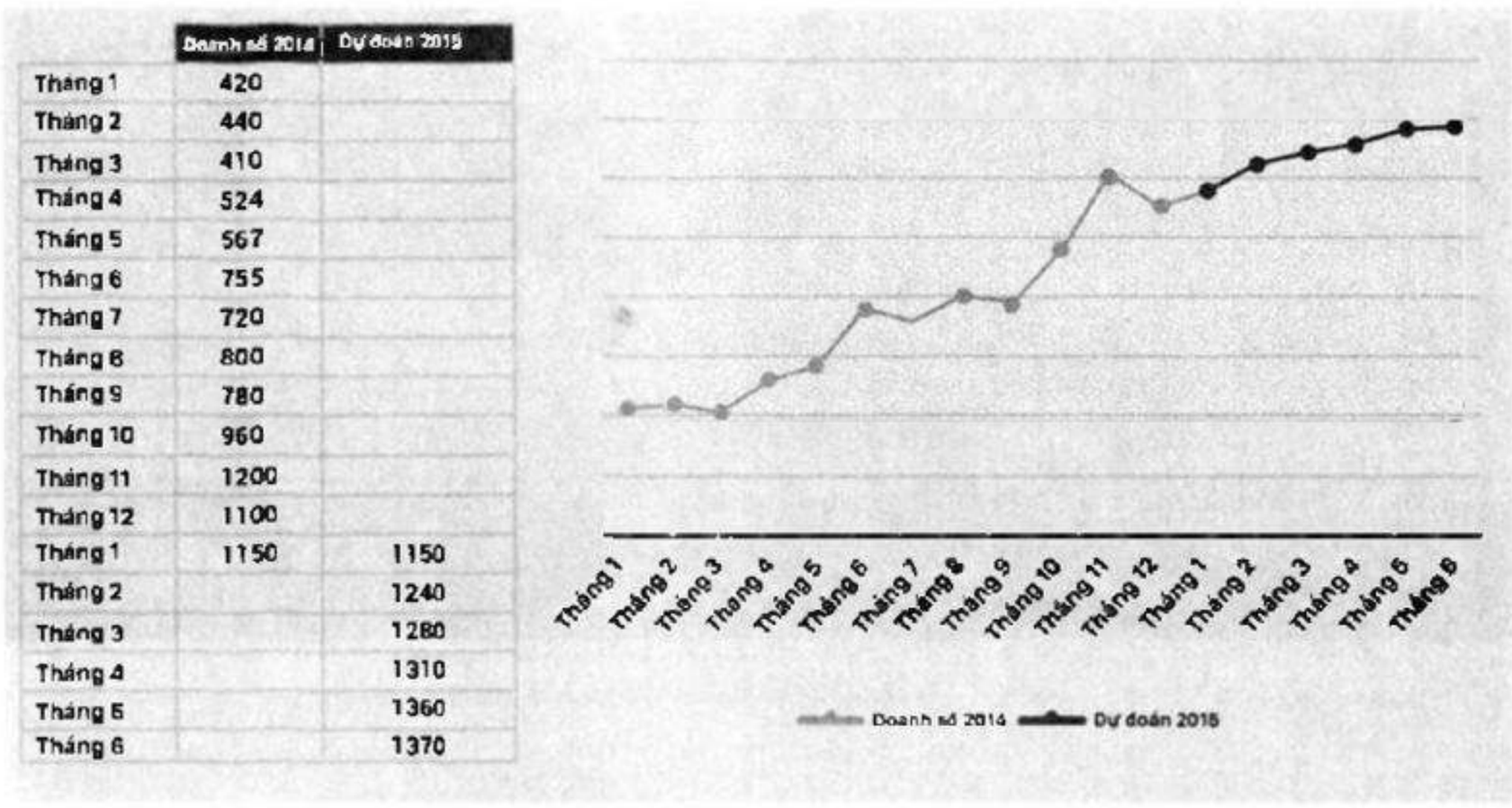
Tờ *The Guardian* ở Anh nổi tiếng với biệt tài biểu diễn các dữ liệu và thông tin phức tạp thành những infographics rất dễ hiểu với các tập dữ liệu có sẵn công khai. Chẳng hạn, Hình 4.13 là cái nhìn tổng thể thể hiện mức chi tiêu và thu nhập của chính phủ Anh năm 2014 do *The Guardian* thực hiện.

Hình 4.14 là một hình ảnh trực quan khác, thể hiện tình hình di cư thuần túy trong năm năm qua.

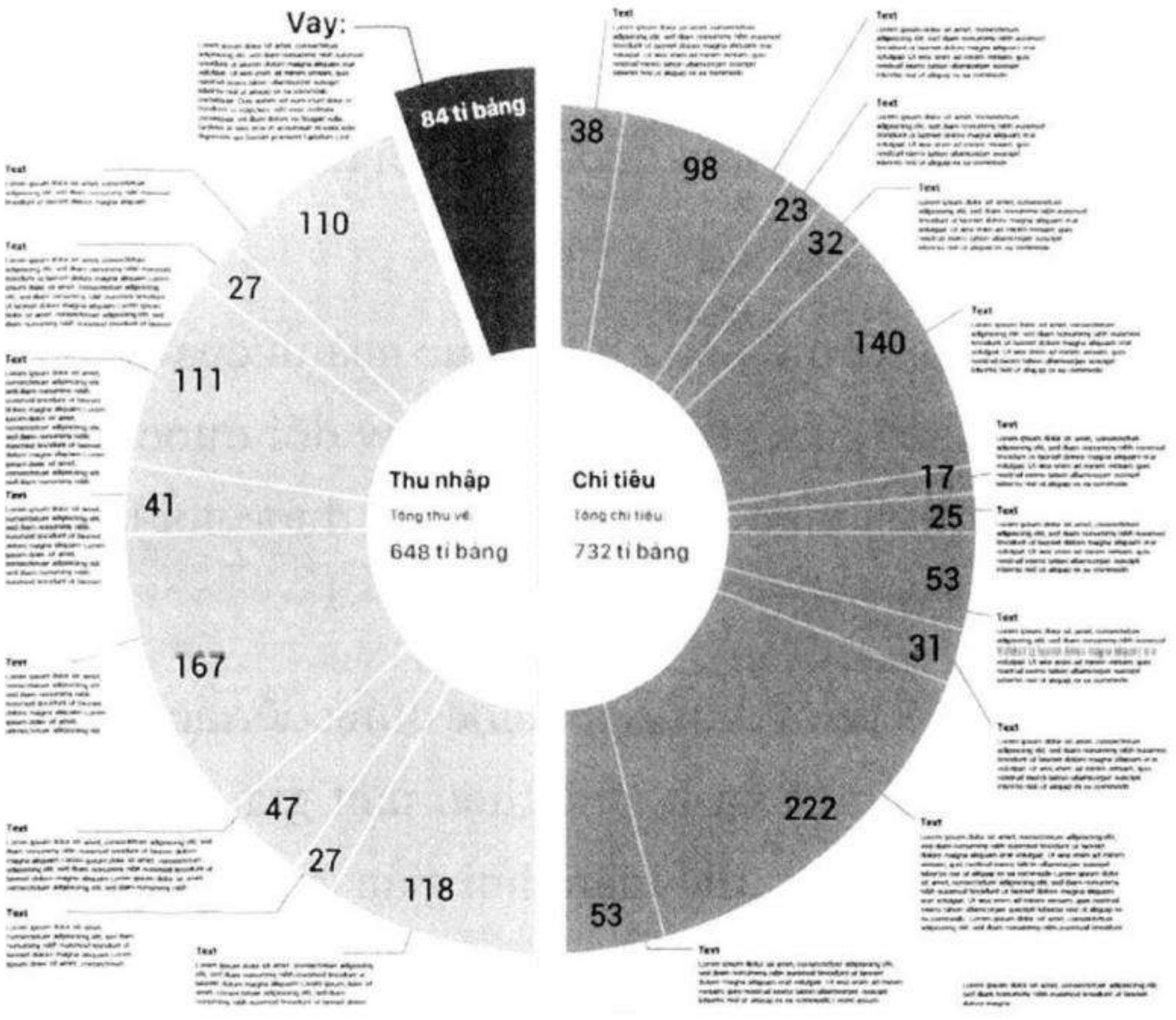
Có nhiều phương án trực quan hóa dữ liệu, nhưng nguyên tắc cơ bản trong quá trình tìm kiếm và lựa chọn có thể được tóm tắt thành một "phương châm tìm kiếm thông tin trực quan" như sau: tổng quan trước, phóng to và lọc, sau đó đến chi tiết theo yêu cầu.

Hiện nay có nhiều công cụ trực quan hóa rất hiệu quả, và thực ra nhiều công cụ trong đó đã được đưa vào các phần mềm trực quan hóa phổ biến như Tableau hay QlikView. Ở đây chúng tôi xin giới thiệu nhanh một danh sách nhỏ các biểu đồ trực quan hóa cơ bản và phổ biến, được sử dụng theo nhiều hình thức khác nhau nhằm trình bày thông tin rút ra được từ các tập dữ liệu.

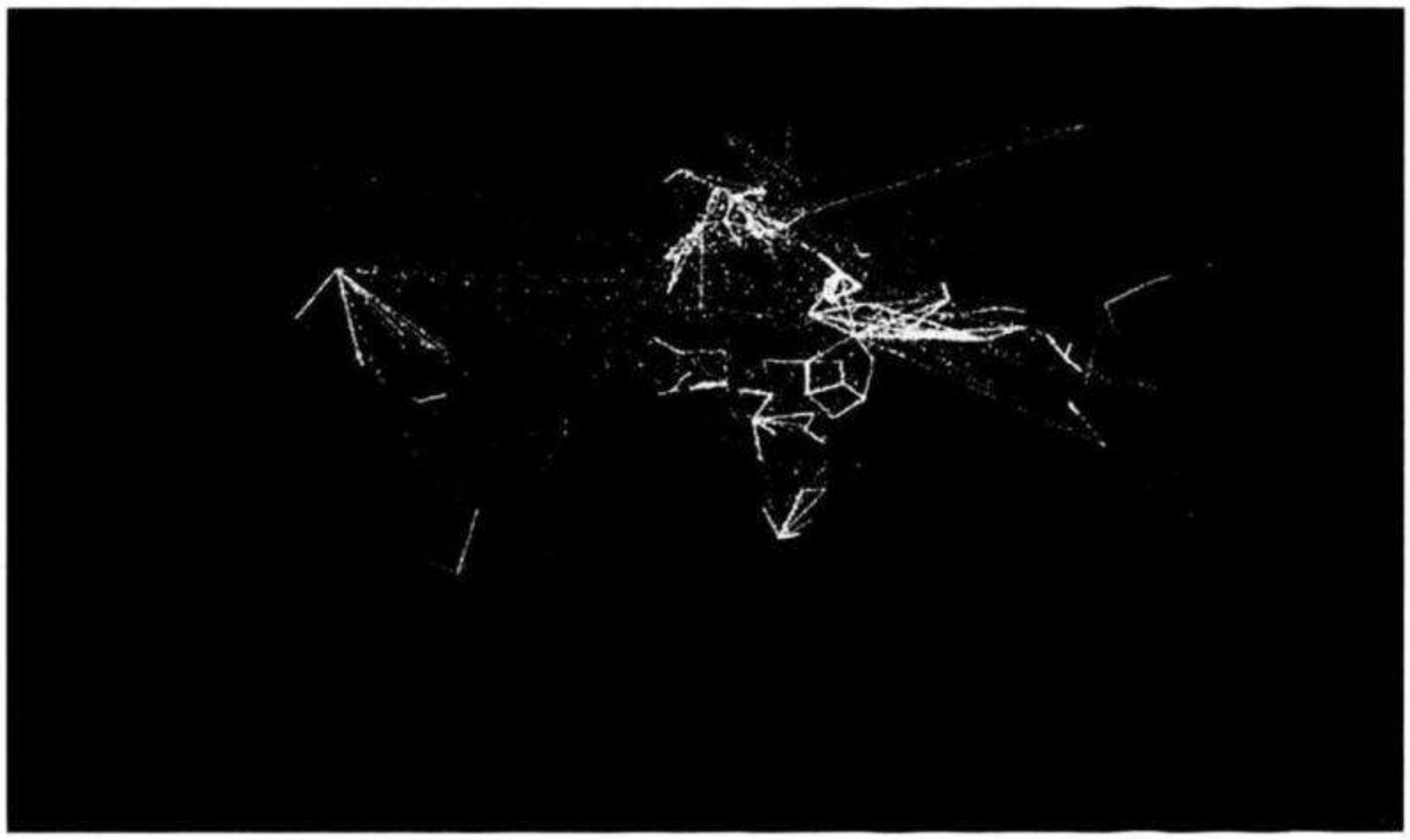
HÌNH 4.12 Ví dụ về số liệu dự đoán doanh số được trình bày trên biểu đồ đường



HÌNH 4.13 Tình hình chi tiêu và thu nhập của chính phủ Anh năm 2014



HÌNH 4.14 Tình hình di cư thuần túy trong năm năm qua



Đồ thị/biểu đồ đường về chuỗi thời gian

Biểu đồ đường là một trong những biểu đồ phổ dụng và cơ bản nhất giúp biểu diễn thông tin về nhiều thông số và/hoặc biến số theo một chuỗi các điểm dữ liệu được kết nối bằng đường kẻ với nhiều hình dáng khác nhau.

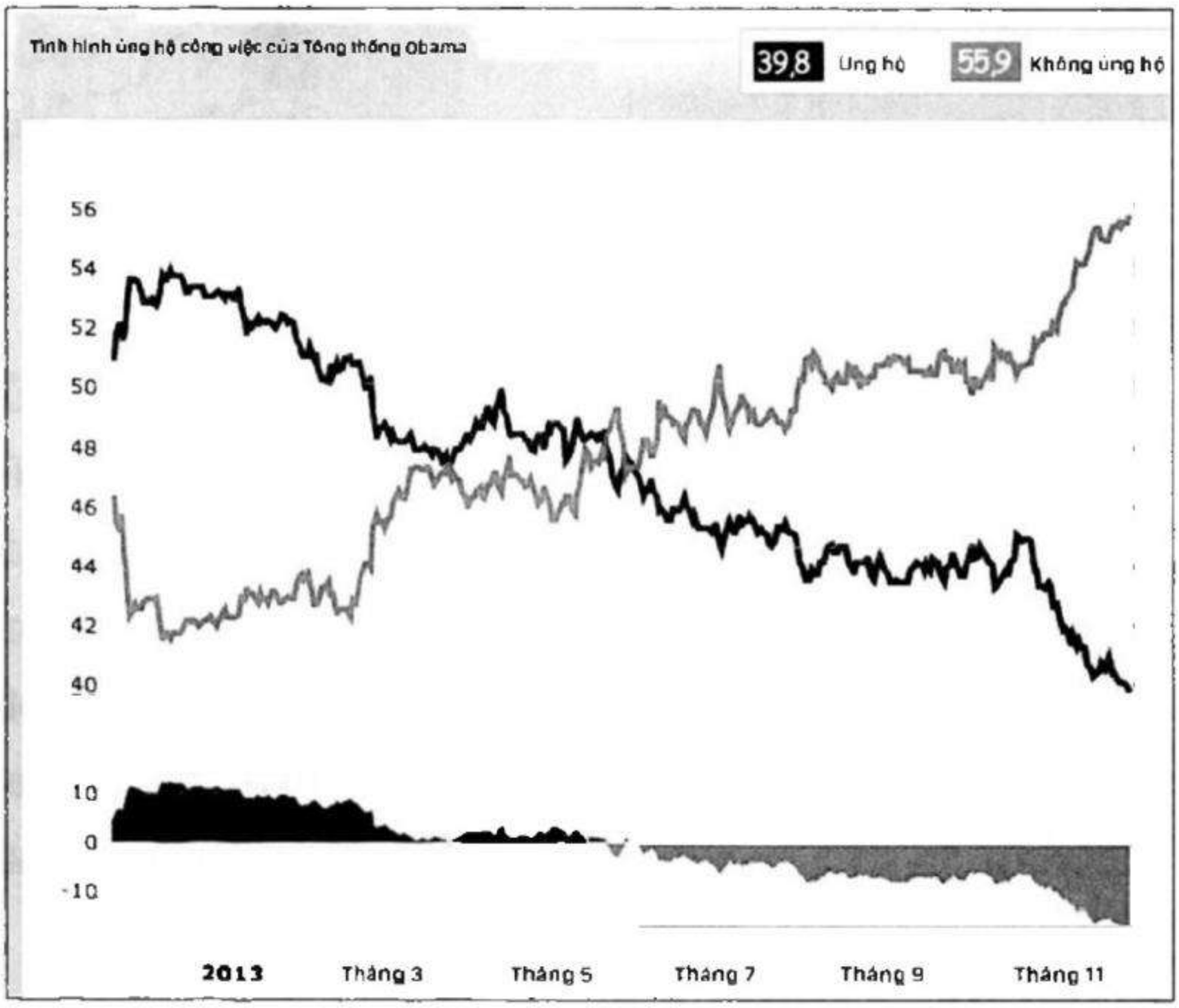
Trong nhiều trường hợp, trục hoành của biểu đồ thể hiện thời gian, theo đó các thay đổi được mô tả theo biến thời gian – vì thế đây còn được gọi là biểu đồ chuỗi thời gian.

Biểu đồ chuỗi thời gian được dùng để xác định sự thay đổi của một xu hướng theo một giai đoạn thời gian – chúng cũng thể hiện thời gian của những diễn biến hay biến động theo mùa và bất thường của các

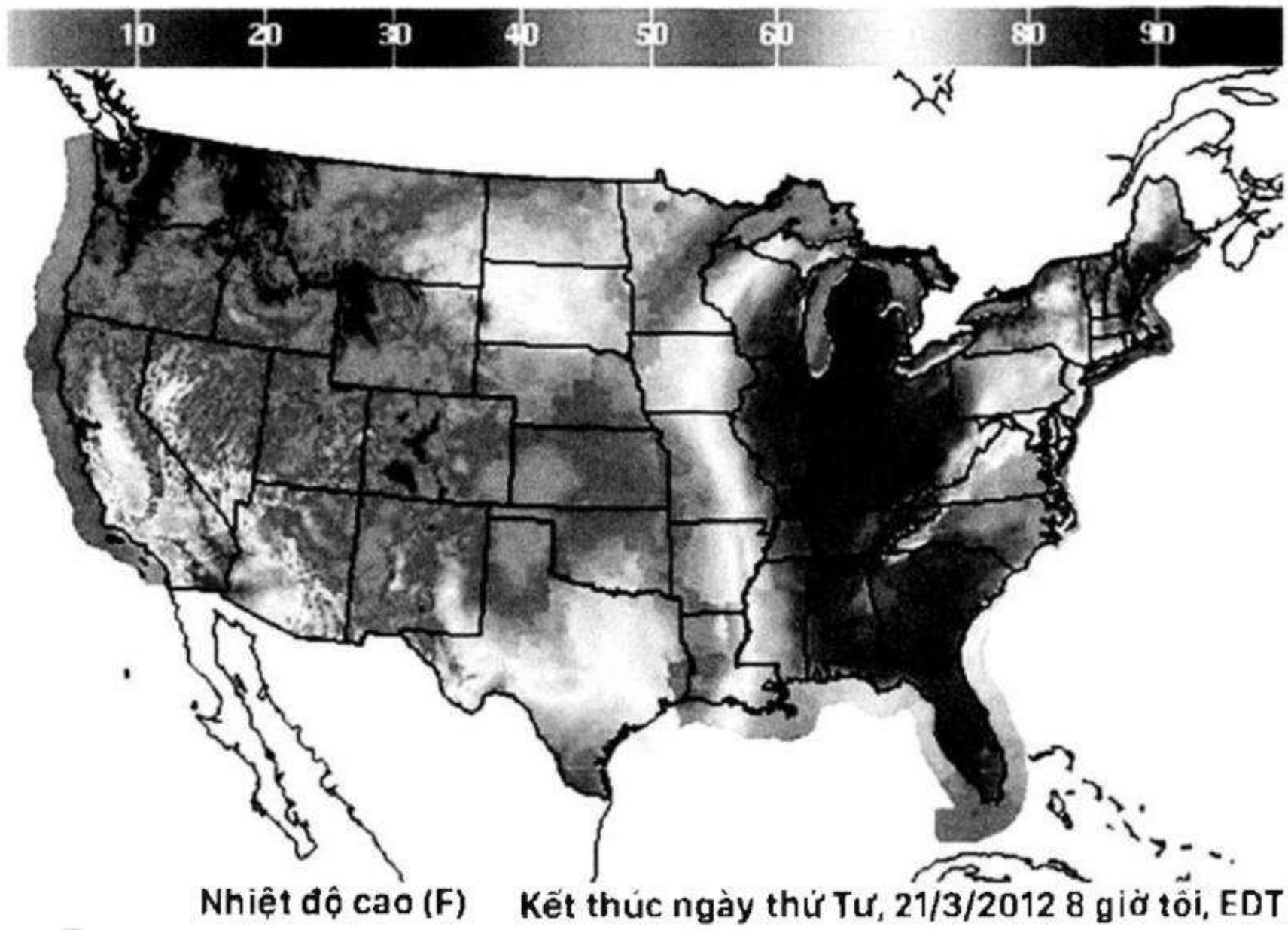
hoạt động/thông tin trong giai đoạn thời gian được phân tích. Có thể lập biểu đồ và phân tích cho nhiều biến khác nhau trong cùng một giai đoạn thời gian.

Hình 4.15 thể hiện tỉ lệ ủng hộ Tổng thống Obama trong năm 2013 dựa trên dữ liệu thu thập được từ ba cuộc trưng cầu dân ý độc lập – chúng ta có thể quan sát thấy rõ ràng những thay đổi về xu hướng ở đây.

HÌNH 4.15 Tỉ lệ ủng hộ Tổng thống Obama trong năm 2013



HÌNH 4.16 Bản đồ nhiệt thể hiện nhiệt độ trong giai đoạn giữa năm 2012 ở nước Mỹ



Bản đồ nhiệt

Bản đồ nhiệt là kỹ thuật biểu diễn dữ liệu bằng đồ thị sao cho các biến hoặc các giá trị riêng lẻ được thể hiện bằng những màu sắc khác nhau trong một biểu đồ hai chiều. Một giá trị từ tập dữ liệu, chẳng hạn doanh thu, đại diện cho kích thước biểu đồ, và một giá trị khác, chẳng hạn tỉ lệ thay đổi, quyết định màu sắc. Bản đồ cây, bản đồ địa lý, biểu đồ tần suất, biểu đồ lưới... là một số bản đồ nhiệt phổ biến.

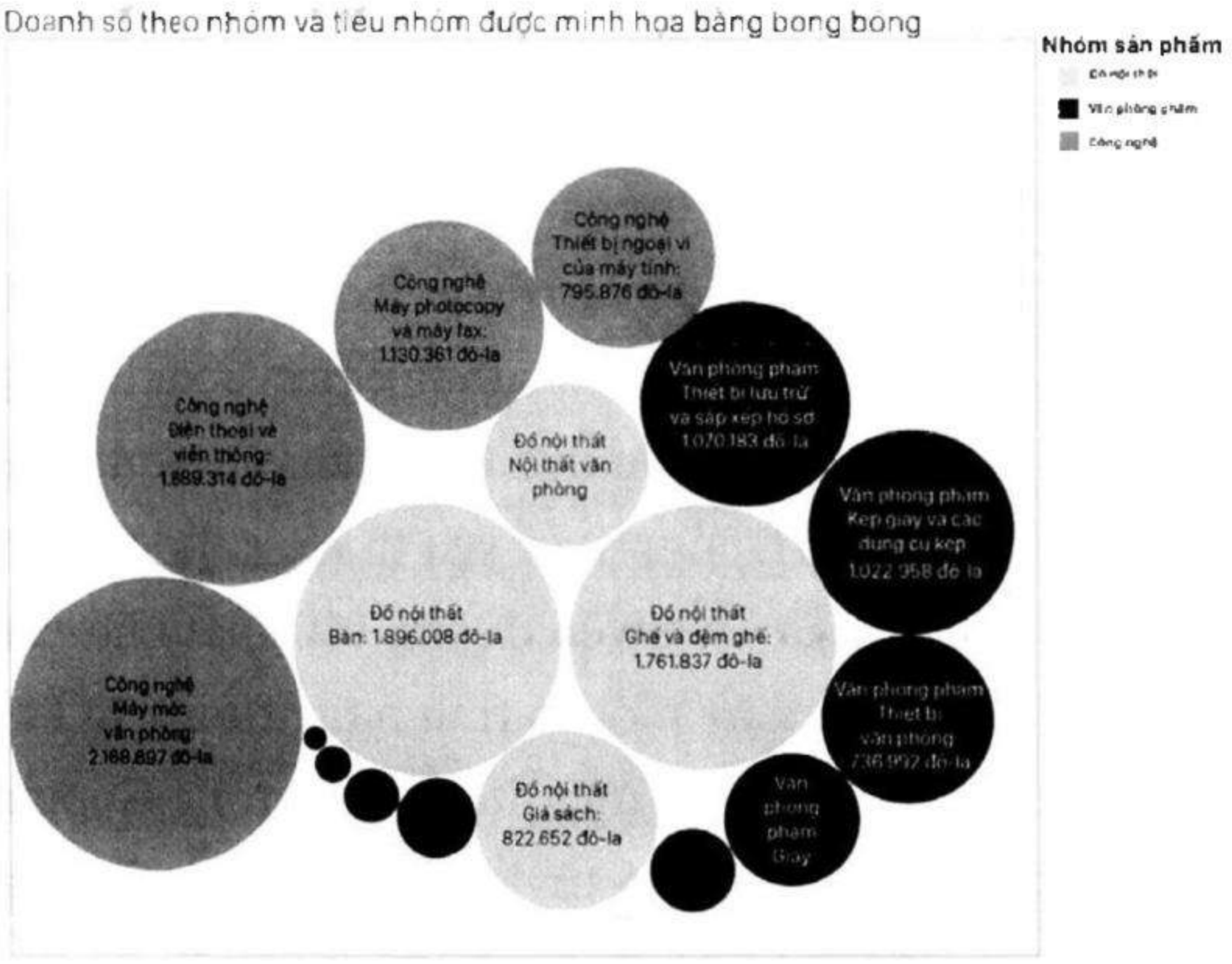
Ví dụ về bản đồ nhiệt ở Hình 4.16 biểu diễn nhiệt độ vào giữa năm 2012 ở Mỹ.

Biểu đồ bong bóng

Biểu đồ bong bóng là kỹ thuật trực quan hóa dữ liệu ba chiều, trong đó hai giá trị đại diện cho vị trí của điểm dữ liệu (x, y) và giá trị của chúng đại diện cho kích thước của quả bong bóng (bán kính của quả bong bóng). Trong một số trường hợp, có thể coi biểu đồ bong bóng là phần mở rộng của biểu đồ phân tán, trong đó mật độ hoặc kích thước được cung cấp như là giá trị thứ ba.

Hình 4.17 biểu diễn trực quan hóa tình hình doanh số bán hàng của một doanh nghiệp hoạt động trong nhiều mảng lĩnh vực khác nhau thông qua biểu đồ bong bóng nhiều màu sắc.

HÌNH 4.17 Một tổ chức hoạt động đa ngành nghề sử dụng biểu đồ bong bóng nhiều màu sắc



Phân tích dự đoán

Năm 2013, Bộ Giao thông Vận tải Anh đã tổ chức cuộc thi hackathon tại Newcastle, Vương quốc Anh. Nếu bạn chưa quen với khái niệm này, chúng tôi xin giới thiệu vắn tắt như sau: Hackathon là một sự kiện ngắn ngày dành cho các lập trình viên; các thí sinh tham gia cuộc thi này sẽ xây dựng một sản phẩm hoặc mô hình sản phẩm khả thi ở mức tối thiểu xung quanh một ý tưởng hoặc nhằm giải quyết một vấn đề do nhà tổ chức đặt ra. Trong cuộc thi hackathon ở Anh, các lập trình viên được giao cho một đề bài rất thú vị: Họ được cấp quyền truy cập vào kho dữ liệu mở về tình hình giao thông vận tải và được yêu cầu chuẩn bị một mô hình dự đoán các hoạt động liên quan đến tình hình giao thông công cộng ở Anh. Một nhóm nhỏ gồm các kỹ sư tài năng đã làm việc xuyên đêm để phát triển một hệ thống sử dụng dữ liệu giao thông nhằm tránh các tuyến đường có lưu lượng xe cộ qua lại đông đúc vào giờ cao điểm trong ngày. Từ dữ liệu cho trước, họ rút ra được thông tin về xu hướng và định vị được các con đường đông đúc và trả phí lưu thông, sau đó dựa trên một số yếu tố khác, họ đã tìm ra được cách dự đoán tuyến đường tốt nhất cho các lái xe. Bạn có thể thấy một số tính năng tương tự khi tìm kiếm đường đi trên ứng dụng chỉ đường Google Maps, nó sẽ gợi ý một số tuyến đường và thời điểm

phù hợp dựa trên hoạt động phân tích dữ liệu theo thời gian thực. Để làm được điều đó, các hệ thống này thực hiện phân tích dữ liệu quá khứ để nhận biết các xu hướng và áp dụng thông tin rút ra vào dữ liệu theo thời gian thực, từ đó dự đoán được những gì sẽ xảy ra tiếp theo và đưa ra những phương án phù hợp cho người dùng.

Ứng dụng tìm kiếm có khả năng tự hoàn thiện câu hỏi hoặc tính năng tự động sửa lỗi đánh máy của Google – Ý bạn là...? – được xây dựng xung quanh việc phân tích dữ liệu của người dùng trong quá khứ; tính năng gợi ý kết bạn của Facebook được xây dựng dựa trên việc phân tích mối liên hệ giữa một người dùng với người dùng khác thông qua nhiều tổ chức, hoạt động, hoặc các bạn bè khác của hai người. Tương tự như Facebook, LinkedIn cũng phát triển thuật toán dự đoán riêng.

Một trong những công cụ gợi ý phổ biến nhất là công cụ do Amazon xây dựng dựa trên cơ sở phân tích dữ liệu và phân tích hành vi của các khách hàng trong quá khứ. Trên cơ sở xu hướng mua sắm trước đây, công cụ gợi ý của Amazon sẽ dự đoán những mặt hàng người dùng có khả năng mua nhất.

Phân tích dự đoán là chiếc chén thánh mà lĩnh vực phân tích dữ liệu lớn ngày đêm truy tìm, và hiện đã có rất nhiều kỹ thuật cũng như thuật toán được

xây dựng để đạt được mục tiêu này trong nhiều tổ chức và ngành dọc khác nhau.

TÓM TẮT

Ở các chương trước, chúng ta đã cùng nhau tìm hiểu về tầm quan trọng của phân tích dữ liệu lớn, và hiểu được vì sao kiến tạo giá trị lại là một trong những mục tiêu quan trọng nhất của bất kỳ doanh nghiệp nào. Trong chương này, chúng ta đã mở rộng việc học về phân tích dữ liệu lớn bằng cách khám phá những kỹ thuật, công cụ, cũng như bàn đến một số ví dụ liên quan đến kỹ thuật phân tích có thể giúp nhà quản lý xác định những kiến thức quý giá ẩn giấu trong biển dữ liệu và vận dụng kiến thức ấy để kiến tạo nên giá trị cho tổ chức của mình. Đây là một số rất ít những công cụ và kỹ thuật quan trọng sẽ giúp ích cho bạn trong hầu hết mọi dự án liên quan đến phân tích dữ liệu lớn, và chúng tôi bàn đến chúng ở chương này nhằm giúp bạn hiểu rõ về cách sử dụng và tính khả thi của chúng trong các trường hợp khác nhau. Mỗi kỹ thuật đều có điểm riêng biệt, và có những kỹ thuật phù hợp hơn với bạn. Vai trò của các kỹ thuật này là giúp chúng ta rút ra được các xu hướng và thực hiện phân tích từ cơ sở dữ liệu đã được tạo ra và thu thập. Khi bạn phụ trách việc triển khai một dự án dữ liệu lớn và cho rằng mình không tiếp cận được tất cả các

loại dữ liệu cần thiết, khi đó bạn có thể phân tích một số công cụ và dữ liệu đại diện bên ngoài để chạy phân tích. Có thể bạn không cần đích thân sử dụng hết tất cả các kỹ thuật hay công cụ này hoặc trực tiếp triển khai chúng, vì đây sẽ là công việc của đội ngũ kỹ thuật, nhưng trên cương vị nhà quản lý, bạn phải hiểu để có thể bàn luận các vấn đề này với họ nhằm đạt được các mục tiêu của dự án một cách hiệu quả.

Như vậy, tới lúc này bạn đã biết về dữ liệu lớn, các kỹ thuật và công nghệ phân tích, việc kiến tạo giá trị, và bạn cũng đã hiểu về một số công cụ quan trọng. Bây giờ là lúc bạn tìm hiểu về bài học quan trọng nhất của cuốn sách này: mô hình C-ADAPT, đề tài mà chúng tôi sẽ bàn đến ở chương tiếp theo.

5

GIỚI THIỆU MÔ HÌNH C-ADAPT

THIẾT KẾ VÀ TRIỂN KHAI MÔ HÌNH

Ở các chương trước, chúng ta đã thấy được một thực tế không thể chối cãi rằng dữ liệu lớn là một cơ hội lớn để các nhà quản lý kiến tạo giá trị tài chính và phi tài chính cho tổ chức của họ. Chúng ta cũng đã tìm hiểu về một số công cụ và kỹ thuật phân tích, đồng thời tham khảo một số ví dụ để các nhà quản lý có thể học hỏi và đủ tự tin để tiến bước trên con đường này. Bất kể lĩnh vực ngành nghề nào, theo ngành dọc hoặc ngành ngang, đều có sự xuất hiện của dữ liệu và những cơ hội gắn liền với chúng. Nhưng ngay cả khi đã nhận thức được điều đó, nhiều nhà quản lý vẫn gặp khó khăn trong việc tận dụng cơ hội này và định hình nên cách tiếp cận đối với vấn đề kiến tạo giá trị – bất luận giá trị tài chính hay phi tài chính.

Thông thường, các nhà quản lý kinh doanh không có kiến thức nền về kỹ thuật cảm thấy rằng họ phải phụ thuộc nhiều vào các đội ngũ hoặc chuyên gia kỹ thuật khi những dự án như dự án dữ liệu lớn bắt đầu phát huy ảnh hưởng trong tổ chức. Về cơ bản, sự phụ thuộc này mang lại những thách thức và rào cản lớn cho các nhà quản lý mong muốn khai thác các cơ hội này theo cách tốt nhất có thể.

Ở Chương 4, chúng tôi đã nhắc đến ví dụ về một công ty phần mềm có trụ sở tại London – trong hầu hết các tổ chức, các nhà quản lý không có kiến thức chuyên môn về kỹ thuật đều sẽ rơi vào tình huống tương tự như vị giám đốc tài chính và giám đốc kinh doanh trong ví dụ đó. Khi nghĩ đến những tình huống như vậy trong các doanh nghiệp, và nhận thức rõ ràng rằng các nhà quản lý không thể lĩnh hội được những kiến thức chuyên môn sâu về kỹ thuật trong một khoảng thời gian ngắn, chúng ta thấy ở đây nảy sinh yêu cầu về một khung tiếp cận đơn giản hơn cho những nhà quản lý đang có kế hoạch triển khai các dự án dữ liệu lớn. Một mô hình và quy trình có hệ thống giúp nhà quản lý tính toán đến chuyện xây dựng chiến lược cho các dự án dữ liệu lớn sẽ trang bị cho họ nhiều lợi thế lớn.

Bên cạnh các khía cạnh kỹ thuật và phi kỹ thuật của quá trình kiến tạo giá trị, chúng ta cũng phải luôn

luôn lưu ý đến phương thức mà chúng ta dự định sử dụng để đo lường giá trị được tạo ra. Hơn nữa, như đã đề cập ở Chương 3, cũng cần nhận thức được rằng sự chênh lệch giữa giá trị được tạo ra và chi phí để tạo ra giá trị đó, bao gồm tất cả các quy trình như chuẩn bị và duy trì dự án dữ liệu lớn, phải có giá trị dương – nếu không, dù có sử dụng đội ngũ và các nguồn lực tốt nhất có thể cho dự án dữ liệu lớn, đồng thời chứng minh được rằng doanh thu tổng đã cải thiện, thì doanh nghiệp rốt cuộc vẫn không thể tạo ra giá trị có đóng góp tích cực cho sự phát triển của tổ chức của họ. Các nhà quản lý phải nghiêm túc đề cao giai đoạn thử nghiệm hoặc đo lường này, bởi nếu giá trị tổng mà dự án phân tích dữ liệu lớn (hoặc bất kỳ dự án nào khác) mang lại có giá trị âm, và đây là quá trình mang tính lặp đi lặp lại, khi đó xét về lâu dài, tổ chức sẽ chịu những khoản thua lỗ tích lũy lớn, có thể đe dọa đến tương lai của doanh nghiệp.

Trong quá trình tham gia vào nhiều dự án và làm việc với nhiều nhà quản lý, chúng tôi nhận ra rằng cần cung cấp cho nhà quản lý một mô hình và khung có hệ thống nhằm giúp họ xác định được những thách thức và cơ hội có thể tiếp cận bằng phân tích dữ liệu lớn, bất kể lĩnh vực hoạt động của họ là gì. Thực ra, chúng tôi đã giới thiệu nhiều kỹ thuật để các nhà quản lý và đội nhóm có thể sử dụng khi triển khai các

dự án phân tích dữ liệu lớn, song có lẽ ở đây chúng tôi phải tổng hợp tất cả lại thành một khung chắc chắn và nhất quán. Dựa trên các công việc đã làm và những quan sát rút ra được từ nhiều trường hợp sử dụng dữ liệu lớn khác nhau, chúng tôi đã xây dựng được một mô hình có thể giúp bất kỳ nhà quản lý nào trong việc triển khai và dẫn dắt các dự án dữ liệu lớn. Giai đoạn cuối cùng của mô hình này sẽ giúp bạn đo lường giá trị được tạo ra, qua đó bảo đảm rằng toàn bộ dự án dữ liệu lớn đang được triển khai hiệu quả.

Trong chương này, chúng tôi sẽ giới thiệu với bạn đọc hai giá trị lớn của cuốn sách: mô hình C-ADAPT và bảng tính C-ADAPT.

1. **Mô hình C-ADAPT** về kiến tạo giá trị dữ liệu lớn

   Mô hình toàn diện này được phát triển dành riêng cho các nhà quản lý để giúp họ xác định các cơ hội và thách thức, định nghĩa dự án cần làm với những mục tiêu rõ ràng, và triển khai chúng để gặt hái được những kết quả tốt. Mô hình này được thiết kế và trình bày theo hướng đơn giản, dễ hiểu để bất kỳ ai, có chuyên môn về bất kỳ mảng nào trong doanh nghiệp, đều không những nhanh chóng tiếp thu được mà còn có thể triển khai áp dụng

nó vào bất kỳ dự án nào liên quan đến dữ liệu lớn. Đây là một mô hình quy trình lặp lại với sáu giai đoạn chính – chúng tôi sẽ trình bày chi tiết về mô hình này và cách sử dụng nó trong các phần tiếp theo. Mô hình này thu gọn lại các khía cạnh chính trong một dự án dữ liệu lớn để nhà quản lý có thể bắt tay vào thực hiện dự án với những mục tiêu ở cấp độ cao, và kết thúc dự án với kết quả đo lường rất rõ ràng về giá trị mà nó tạo ra, sau đó quyết định xem cần thực hiện những bước thay đổi hay cải thiện gì cho vòng lặp tiếp theo. Tuy nhiên, điều quan trọng cần ghi nhớ là chúng ta phải ghi chép lại cẩn thận tất cả các khoản chi phí về mặt tài chính và phi tài chính trước khi đạt đến giai đoạn cuối cùng là *Kiểm định*.

2. **Bảng tính C-ADAPT** của quá trình kiến tạo giá trị dữ liệu lớn

   Đúng như tên gọi, đây là bảng tính mà nhà quản lý có thể sử dụng để xác định một cách có hệ thống các yếu tố bổ sung trong dự án và lập bản đồ cho chúng với những phần khác nhau trong mô hình C-ADAPT nhằm triển khai dự án thành công. Cũng có thể sử dụng bảng tính này làm công cụ tập luyện để trau dồi các kỹ năng cần thiết nhằm xử trí với các dự án dữ

liệu phức tạp. Chúng tôi cũng khuyến khích bạn nên sử dụng bảng tính này để vận dụng vào các dự án đã làm hoặc các trường hợp ví dụ điển hình để có thể nhanh chóng làm chủ mô hình và các quy trình này. Trong chương sau, chúng tôi sẽ cung cấp một số trường hợp ví dụ điển hình để bạn thực hành ứng dụng mô hình này, qua đó bạn sẽ thấy được rằng nó có thể mang lại thành công cho bất kỳ loại hình dự án nào.

Trước khi đi vào phần kế tiếp và bắt tay vào tập sử dụng mô hình và bảng tính C-ADAPT, chúng tôi có một lưu ý: Khi giới thiệu các trường hợp ví dụ điển hình về việc các tổ chức bên ngoài đã gặt hái được thành công ra sao với dữ liệu lớn, người ta hầu như chỉ nói đến việc dự án đã hoàn thành các mục tiêu đặt ra trước khi bắt đầu – trong phần lớn các trường hợp đều khuyết phần tính toán cuối cùng đối với giá trị được tạo ra từ dự án đó, bởi chúng ta không có được nguồn dữ liệu nội bộ (đa phần là dữ liệu tài chính) để đo lường giá trị tạo ra. Nhưng khi bạn, trên cương vị một nhà quản lý thiên về thực hành, thực hiện bất kỳ dự án nào tương tự trong tổ chức của mình, hãy nhớ áp dụng các công thức tính toán và các phương pháp như NPV (đã mô tả ở Chương 3) nhằm bảo đảm rằng dự án bạn phụ trách rốt cuộc sẽ

tạo được giá trị tài chính dương cho tổ chức, và nó không chỉ là dự án bạn làm cho vui.

MÔ HÌNH C-ADAPT TRONG QUÁ TRÌNH KIẾN TẠO GIÁ TRỊ

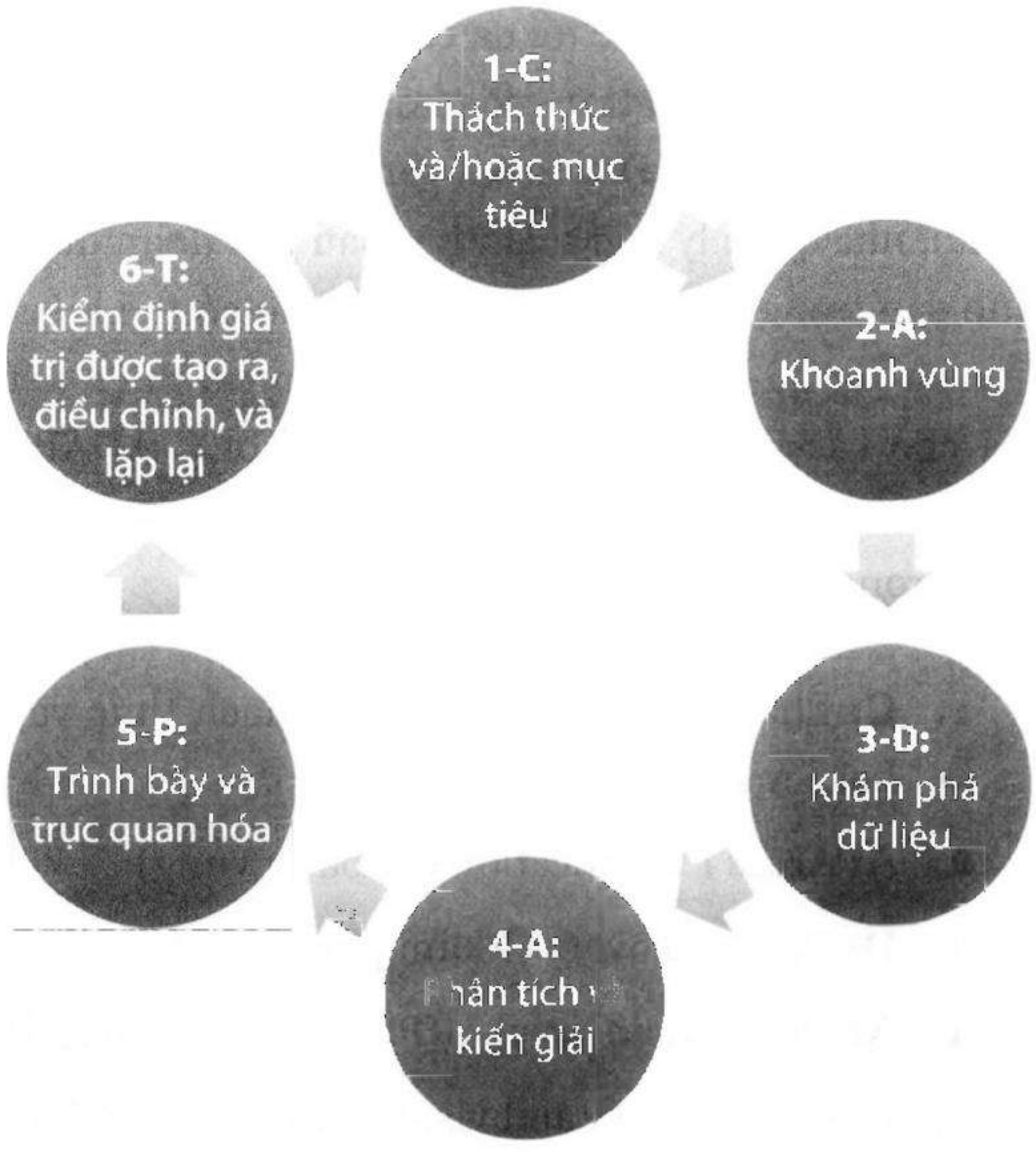
Mô hình C-ADAPT được chúng tôi phát triển sau khi làm việc với nhiều khách hàng lớn nhỏ khác nhau trong các dự án phân tích dữ liệu lớn của họ. Đây là một cách tiếp cận có hệ thống mà bất kỳ nhà quản lý hay nhà lãnh đạo doanh nghiệp nào cũng có thể áp dụng để tận dụng được tối đa cơ hội thực hiện dự án dữ liệu lớn mà họ có trong tay hoặc để giải quyết bất kỳ thách thức nào. Đây là một mô hình mang tính lặp lại, nhà quản lý nên tuân theo sáu giai đoạn được định sẵn trong bất kỳ dự án nào, sau đó có thể kiểm định và lặp lại các giai đoạn trên nhằm cải thiện kết quả cho các vòng lặp tiếp theo. Tên gọi C-ADAPT được cấu thành từ chữ cái đầu của tên gọi sáu giai đoạn trong mô hình này:

1. **C:** Challenges and/or goals (Thách thức và/hoặc mục tiêu)
2. **A:** Areas identification (Khoanh vùng)
3. **D:** Data discovery (Khám phá dữ liệu)
4. **A:** Analysis and insights (Phân tích và kiến giải)
5. **P:** Presentation and visualization (Trình bày và trực quan hóa)

6. **T:** Testing the value created, refine and repeat (Kiểm định giá trị được tạo ra, điều chỉnh, và lặp lại)

C-ADAPT là một mô hình tuần tự mang tính lặp lại (tham khảo về lưu đồ logic của mô hình này ở Hình 5.1).

HÌNH 5.1 Lưu đồ logic của mô hình lặp lại C-ADAPT



1. C (Thách thức và/hoặc mục tiêu)

Đây là giai đoạn đầu tiên của mô hình, trong đó chữ C đại diện cho từ *challenges* (thách thức) – tức một danh sách các vấn đề, đề tài, hay thách thức chính mà doanh nghiệp cần xử lý. Ngoài ra, đây cũng có thể là một danh sách các mục tiêu hoặc cơ hội mới mà doanh nghiệp mong muốn tạo ra hoặc xử lý. Hiển nhiên, giai đoạn này cho thấy rằng chúng ta có thể sử dụng mô hình này để giải quyết các vấn đề trong các dự án hoặc quy trình hiện tại, hoặc để tạo ra những cơ hội mới nhằm đạt được các mục tiêu kinh doanh mới với những kết quả mang tinh thần đổi mới sáng tạo. Đối với nhà quản lý, phải xác định được những thách thức cần xử lý hoặc những mục tiêu cần đạt được trước khi bắt tay vào thực hiện các khâu tiếp theo trong dự án. Ở giai đoạn này, toàn bộ đội dự án có thể tập trung lại để vạch ra một danh sách các mục tiêu (nếu có thể). Trong một số trường hợp, chỉ riêng nhà quản lý đảm nhiệm phần xác định mục tiêu ở giai đoạn này – và điều đó cũng không có vấn đề gì cả.

Để tìm hiểu rõ hơn về mô hình này, hãy cùng bàn đến một ví dụ giả định. Giả sử bạn đang phụ trách bộ phận kinh doanh của một công ty thương mại điện tử tên là E-COMM, chuyên bán các sản phẩm điện tử ở đất nước bạn (ở đây là nước Anh). Cổng web của E-COMM hiện đang vận hành trên một hệ thống

thương mại điện tử tiêu chuẩn giống như Shopify hay Magento; từ phần quản trị của website này, người dùng có thể dễ dàng quản lý sản phẩm, hoạt động thanh toán... Mục tiêu chính của bạn là gia tăng khả năng tạo lợi nhuận cho công ty. Bạn và đội ngũ thực hiện dự án của bạn cần đưa ra một kế hoạch và chiến lược triển khai, sau đó thực hiện nó để đạt được kết quả tốt nhất có thể. Bước đầu tiên là xác định mục tiêu (hoặc nhóm mục tiêu) chính của dự án, và đánh dấu chúng là chữ "C" trong mô hình C-ADAPT.

Một điều thú vị cần lưu ý: Có nhiều yếu tố có thể mang lại tác động cuối cùng đến khả năng sinh lợi của doanh nghiệp – có thể đó là các yếu tố liên quan đến doanh thu gộp hoặc lợi nhuận. Giả sử công ty E-COMM kiếm được 5 triệu bảng lợi nhuận mỗi năm, và đội ngũ của bạn quyết định thực hiện một dự án dữ liệu lớn với chi phí 3 triệu bảng, như vậy chúng ta phải đợi đến khi E-COMM kiếm được ít nhất 8 triệu bảng thì mới đủ cơ sở để khẳng định rằng đây là một dự án thành công.

Chúng tôi sẽ thực hiện phép tính lợi nhuận này ở giai đoạn thứ 6 của mô hình C-ADAPT, tức là giai đoạn kiểm định giá trị được tạo ra, điều chỉnh, và lặp lại.

Đây là những chữ C (thách thức) của dự án: Gia tăng khả năng sinh tạo lợi nhuận của E-COMM.

2. A (Khoanh vùng)

Giai đoạn tiếp theo của mô hình C-ADAPT là chữ "A", đại diện cho các khía cạnh của doanh nghiệp có liên quan đến hoặc chịu ảnh hưởng từ các thách thức hoặc mục tiêu đã được xác định ở giai đoạn trước. Đây là những khía cạnh có thể thiết lập mối liên hệ trực tiếp hoặc gián tiếp với danh sách các thách thức của bạn – và trong một số trường hợp, bạn sẽ phải nghiên cứu một chút để phát hiện ra những khía cạnh ẩn có thể tác động đến các mục tiêu cuối cùng. Bạn cũng có thể lập danh sách các khía cạnh này dựa trên cơ sở là những giả định sau khi đã bàn bạc và hội ý với các nhà quản lý khác, những người đứng đầu các bộ phận nghiệp vụ, cùng các thành viên khác trong đội ngũ của bạn. Danh sách càng dài thì càng tốt cho bạn, bởi bạn có thể loại bỏ một số khía cạnh trong giai đoạn này, nếu sau khi tìm hiểu bạn thấy rằng chúng không phù hợp.

Khi kết hợp những gì đã khoanh vùng được với dự án của chúng ta, trong đó mục tiêu chính là gia tăng khả năng sinh lợi nhuận của E-COMM, chúng ta có thể cân nhắc đến các cách sau đây để đạt được mục tiêu như đã bàn đến ở Chương 3. Tại Chương 3, chúng tôi đã nói rằng việc gia tăng 1% mức giá thành sản phẩm có thể giúp doanh nghiệp gia tăng lợi nhuận 11% – tính tới nay, tăng giá thành sản phẩm vẫn là một trong những phương pháp hiệu quả nhất

nhuận. Tương tự thế, chúng ta cũng; tự thế, chúng ta cũng
các phương án kiến tạo giá trị khác – n kiến tạo giá trị khác –
h sách các yếu tố có thể tác động đến ı tố có thể tác động đến
ợi nhuận mà chúng tôi đã bàn đến ở chúng tôi đã bàn đến ở

á 1%: tăng lợi nhuận 11% nhuận 11%

ı phí biến đổi 1%: tăng lợi nhuận 7,8% 1%: tăng lợi nhuận 7,8%

lượng khách hàng 1%: tăng lợi nhuận hàng 1%: tăng lợi nhuận

n chi phí cố định 1%: tăng lợi nhuận ịnh 1%: tăng lợi nhuận

tăng giá thành sản phẩm đang bán, h sản phẩm đang bán,
đạt được khả năng tạo lợi nhuận cao năng tạo lợi nhuận cao
ợi nhuận sẽ gia tăng theo từng mốc; ia tăng theo từng mốc;
ể ảnh hưởng đến số lượng hàng hóa đến số lượng hàng hóa
giá cao có thể là rào cản khiến một số là rào cản khiến một số
ần chừ khi quyết định mua. Các công yết định mua. Các công
có thể lợi dụng động thái này và đẩy ng động thái này và đẩy
keting cho sản phẩm của họ tập trung phẩm của họ tập trung
nức giá thấp hơn, hoặc coi đây là cơ hơn, hoặc coi đây là cơ
ng giá sản phẩm của họ. Ngược lại, ẩm của họ. Ngược lại,
ếp cận khác từ góc độ chi phí, nhất là góc độ chi phí, nhất là
ợp thị trường có tính cạnh tranh cao có tính cạnh tranh cao
ạy cảm về giá. Việc giảm chi phí sản á. Việc giảm chi phí sản
hoặc logistics có thể giúp giảm chi phí có thể giúp giảm chi phí

bán sản phẩm tổng th
quy trình đều tồn tại và
chức – đàm phán được
cung cấp sản phẩm ho
thô, linh kiện lắp ráp,
có thể giúp doanh nghi
nhuận. Bằng cách này,
và không ai thấy được t
thủ cạnh tranh cũng k
làm cơ hội để hành độn

Bảng 5.1 tóm tắt lạ
trong các yếu tố thúc
chắc chắn đi kèm với tù

Bạn cũng có thể cân
trình hoặc cải thiện hiệ
chu kỳ sản xuất hoặc gi
trình. Nếu có thể tối ưu
lượng hàng hóa bán đư
cao hơn cho E-COMM.
cạnh chức năng có mối l
với mục tiêu cải thiện k
bao gồm: marketing và
đặt sản phẩm...); bán hà
lý mối quan hệ với nhà
và các yếu tố bên ngoài
cả trong mục chữ A nà
hành, nhà cung cấp, đối

bán sản phẩm tổng thể bởi phần lớ
quy trình đều tồn tại và vận hành sẵ
chức – đàm phán được mức giá tốt
cung cấp sản phẩm hoặc dịch vụ, từ
thô, linh kiện lắp ráp, đến vận chuy
có thể giúp doanh nghiệp nâng cao k
nhuận. Bằng cách này, vì bạn vẫn giữ
và không ai thấy được tác động từ bê
thủ cạnh tranh cũng không thể lấy
làm cơ hội để hành động.

Bảng 5.1 tóm tắt lại những điểm
trong các yếu tố thúc đẩy giá trị và
chắc chắn đi kèm với từng yếu tố.

Bạn cũng có thể cân nhắc đến việ
trình hoặc cải thiện hiệu suất vận hà
chu kỳ sản xuất hoặc giảm bớt sự lãn
trình. Nếu có thể tối ưu hóa phần ch
lượng hàng hóa bán được sẽ mang lạ
cao hơn cho E-COMM. Như vậy, dan
cạnh chức năng có mối liên hệ trực tiế
với mục tiêu cải thiện khả năng sinh
bao gồm: marketing và khuyến mãi
đặt sản phẩm...); bán hàng; hiệu suất
lý mối quan hệ với nhà cung cấp; đối
và các yếu tố bên ngoài khác. Chúng
cả trong mục chữ A này: marketing
hành, nhà cung cấp, đối thủ cạnh tran

HÌNH 5.1 Yếu tố thúc đẩy giá trị và thứ bậc trong kiến tạo giá trị

Yếu tố thúc đẩy giá trị	*Thứ bậc trong kiến tạo giá trị*	*Mức độ không chắc chắn khi hấp thụ giá trị*
Doanh thu	Khó dự đoán và khó định lượng	Đây là yếu tố thúc đẩy giá trị kém chắc chắn nhất, vì nó đưa ra những giả định về sự tăng trưởng về số lượng khách hàng hoặc sự gia tăng giá thành sản phẩm ở mức khối lượng hàng hóa nhất định.
Chi phí	Có thể dự đoán và có thể định lượng	Nhìn chung, chi phí là vấn đề nội bộ và yếu tố này mang lại sự chắc chắn cao hơn so với những giả định về việc gia tăng doanh thu.
Vốn lưu động	Có thể dự đoán và có thể định lượng	Vốn lưu động chủ yếu nằm trong tầm kiểm soát của doanh nghiệp, do đó yếu tố này có độ chắc chắn khi dự đoán cao hơn so với doanh thu.
Tài sản cố định	Có thể dự đoán nhưng có thang thời gian dài hơn	Nhìn chung, tài sản cố định là yếu tố mà doanh nghiệp nắm rõ. Nếu giá trị này được tạo ra bằng cách giảm bớt (nghĩa là loại bỏ) các tài sản cố định, khi đó sẽ xuất hiện sự không chắc chắn về giá trị bán lại tài sản và thời gian cần thiết để loại bỏ bớt tài sản đó.

3. D (Khám phá dữ liệu)

Giai đoạn tiếp theo của mô hình C-ADAPT là chữ "D", đại diện cho cụm từ *data discovery*, nghĩa là khám

phá dữ liệu cho các khía cạnh đã được lựa chọn và khoanh vùng. Ở giai đoạn này, nhiệm vụ của bạn là liệt kê tất cả các nguồn dữ liệu và loại dữ liệu liên quan đến những khía cạnh đó. Khi làm việc với dữ liệu, có một lưu ý quan trọng cần nhớ là một số tập dữ liệu đã có sẵn trong nội bộ tổ chức của bạn, và một số lại đang tồn tại bên ngoài tổ chức. Một số dữ liệu có thể đã được định nghĩa và có cấu trúc rõ ràng, nhưng một số khác lại chưa được xác định rõ, hay tồn tại dưới dạng phi cấu trúc, hoặc nằm phân tán rải rác khắp nơi. Dĩ nhiên, có thể bạn không có điều kiện để tiếp cận trực tiếp tất cả các nguồn dữ liệu (chủ yếu là các dữ liệu bên ngoài và một số dữ liệu nội bộ), nhưng ở giai đoạn này, bạn chỉ cần xác định và liệt kê càng nhiều nguồn dữ liệu càng tốt. Một số ví dụ về các tập dữ liệu bên ngoài tổ chức: nội dung các cuộc trao đổi của mọi người nói về công ty bạn, sản phẩm của bạn, đối thủ cạnh tranh của bạn, hoặc những sản phẩm bổ trợ trên các nền tảng mạng xã hội như Twitter hay Facebook. Phần lớn các dữ liệu bên ngoài này sẽ là dữ liệu phi cấu trúc, và bạn có thể tạo giá trị cũng như tạo sự khác biệt từ các dữ liệu phi cấu trúc đó. Trong hầu hết các công cụ tình báo doanh nghiệp hiện nay đều đã có sẵn tính năng phân tích dữ liệu có cấu trúc trong nội bộ, và đề tài này cũng đã được bàn đến chi tiết trong nhiều cuốn sách trước đây. Vì vậy, trong cuốn sách này, chúng tôi chủ yếu tập trung vào tiềm

năng còn ẩn giấu của dữ liệu phi cấu trúc, đồng thời bàn đến cách vận dụng các công nghệ và công cụ đã gợi ý ở Chương 4 vào quá trình phân tích dữ liệu phi cấu trúc. Có thể sử dụng các công cụ theo dõi mạng xã hội hiện có trên thị trường để thu thập một số dữ liệu phi cấu trúc. Những dữ liệu này cũng có thể được thu thập thông qua việc sử dụng các quy trình phản hồi khách hàng trực tuyến hoặc phi trực tuyến. Chúng ta cũng có thể phân nhóm các tập dữ liệu này là nội bộ hoặc bên ngoài.

Trong ví dụ về công ty E-COMM, chúng ta có thể gắn các dữ liệu sau đây với các khía cạnh đã khoanh vùng từ giai đoạn trước của mô hình C-ADAPT:

Marketing: Xét từ góc độ marketing, bạn có thể xác định các thông tin và dữ liệu chính về hồ sơ khách hàng, phân khúc thị trường/thông tin nhân khẩu học, xu hướng mua sắm, yếu tố kích thích mua sắm, quy trình mua sắm... Có thể thu thập các dữ liệu này từ dữ liệu về doanh số đang tồn tại sẵn trong hệ thống CRM (Customer Relationship Management – Quản lý quan hệ khách hàng) của doanh nghiệp. Sẽ có cả dữ liệu từ những phản ứng của khách hàng trước các chương trình khuyến mãi mà bạn từng thực hiện trong quá khứ, và từ những phản ứng của khách hàng mà bạn thu thập được qua dữ liệu về các chiến dịch marketing và các hoạt động thực hiện trong giai đoạn trước

khi bán hàng. Trong trường hợp các chương trình khuyến mãi thực hiện trực tuyến trong không gian số, những dữ liệu này có thể tồn tại trên các nền tảng quản lý chiến dịch như Google Ads, Google Analytics, Facebook, và Bing Advertising. Nếu liên quan đến các chương trình thực hiện ngoài đời thực, khi đó đội ngũ marketing thường sẽ chịu trách nhiệm kiểm soát các hệ thống quản lý dữ liệu nội bộ để cập nhật theo phương thức thủ công... Với các cuộc trao đổi trong không gian mạng xã hội như Facebook và Twitter, bạn có thể thu thập dữ liệu thông qua các công cụ theo dõi như Hootsuite hoặc Odimax. Dữ liệu đến từ mạng xã hội phần lớn sẽ là dữ liệu phi cấu trúc, do đó chúng ta sẽ phải sử dụng các kỹ thuật phân tích văn bản hoặc phân tích cảm xúc để có thể tạo ra giá trị từ nguồn dữ liệu này. Những người giữ vai trò quản trị viên của website sẽ nắm được dữ liệu về chi phí và giá thành của sản phẩm được bán – nếu sử dụng Shopify, Magento, hoặc Wordpress, bạn sẽ có thể dễ dàng truy cập vào phần quản trị trên cổng web và tải xuống dữ liệu quản lý danh mục sản phẩm.

Bán hàng: Khi chúng tôi nghiên cứu cách cải thiện khả năng sinh lợi nhuận cho bất kỳ công ty nào, doanh số cũng trở thành một trong những phần việc quan trọng nhất của dự án. Thông thường, việc phân tách hoàn toàn dữ liệu đến từ hoạt động bán hàng

và dữ liệu đến từ hoạt động marketing là một nhiệm vụ rất khó khăn, do đó, có thể bạn sẽ gặp trường hợp một số tập dữ liệu hoặc nguồn dữ liệu xuất hiện ở cả hai khía cạnh bán hàng và marketing. Với các dữ liệu như dữ liệu về sản phẩm được ưa chuộng nhất xét về số lượng sản phẩm đã bán ra, và dữ liệu về sản phẩm có lãi nhất dựa trên cách tính biên lợi nhuận gộp, bạn có thể thu thập chúng thông qua các dữ liệu ở CRM và SCM. Những sản phẩm có ít người quan tâm nhất nằm ở trang chủ hoặc trang quảng cáo và không tạo ra kết quả khả quan nào có thể sẽ được điều chỉnh lại – có thể tìm ra thông tin này bằng cách thực hiện nghiên cứu kết hợp giữa dữ liệu ở hệ thống CRM và dữ liệu của bộ phận marketing. Giả sử khách hàng xem sản phẩm trên mạng, sau đó gọi điện đến cho bộ phận bán hàng hoặc trao đổi với nhân viên bán hàng/ nhân viên hỗ trợ đang trực trên website để hỏi chi tiết thêm về sản phẩm rồi quyết định đặt mua qua điện thoại – trong trường hợp này, dữ liệu sẽ được lưu ở hệ thống MIS (Management Information System – Hệ thống thông tin quản lý) của đội ngũ bán hàng và đội ngũ hỗ trợ. Khi hoạt động bán hàng diễn ra trên điện thoại hoặc ô cửa sổ trò chuyện trực tuyến, bạn sẽ phải truy cập được vào nơi lưu trữ cuộc trao đổi giữa khách hàng với nhân viên bán hàng và chuyển nội dung cuộc trao đổi đó thành văn bản nếu cần thiết (thực ra, nhiều gói phần mềm cũng thực hiện được tính

năng này). Các dữ liệu này, chủ yếu là phi cấu trúc, sẽ giúp bạn xác định xem kênh bán hàng nào thực hiện được bao nhiêu giao dịch bán hàng, và tỉ lệ chuyển đổi thành khách hàng của các cuộc trao đổi qua điện thoại và ứng dụng nhắn tin nhanh là bao nhiêu. Các nguồn cung cấp dữ liệu MIS (hay CRM) cũng sẽ ghi lại những dữ liệu trao đổi giữa khách hàng và nhân viên, sau đó rút ra được những thông tin sâu sắc như: trên thị trường đang bàn luận về điều gì, đề tài được bàn luận sôi nổi nhất là gì – những ý kiến phàn nàn hay thông tin... Doanh nghiệp cũng có thể phải tìm hiểu về kết quả làm việc của những thành viên trong các đội bán hàng và hỗ trợ nếu tỉ lệ chuyển đổi thành khách hàng nói chung của đội có phần cao hơn so với kết quả của các thành viên vốn có hiệu suất làm việc kém – những thông tin này có thể được lưu trữ trong hệ thống quản lý hiệu suất công việc của nhân viên.

Vận hành: Cần xác định tất cả các điểm tiếp xúc để xem điểm nào gắn với hoạt động bán hàng và hỗ trợ, điểm nào gắn với vận hành. Về cơ bản, vận hành là các trung tâm chi phí[1], do đó quy trình trong tổ chức càng được tối ưu và hoạt động thông suốt, bạn càng dễ dàng cắt giảm được các chi phí liên quan. Bạn có

---

[1]Trung tâm chi phí: Là bộ phận hoặc phòng ban trong một tổ chức không sản xuất hay cung cấp dịch vụ trực tiếp nhưng vẫn tốn chi phí vận hành. (ND)

thể lập bản đồ cho toàn bộ quá trình từ lúc nhận sản phẩm hoặc phụ tùng lắp ráp trong nhà kho đến hoàn thành đơn hàng và bàn giao sản phẩm theo các đơn hàng khác nhau. Việc xác định tất cả các quy trình vận hành và tiếp cận các dữ liệu liên quan đến yêu cầu về nguồn lực và giới hạn về sản xuất và sản lượng đầu ra sẽ giúp bạn thực hiện được tối ưu hóa quy trình khi cần thiết. Bạn cũng có thể nghĩ đến chuyện xác định các khía cạnh hiện đang tốn kém nhiều chi phí hơn so với các khía cạnh khác để thực hiện những thay đổi phù hợp. Nên tìm hiểu kỹ về việc phải có bao nhiêu nguồn lực tham gia vào toàn bộ quá trình này, và đâu là những điểm tiếp xúc bên trong và bên ngoài tổ chức trong quá trình bàn giao sản phẩm, bao gồm cả khâu đóng gói và vận chuyển hàng.

Để cải thiện khâu vận hành, bạn cũng có thể cân nhắc đến việc phân tích các file nhật ký hệ thống, các email mã hóa... – ở đây, kỹ thuật phân tích ngữ nghĩa sẽ giúp bạn rút ra được những kiến giải chính nhằm tiết giảm chi phí bằng cách cải thiện chất lượng vận hành.

Nhà cung cấp: Hệ thống thu mua sẽ giúp bạn tiếp cận với tất cả các nhà cung cấp, thông tin chi tiết về các hoạt động giao dịch của họ, các loại hàng hóa/dịch vụ mà họ cung cấp cùng với các điều khoản, điều kiện và chi phí trong hợp đồng... Đây cũng là một khía

cạnh quan trọng bởi nếu có thể, bạn sẽ có cơ hội để đàm phán lại các điều khoản và chi phí hàng hóa với các nhà cung cấp hiện tại, qua đó tạo ảnh hưởng trực tiếp đến khả năng sinh lợi nhuận của công ty bằng cách vẫn cung cấp các sản phẩm tương tự như trước ở mức giá cũ. Hãy liệt kê tất cả các nhà cung cấp, chiến lược định giá sản phẩm của họ, biên lợi nhuận đối với hàng hóa mà họ cung cấp, lịch trình giao hàng đã thống nhất... Đôi khi bạn cũng nên sử dụng các dữ liệu bên ngoài thu thập được từ website, mạng xã hội, hoặc các nguồn khác về nhà cung cấp và những sản phẩm của họ – biết đâu nhờ vậy bạn phát hiện được thêm thông tin về các mối quan hệ của họ cũng như những sản phẩm họ chào bán cho các bên khác trên thị trường – thông tin này có thể trang bị cho bạn một vài luận điểm hữu ích giúp bạn đàm phán được thỏa thuận hợp tác có lợi hơn. Cũng có thể thu thập thông tin từ các thị trường khác nếu bạn sẵn sàng nhập nguồn sản phẩm từ nước ngoài. Trong hầu hết các trường hợp, chi phí sản phẩm sẽ thấp hơn nếu bạn có thể nhập khẩu hàng hóa từ các quốc gia có chi phí thấp như Trung Quốc hay Ấn Độ. Vì E-COMM bán các sản phẩm điện tử trên mạng, nên khả năng cao là bạn có thể nhập một số nguồn sản phẩm, thậm chí tất cả, từ nước ngoài – trong trường hợp đó, dữ liệu thu thập được từ những thị trường khác về các nhà cung cấp cũng cần được tính đến trong quá trình phân tích.

Đối thủ cạnh tranh: Nghiên cứu về đối thủ cạnh tranh bao giờ cũng là một yếu tố hết sức quan trọng trong bản phân tích về khả năng sinh lợi nhuận của doanh nghiệp. Nếu khách hàng tiềm năng không mua hàng của bạn, thì họ có thể mua sản phẩm mà bạn cung cấp (sản phẩm thay thế) ở đâu, bằng cách nào, và vì sao? Hiện nay đã có nhiều sản phẩm phân tích đối thủ cạnh tranh dành cho doanh nghiệp, bao gồm cả các doanh nghiệp hoạt động trong lĩnh vực thương mại điện tử như E-COMM. Các công cụ như SEMRush có thể giúp bạn lấy được thông tin so sánh về các hoạt động marketing trên mạng của công ty bạn và các đối thủ cạnh tranh – khách hàng đang tìm kiếm cái gì, tìm kiếm như thế nào, và đối thủ của bạn đang đáp lại nhu cầu đó ra sao, tất cả đều là những thông tin hết sức quan trọng đối với bất kỳ doanh nghiệp nào. Nếu không có sẵn thông tin về danh mục sản phẩm, các đặc điểm sản phẩm, chính sách định giá, chiến lược marketing, các hoạt động quảng bá... của đối thủ, hãy thu thập bằng các công cụ tự động, bởi các thông tin này rất có giá trị trong quá trình phân tích. Hầu hết các thông tin này đều là dữ liệu phi cấu trúc từ nguồn bên ngoài – các công cụ theo dõi mạng xã hội và hoạt động nghiên cứu thị trường có thể là những giải pháp tuyệt vời giúp bạn thu thập các dữ liệu này, sau đó bạn hãy phân tích dữ liệu thu

được bằng kỹ thuật phân tích văn bản và phân tích ngữ nghĩa như đã bàn đến ở Chương 4.

Các yếu tố bên ngoài: Hầu như ai trong chúng ta cũng biết đến phân tích PESTEL (nếu không, hãy xem Hình 5.2 và tham khảo phần giới thiệu ngắn bên dưới). Bạn cũng cần thực hiện phân tích PESTEL cho E-COMM bởi nhiều yếu tố bên ngoài cũng có tác động trực tiếp đến khả năng sinh lợi nhuận của doanh nghiệp. Những dữ liệu liên quan đến các yếu tố về chính trị, môi trường, xã hội, công nghệ và pháp lý sẽ mang lại cho bạn rất nhiều kiến giải. Chẳng hạn, nếu E-COMM nhập khẩu một số sản phẩm từ các nước bên ngoài khối Liên minh châu Âu, ví dụ như Trung Quốc, khi đó chi phí cập bến[1] ở Anh có thể sẽ chịu ảnh hưởng bởi các sự kiện chính trị như Brexit[2]. Sau cuộc trưng cầu dân ý với kết quả ủng hộ cho Brexit, tại thời điểm cuốn sách này được viết, nước Anh đang chuẩn bị rời khỏi Liên minh châu Âu, và sự kiện này sẽ có tác động tiêu cực đến giá trị đồng bảng của Anh. Điều này nghĩa là với cùng một lượng đồng bảng Anh như trước, bạn chỉ đổi được một lượng đô-la ít hơn, hệ quả là hầu hết các hoạt động xuất nhập khẩu bên ngoài

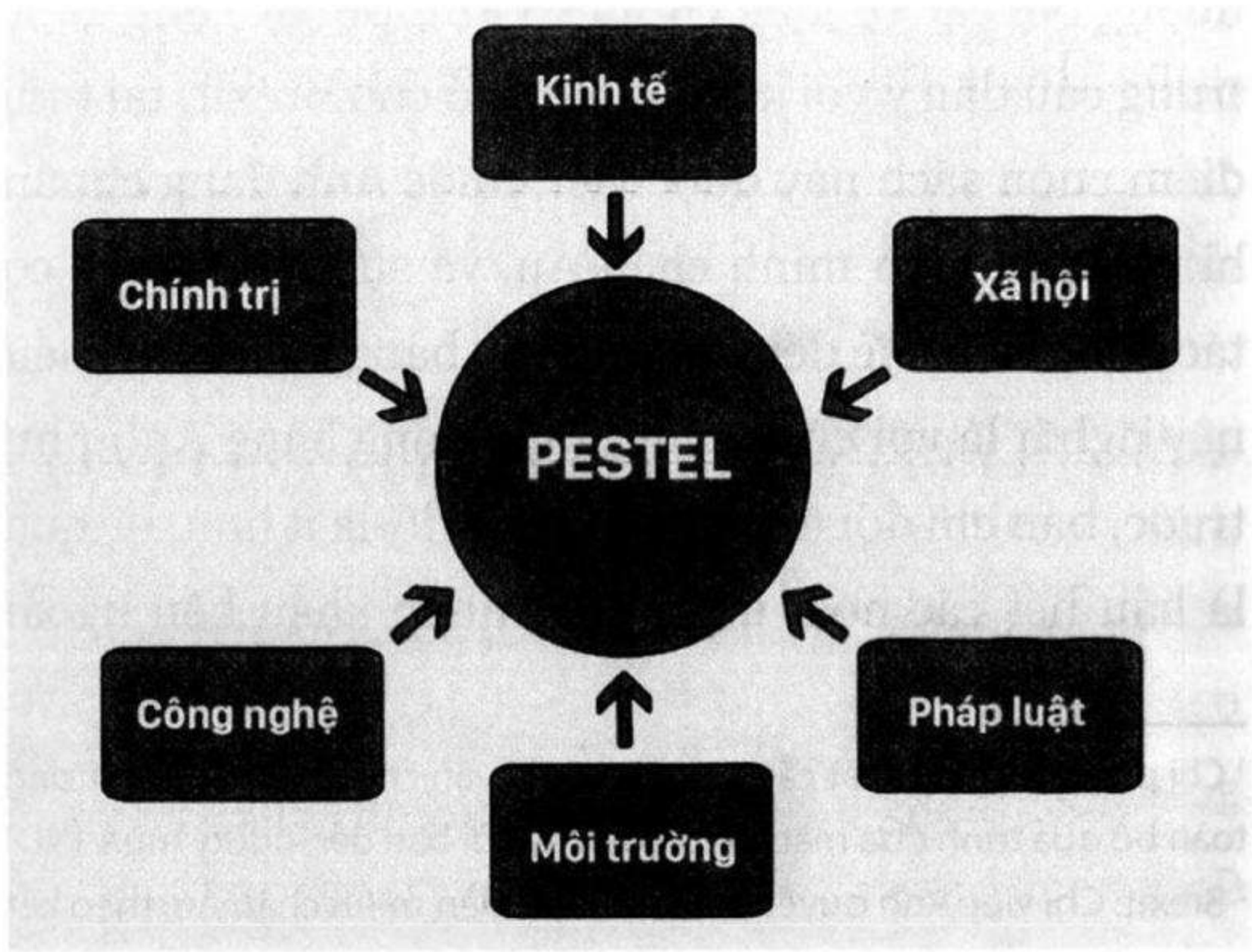
---

[1] Chi phí cập bến: Là tất cả các loại chi phí vận chuyển phát sinh trong toàn bộ quá trình đưa mặt hàng đó từ người bán đến người mua. (ND)

[2] Brexit: Chỉ việc Anh quyết định rút khỏi Liên minh châu Âu, theo kết quả cuộc trưng cầu dân ý năm 2016. (ND)

khối Liên minh châu Âu đều được thực hiện bằng đồng đô-la Mỹ, và để mua bất kỳ loại hàng hóa nào từ Trung Quốc hay Ấn Độ, các công ty của Anh như E-COMM sẽ phải bỏ ra nhiều tiền hơn so với trước đây. Sự thay đổi này sẽ có tác động trực tiếp đến chi phí sản phẩm, dẫn tới biên lợi nhuận thấp hơn. Những thay đổi về chính trị như vậy cũng ảnh hưởng lớn đến thị trường. Người dân Anh cố gắng thắt lưng buộc bụng hơn, và xu hướng này cũng gây ảnh hưởng tới doanh thu gộp của các doanh nghiệp. Một dấu hiệu rõ ràng là giá nhà đất ở Anh đã quay đầu đi xuống sau cuộc trưng cầu dân ý ủng hộ cho Brexit.

HÌNH 5.2 Khung phân tích PESTEL



Lưu ý: Phân tích PESTEL là một khung hoặc công cụ được các chuyên gia marketing sử dụng để phân tích và giám sát các yếu tố liên quan đến môi trường vi mô (môi trường marketing bên ngoài tổ chức) có tác động đến doanh nghiệp.

Các sự kiện và hoạt động bên ngoài khác như các chiến dịch "Made in China versus Made in the UK" (Sản xuất tại Trung Quốc hay sản xuất tại Anh) vẫn luôn tác động đến các doanh nghiệp tùy theo mô hình hoạt động của họ. Những doanh nghiệp bán sản phẩm sản xuất tại Anh có sự gia tăng doanh thu trong thời gian gần đây, trong khi các doanh nghiệp khác lại sụt giảm doanh thu. Chiến dịch bảo vệ môi trường gần đây nhất của tổ chức Hòa bình Xanh có tác động gì đến các sản phẩm làm từ nhựa của bạn? Đây là một mảng thông tin rất cần thiết, bởi phần lớn các sản phẩm của E-COMM đều sử dụng nhựa làm nguyên liệu chính.

Như vậy, có rất nhiều yếu tố bên ngoài và luồng dữ liệu đi kèm để bạn cân nhắc đưa vào trong phân tích của mình.

Phần lớn dữ liệu liên quan đến các hoạt động bên ngoài như vậy đều tồn tại trên website, các cổng tin tức, hoặc mạng xã hội – và chúng thường tồn tại ở dạng thức văn bản thô thuần túy; trong trường hợp này, hãy thực hiện phân tích cảm xúc với các đề tài

hoặc từ khóa phổ biến như Brexit, như vậy bạn sẽ xác định được các xu hướng và hoạt động có thể ảnh hưởng đến doanh nghiệp của mình trong tương lai.

Tới đây, hãy cùng nhau tóm tắt lại danh sách các chữ D (dataset – tập dữ liệu) nên đưa vào trong dự án này: dữ liệu marketing và bán hàng từ hệ thống CRM và các nguồn khác (dữ liệu có cấu trúc và phi cấu trúc), Google Adwords, Google Analytics, Facebook Advertising, dòng thời gian trên Facebook và Twitter (dữ liệu phi cấu trúc), dữ liệu quản lý danh mục sản phẩm (dữ liệu có cấu trúc), hệ thống MIS cho đội ngũ bán hàng và đội ngũ hỗ trợ, dữ liệu về nội dung các cuộc trao đổi qua điện thoại/ứng dụng trò chuyện trực tuyến (dữ liệu phi cấu trúc), dữ liệu về thu mua (dữ liệu có cấu trúc), dữ liệu về cung ứng trên thị trường mở, dữ liệu về cung ứng nước ngoài, dữ liệu nghiên cứu thị trường và đối thủ cạnh tranh trong không gian trực tuyến, dữ liệu về các yếu tố bên ngoài (dữ liệu có cấu trúc và phi cấu trúc)...

4. A (Phân tích và kiến giải)

Trong các giai đoạn trước, chúng ta đã cùng nhau chuẩn bị một danh sách các nguồn dữ liệu nên đưa vào trong phần phân tích. Giai đoạn tiếp theo sẽ là chữ "A" đại diện cho các phân tích và kiến giải rút ra được từ quá trình phân tích dữ liệu. Trước khi bắt

tay vào khâu phân tích, bạn cần sắp xếp lại tất cả các nguồn dữ liệu theo đúng trình tự, ở đúng định dạng nhằm bảo đảm sự đồng bộ và sẵn sàng cho việc phân tích. Trước khi giai đoạn phân tích thực sự bắt đầu, các doanh nghiệp thường thực hiện một quy trình gọi là ETL đối với các tập dữ liệu. ETL là từ viết tắt bao gồm chữ cái đầu của các từ Extract (trích xuất), Transform (biến đổi) và Load (tải) – đây là ba chức năng trong cơ sở dữ liệu được kết hợp lại thành một công cụ để chuyển dữ liệu từ cơ sở dữ liệu này sang cơ sở dữ liệu khác. Trích xuất là quá trình đọc dữ liệu từ một cơ sở dữ liệu. Ở giai đoạn này, nhà quản lý có thể huy động sự hỗ trợ của đội ngũ kỹ thuật để họ thiết lập luồng chảy dữ liệu từ các nguồn dữ liệu khác nhau (như mô tả ở giai đoạn vừa qua) và chạy ETL trên đó để chuẩn bị sẵn sàng cho chúng trong khâu phân tích thực sự.

Như đã thấy ở giai đoạn vừa qua, phần lớn các dữ liệu được cân nhắc đưa vào khâu phân tích đều là dữ liệu phi cấu trúc, nên các kỹ thuật phân tích văn bản, phân tích ngữ nghĩa, và phân tích cảm xúc sẽ được sử dụng theo nhiều hình thức khác nhau – nhưng trước khi thực hiện ETL, chúng ta sẽ phải triển khai các khâu làm sạch dữ liệu, cấu trúc dữ liệu, và lưu trữ dữ liệu.

Sau khi dữ liệu đã được làm sạch và định dạng lại, và các quy trình cũng đã sẵn sàng để thực hiện

ETL trên dữ liệu theo thời gian thực, bây giờ bạn có thể bắt tay vào sử dụng các công cụ và kỹ thuật cần thiết cho việc phân tích. Lúc này, tất cả các tập dữ liệu đều phải được phân tích thông qua nhiều kỹ thuật khác nhau như đã bàn đến ở các chương trước; mục đích là sử dụng các kiến giải và kết quả rút ra để đạt được mục tiêu cải thiện khả năng sinh lợi nhuận cho E-COMM. Nhiều nhà quản lý chưa thực sự cảm thấy tự tin khi sử dụng các phương pháp phân tích tinh vi và phức tạp ở giai đoạn này; trong trường hợp đó, họ có thể tìm kiếm sự hỗ trợ của đội ngũ kỹ thuật để tiến hành phân tích các tập dữ liệu. Tuy vậy, ngay cả một nhà quản lý không có chuyên môn kỹ thuật cũng có thể chuẩn bị sẵn một danh sách các giả định dựa trên những dữ liệu đã thu thập và yêu cầu đội kỹ thuật kiểm chứng các giả định đó song song với việc tìm kiếm những thông tin giá trị từ dữ liệu. Nếu đã đọc toàn bộ các chương trước, có lẽ bạn cũng đã có hiểu biết tương đối về những kỹ thuật có thể sử dụng đối với các tập dữ liệu khác nhau cho các mục đích khác nhau. Dựa trên kiến thức đó, bạn có thể trao đổi với đội ngũ của mình để họ giúp bạn đưa ra được những kết quả nhằm trả lời thỏa đáng cho các giả định bạn đã đặt ra. Trên cương vị nhà quản lý, bạn có thể yêu cầu đội ngũ kỹ thuật tìm ra tác động hoặc xu hướng bán hàng trong mùa Giáng sinh hoặc khi đối thủ cạnh tranh tung ra sản phẩm mới – về cơ bản, bạn đang

cung cấp cho họ một khung thời gian và yêu cầu họ thiết lập sự liên hệ, mô hình xu hướng, và mối tương quan giữa các tập dữ liệu marketing và bán hàng khác nhau được thu thập từ nhiều nguồn khác nhau, ví dụ hệ thống CRM, dữ liệu phản hồi của khách hàng được lưu trữ trong các phiếu ghi ý kiến đóng góp in ra giấy. Bạn cùng có thể có được thông tin về mức độ nhạy cảm về giá của thị trường và tác động của sự thay đổi giá cả gần đây nhất đối với kết quả bán hàng – phân tích dạng này cũng có thể được thực hiện trên các dữ liệu marketing và bán hàng thu thập từ nhiều nguồn khác nhau. Khi chính phủ thực thi một số thay đổi về chính sách thuế đối với hàng nhập khẩu, biên lợi nhuận của bạn bị ảnh hưởng ra sao, và khách hàng có phản ứng tiêu cực trước sự gia tăng nhẹ về giá thành sản phẩm không? Trong ví dụ giả định nêu trên, khi phân tích dữ liệu thu thập được từ các nguồn marketing, bạn có thể thiết lập được một số mối tương quan giữa các chiến dịch do đội ngũ marketing thực hiện, doanh thu nhận được từ hoạt động bán hàng, và cảm xúc cũng như sự đánh giá của mọi người về sản phẩm và chiến dịch do đội ngũ marketing thực hiện. Cũng có thể thực hiện phân tích thêm đối với các tập dữ liệu nhận được từ đội ngũ vận hành nhằm xác định những điểm yếu cần khắc phục và tìm ra những cơ hội tiết giảm chi phí – bằng cách này, cuối cùng bạn sẽ cải thiện được khả năng sinh lợi nhuận cho công ty.

Ở giai đoạn này, tranh thủ sự hỗ trợ của đội ngũ là điều hết sức quan trọng, dù bạn có hiểu biết về chuyên môn kỹ thuật và khá tự tin trong mảng này. Vì thế, hãy cùng liệt kê ra tất cả các kỳ vọng về kết quả phân tích hoặc những câu hỏi về các giả định mà bạn muốn tìm câu trả lời – dưới đây là danh sách các chữ A cho công ty E-COMM trong ví dụ giả định của chúng ta dựa trên các nguồn dữ liệu tiềm năng:

- Đâu là những sản phẩm chủ đạo mang lại biên lợi nhuận cao nhất và đâu là những yếu tố thúc đẩy doanh số bán hàng cho các sản phẩm này?
    - Phân tích dữ liệu có cấu trúc.
- Đâu là các sản phẩm phổ biến nhất và ít phổ biến nhất xét theo khối lượng doanh số bán hàng, điểm giá, biên lợi nhuận gộp, và mức độ phổ biến trên mạng xã hội?
    - Phân tích dữ liệu có cấu trúc và phi cấu trúc.
- Đâu là các khu vực địa lý tốt nhất cho chúng ta xét theo tỉ lệ chuyển đổi bán hàng?
    - Phân tích dữ liệu có cấu trúc và phi cấu trúc.
- Chúng ta có thể đưa thêm điều gì vào cách chào bán sản phẩm của mình hay không – chẳng hạn tính năng gợi ý sản phẩm?

- Phân tích dữ liệu có cấu trúc và phi cấu trúc.

- Loại quảng cáo nào và nền tảng nào mang lại kết quả tốt nhất xét về mặt doanh số bán hàng và doanh thu?
  - Phân tích dữ liệu có cấu trúc và phi cấu trúc.
- Chúng ta chi tiêu bao nhiêu cho hoạt động giao sản phẩm tới tay khách hàng, và liệu có thể tiết kiệm được khoản chi phí nào trong quá trình đó không?
  - Phân tích dữ liệu có cấu trúc.
- Thời điểm tốt nhất (tháng/ngày/giờ) cho chúng ta triển khai các chương trình giảm giá hoặc các chiến dịch quảng bá sản phẩm khác là bao giờ?
  - Phân tích dữ liệu có cấu trúc và phi cấu trúc.
- Nhà cung cấp nào có chất lượng phục vụ tốt nhất, sản phẩm chất lượng nhất, và đòi hỏi mức chi phí thấp nhất?
  - Phân tích dữ liệu có cấu trúc.
- Chúng ta dành ra bao nhiêu nguồn lực/thời gian cho các quá trình sau bán hàng cho đến khi giao sản phẩm tới tay khách hàng?
  - Phân tích dữ liệu có cấu trúc và phi cấu trúc.

- Đâu là những câu hỏi phổ biến mà khách hàng hay nêu ra trong các cuộc gọi điện hoặc trò chuyện trực tuyến?
    - Phân tích dữ liệu phi cấu trúc.
- Tỉ lệ chuyển đổi thành khách hàng là bao nhiêu và hiệu suất công việc của các cá nhân trong bộ phận bán hàng/hỗ trợ như thế nào?
    - Phân tích dữ liệu có cấu trúc.
- Những thay đổi trong chính sách thuế của chính phủ sẽ tác động đến tình hình kinh doanh của chúng ta ra sao?
    - Phân tích dữ liệu có cấu trúc.

Trên đây là một số câu hỏi phổ biến và các loại hình phân tích phù hợp với chúng – phân tích dữ liệu có cấu trúc hay phi cấu trúc – mà một doanh nghiệp thực sự có thể vận dụng. Danh sách này có thể rất dài, và đội ngũ kỹ thuật sẽ bắt tay vào quá trình tìm kiếm câu trả lời cho các câu hỏi này bằng cách thực hiện nhiều kỹ thuật phân tích khác nhau trên nhiều tập dữ liệu khác nhau dựa trên các loại hình và kỹ thuật phân tích như đã mô tả ở Chương 4. Cuối cùng, họ sẽ báo cáo lại kết quả phân tích theo cách tốt nhất có thể. Chúng tôi sẽ bàn đến chủ đề này ở phần sau.

5. P (Trình bày và trực quan hóa)

Giai đoạn tiếp theo trong mô hình C-ADAPT là chữ "P", đại diện cho Presentation and visualization (trình bày và trực quan hóa). Ở giai đoạn này trong quá trình phân tích, nhà quản lý phải thấy được các kết quả để từ đó quyết định cách mô phỏng lại thông tin thu thập được cũng như kiến giải đã xác định và phân tích ở các giai đoạn tiếp theo. Một bức hình hơn nghìn lời nói – đôi khi, thông tin này xuất hiện trong hình dạng các hình ảnh trực quan, infographics, và bảng tóm tắt thông tin; đối với các nhà quản lý hoặc lãnh đạo doanh nghiệp, điều này sẽ mang lại lợi ích rất lớn. Đôi khi đội ngũ thực hiện dự án còn phải cung cấp kết quả thu được cho các phòng ban khác cùng tham khảo để họ có thể thực hiện thay đổi theo gợi ý của đội ngũ phân tích dữ liệu và nhà quản lý phụ trách dự án. Dựa trên các phân tích cần thiết và câu hỏi đi cùng, cần lựa chọn ra các cách thức phù hợp và hiệu quả để trình bày chúng. Dưới đây là một số ví dụ minh họa cách trình bày kết quả phân tích cho công ty giả tưởng E-COMM.

Khi phân tích để so sánh hiệu quả hoạt động giữa các kênh bán hàng, bạn có thể chuẩn bị một bảng tóm tắt động để mọi người hiểu được tác động của các chiến dịch và kết quả chúng tạo ra. Hình 5.3 chỉ ra vị trí minh họa hiệu suất hoạt động của một cổng

web dựa trên cơ sở nhiều kênh bán hàng khác nhau và nhiều khu vực địa lý khác nhau:

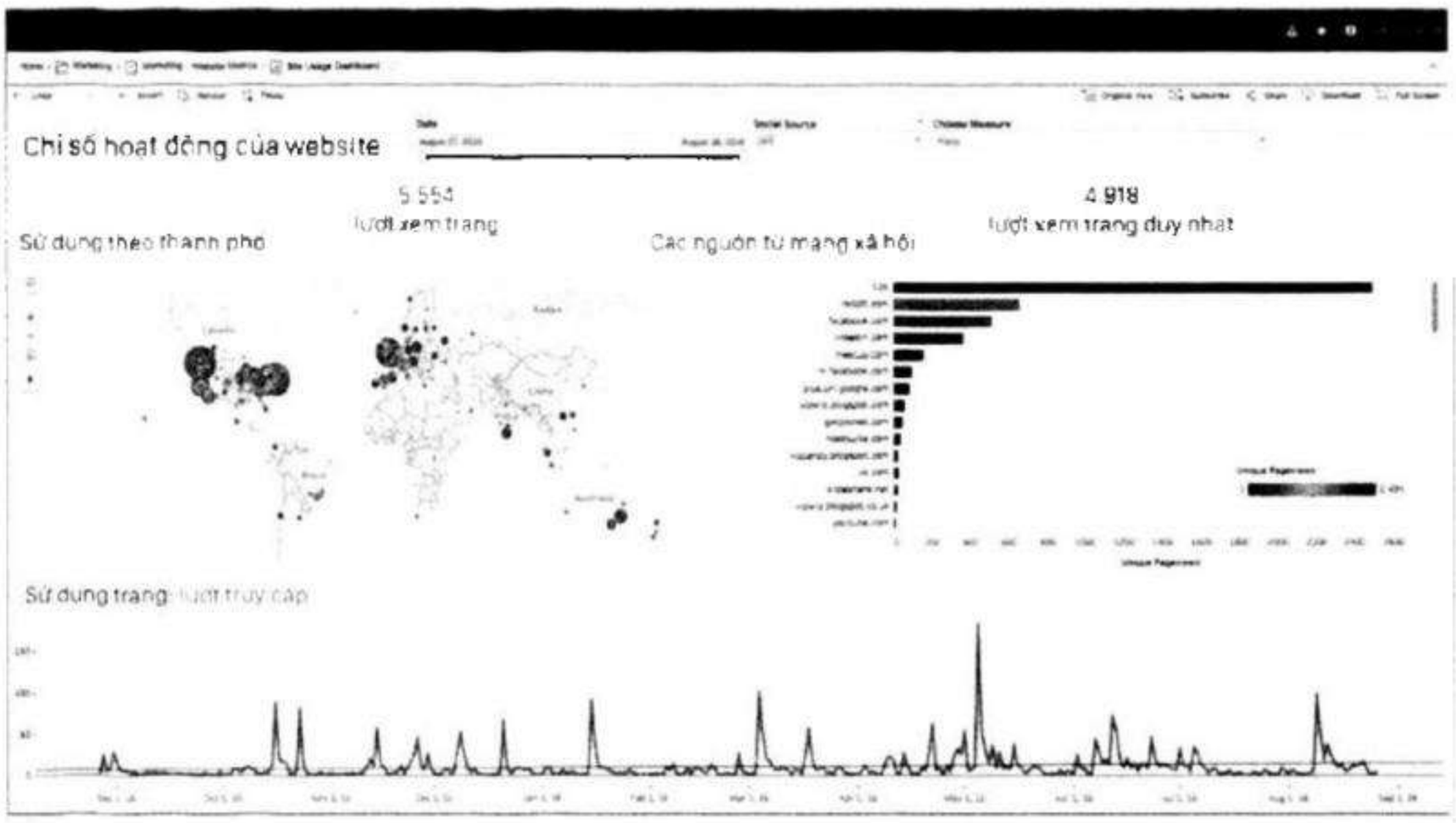
Cũng có thể thực hiện phân tích và biểu diễn hiệu suất hoạt động của từng kênh riêng lẻ nhằm xác định ra cách tốt nhất để tiếp cận với nhiều khách hàng tiềm năng hơn và chuyển họ thành khách hàng – phương pháp này được xây dựng dựa trên tỉ suất ROI, trong đó bạn có thể điều chỉnh việc đầu tư và các chiến lược marketing cho doanh nghiệp. Hình 5.4 là một báo cáo mẫu về kết quả phân tích các hoạt động trên Facebook của một doanh nghiệp.

Sau khi xác định các kênh tạo được nhiều doanh số hơn cho doanh nghiệp, chúng ta có thể tiếp tục minh họa mùa hoạt động tốt nhất cho doanh nghiệp xét từ góc độ doanh thu bán hàng. Một cách hiệu quả để tái hiện thông tin này là sử dụng biểu đồ thanh ngang hoặc biểu đồ đường trình bày thông tin chi tiết về doanh thu bán hàng theo thời gian. Hình 5.5 là một ví dụ mẫu.

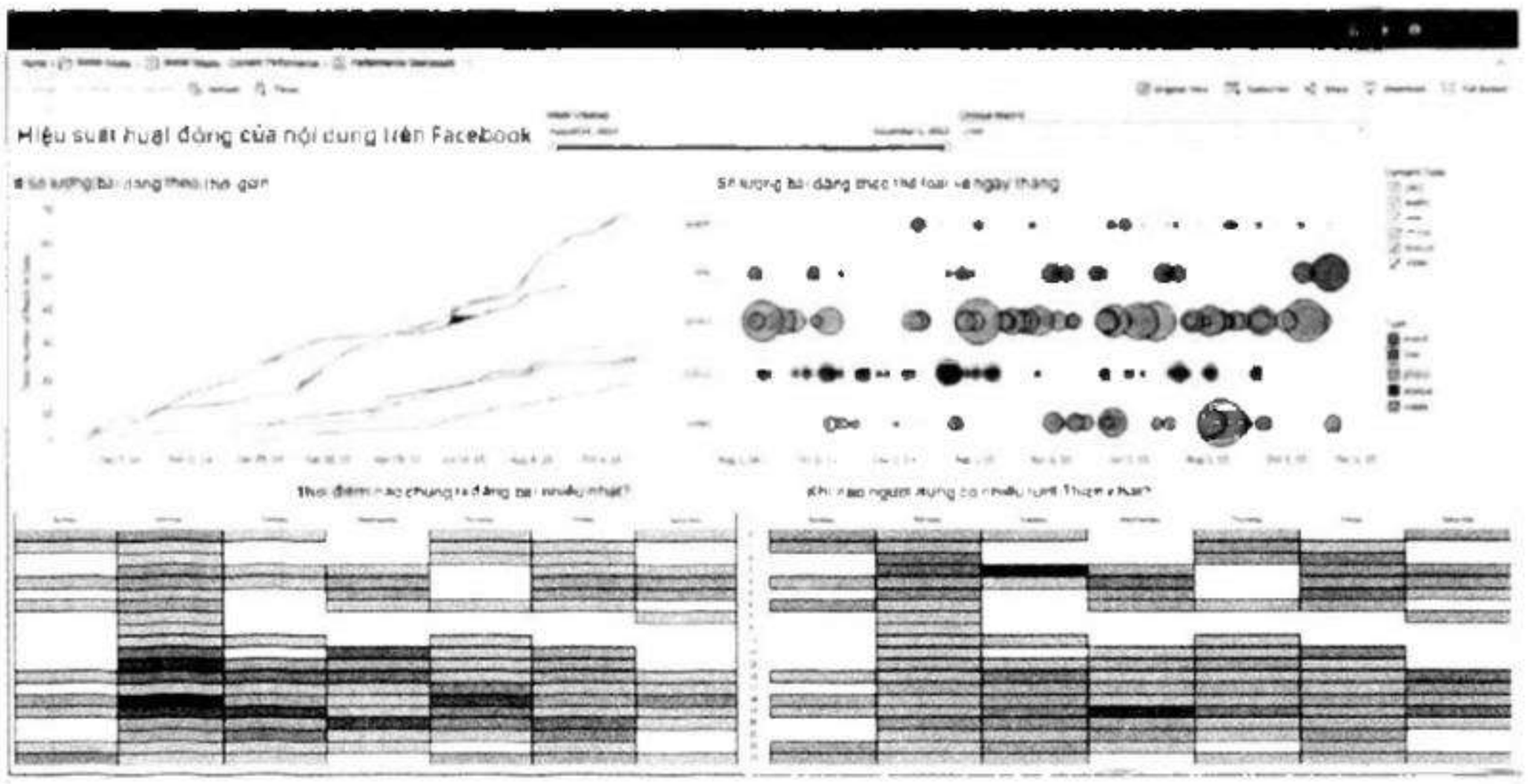
Nếu chúng ta có quyền truy cập vào một hệ thống CRM (chẳng hạn Salesforce), khi đó quá trình phân tích cũng có thể đưa ra được các kết quả phân khúc hợp lý cho sản phẩm và doanh thu bán hàng – ví dụ dưới đây minh họa hiệu quả của các cách chào bán sản phẩm khác nhau trong các quý khác nhau và trong các lĩnh vực ngành nghề và khu vực địa lý khác nhau.

Bạn cũng có thể chuẩn bị một báo cáo tương tự cho E-COMM để hiểu rõ hơn về hiệu quả hoạt động của sản phẩm ở các khu vực và phân khúc khác nhau (xem Hình 5.6).

HÌNH 5.3 Bảng tóm tắt các chỉ số đo lường hoạt động của website



HÌNH 5.4 Bảng tóm tắt hiệu suất hoạt động của nội dung trên Facebook



6. T (Kiểm định giá trị được tạo ra, điều chỉnh, và lặp lại)

Giai đoạn cuối cùng trong mô hình C-ADAPT là chữ "T", đại diện cho cụm từ *testing the value created, refine, and repeat* (kiểm định giá trị được tạo ra, điều chỉnh, và lặp lại). Tuy xếp vị trí sau cùng nhưng trên thực tế, đây mới là giai đoạn quan trọng nhất của mô hình này. Dựa trên những phát hiện mà các bộ phận khác nhau rút ra từ kết quả phân tích và những bài trình bày, các chiến lược hoặc kế hoạch thay đổi sẽ được đưa ra và tác động của chúng sẽ được kiểm định liên tục. Một số hoạt động kiểm định có thể diễn ra ngay trong lúc bạn nhìn vào các báo cáo hoặc bảng tóm tắt được chuẩn bị sau khi hoàn thành khâu phân tích, bởi bạn có thể thực hiện các thay đổi theo thời gian thực đối với các giá trị, từ đó thấy được tác động của chúng ngay trên các đồ thị và bảng biểu. Các công cụ như Tableau mang đến cho nhà quản lý rất nhiều phương án lựa chọn để họ có thể tha hồ mày mò với các loại biểu đồ động và hoạt động theo thời gian thực. Một ví dụ khác về hoạt động kiểm định là việc bạn thực hiện một số thay đổi mỗi khi cần phân tích, có thể bạn sẽ phải chờ mới có được thông tin về các phản ứng hoặc câu trả lời cần thiết. Nếu thực hiện kiểm định thường xuyên và liên tục đưa ra những sự thay đổi cho doanh nghiệp dựa trên đó, nhà quản lý

sẽ đưa ra được những quyết định có căn cứ rõ ràng, từ đó tạo thêm nhiều giá trị cho doanh nghiệp.

Đây là một giai đoạn rất quan trọng của mô hình C-ADAPT này, trong đó nội dung kiểm định không chỉ xoay quanh những thay đổi bạn đã thực hiện hay nhằm tìm hiểu xem nó có củng cố khả năng sinh lợi nhuận hay không (cũng chính là mục tiêu của dự án), mà còn liên quan đến việc kiến tạo giá trị. Trong quá trình tính toán ROIC (tỉ suất lợi nhuận trên vốn đầu tư) để kiểm tra giá trị được tạo ra, bạn cũng nên tính đến những khoản đầu tư dành cho quá trình chuẩn bị, các nguồn lực, và hoạt động triển khai dự án dữ liệu lớn.

Như đã bàn đến ở Chương 3, nếu các kết quả phân tích cho thấy những thay đổi đã thực hiện làm gia tăng doanh thu tổng (và lợi nhuận), thì đây là tin tốt. Nhưng trước khi đưa ra bất kỳ kết luận nào, chúng ta còn phải cân nhắc đến các khoản đầu tư đã bỏ vào dự án dữ liệu lớn, trong đó bao gồm các khoản đầu tư vào cơ sở hạ tầng và các nguồn lực. Cuối cùng, bạn không những phải khôi phục lại chi phí chuẩn bị và triển khai dự án dữ liệu lớn, mà còn phải tạo ra lợi nhuận cao hơn so với kết quả hoạt động của công ty trong giai đoạn trước khi thực hiện dự án này. Như vậy, chúng ta chỉ có thể khẳng định rằng dự án dữ liệu lớn này đã tạo ra giá trị thực sự nếu khả năng sinh lợi

nhuận của công ty lớn hơn trước đây và lớn hơn tổng vốn đầu tư vào quá trình chuẩn bị và triển khai dự án dữ liệu lớn.

Nếu đã quyết định thực hiện một chiến lược marketing, bạn sẽ phải chờ đợi một thời gian và phân tích kết quả trước khi ra quyết định về chiến dịch tiếp theo. Ở giai đoạn này, nhà quản lý cần chuẩn bị một danh sách các kết quả mong muốn, sau đó so sánh chúng với các kết quả đạt được trên thực tế. Như đã đề cập đến ở phần đầu chương này, C-ADAPT là một mô hình lặp lại, dựa trên việc kiểm định kết quả thu được, bạn có thể quyết định đưa vào đó những thay đổi, hoặc điều chỉnh lại các thay đổi trước đó nhằm đạt được kết quả tốt hơn cho các vòng lặp sau. Sau giai đoạn kiểm định, nếu đã có đủ cơ sở để khẳng định rằng chúng ta vừa tạo ra được giá trị thực sự cho tổ chức sau khi đã tính đến tất cả các khoản chi phí rót vào dự án dữ liệu lớn, chúng ta có thể yên tâm chuyển sang giai đoạn tiếp theo – các thách thức và mục tiêu – nhưng ở mỗi vòng lặp, dựa vào kết quả kiểm định, chúng ta lại tiến gần hơn đến các mục tiêu kỳ vọng, mục tiêu điều chỉnh, hoặc mục tiêu mới đặt ra; ở từng giai đoạn, chúng ta phải so sánh giá trị tài chính mà phân tích này tạo ra sau khi giảm chi phí chuẩn bị và vận hành dự án dữ liệu lớn.

Tiếp theo chúng tôi sẽ sử dụng một số dữ liệu giả định để tìm hiểu xem chúng ta có thể tính toán và đi

đến kết luận rằng dự án dữ liệu lớn của công ty giả định E-COMM có tạo ra được giá trị hay không. Có một điều chúng ta phải luôn luôn ghi nhớ khi tính toán lợi nhuận đạt được: bao giờ cũng phải tính đến các khoản đầu tư đã bỏ ra cho dự án này, bao gồm các chi phí chuẩn bị và triển khai dự án, cũng như chi phí thực hiện các thay đổi và gợi ý mà dự án này đưa ra.

Giả sử E-COMM đã đầu tư một số vốn sau đây để khởi động và vận hành vòng lặp đầu tiên của dự án này. Bảng 5.1 cho thấy tổng số vốn đầu tư trong năm năm.

Chúng ta tiếp tục giả định thêm rằng chi phí chuẩn bị ban đầu chỉ bỏ ra một lần và dự án sẽ chạy với nhiều vòng lặp trong suốt năm năm, và mỗi năm chi phí vận hành dự án lại tăng thêm 10%. Do đó, tổng vốn đầu tư cho dự án này sẽ giống như minh họa ở Bảng 5.2.

Bảng 5.2 cho thấy tổng vốn đầu tư trong năm năm là 14,2 triệu đô-la.

Bây giờ, chúng ta giả định rằng với sự hỗ trợ của dự án này và những kiến nghị nó đưa ra, doanh nghiệp sẽ bắt đầu gia tăng doanh số bán hàng và cuối cùng, khả năng sinh lợi nhuận của doanh nghiệp cũng tăng theo – đây chính là mục tiêu tối hậu của dự án này. Nhưng chúng ta cũng không nên quên rằng công ty

này sẽ phải đầu tư 14,2 triệu đô-la chi phí chuẩn bị và vận hành dự án dữ liệu lớn trong năm năm, và dự án này sẽ chỉ tạo ra giá trị nếu công ty kiếm được ít nhất 19,92 triệu đô-la trong năm năm (tính đến cả NPV như đã đề cập ở Chương 3). Bây giờ, chúng tôi sẽ giải thích vì sao công ty lại đạt được 20,31 triệu đô-la. Theo Bảng 5.2, từ năm 1 đến năm 5, công ty sẽ đầu tư tổng cộng một khoản trị giá 14,2 triệu đô-la cho dự án này (bổ sung dòng tổng ở Bảng 5.2). Giả sử khoản đầu tư này giúp họ gia tăng doanh thu từ năm 2, và từ năm 1 tới năm 5, phần doanh thu *gia tăng* của họ lần lượt là 0 đô-la, 1 triệu đô-la, 2,5 triệu đô-la, 5 triệu đô-la, và 10 triệu đô-la, với mức biên lợi nhuận gộp ước tính vào khoảng 40%.

Kết quả trên tương ứng với mức tăng (hoặc giảm) thu nhập hoạt động hằng năm (hay còn gọi là dòng tiền) lần lượt từ năm 1 đến năm 5 là -4 triệu đô-la, -1,2 triệu đô-la, -0,08 triệu đô-la, 2,34 triệu đô-la, và 7,08 triệu đô-la.

Nếu giả sử chi phí vốn là 10%, và tính tỉ lệ khấu hao cho chỉ số NPV (giá trị hiện tại ròng) từ năm 1 đến năm 5 là 10%, thì NPV sẽ là 1,43 triệu đô-la.

Mục tiêu của khoản đầu tư này là gia tăng khả năng sinh lợi nhuận cho công ty, do đó khoản đầu tư ban đầu 14,2 triệu đô-la sẽ phải mang về ít nhất 20,31 triệu đô-la (14,2 x 1,43) để có thể gia tăng khả năng

sinh lợi nhuận. Dựa trên những giả định này, có thể kết luận đây là một dự án nên làm. Chúng ta đang giả định rằng đây là một phân tích ở cấp độ tổng quan, nên bên dưới nó sẽ có một số tiểu mục khác nữa. Như vậy, khả năng sinh lợi nhuận chung thậm chí sẽ còn cao hơn – tuy nhiên, điểm cốt lõi cần rút ra là phải xác định được cách thức phù hợp nhằm đánh giá kết quả thực hiện dự án để đi đến kết luận rằng việc triển khai dự án này sẽ tạo ra giá trị hay phá hủy giá trị.

Mô hình C-ADAPT là một cách tiếp cận rất đơn giản, dễ hiểu, được thiết kế đặc biệt cho các nhà quản lý không có kiến thức chuyên môn về kỹ thuật để họ có thể xác định và triển khai thành công các dự án phân tích dữ liệu lớn. Bên cạnh đó, chúng tôi cũng phát triển một bảng tính để nhà quản lý có thể dùng khi triển khai mô hình C-ADAPT cho dự án của họ đạt được thành công.

BẢNG 5.1 Chi phí dự án dữ liệu lớn
(hằng năm trong 5 năm)

Chi phí cơ sở hạ tầng/chuẩn bị cho dự án dữ liệu lớn (chỉ bỏ ra trong năm 1)	2 triệu đô-la
Chi phí vận hành IT/kho dữ liệu	1 triệu đô-la
Chi phí duy trì đội ngũ chuyên gia phân tích và nhân viên hỗ trợ trong nội bộ gồm 30 người	1 triệu đô-la

BẢNG 5.2 Ví dụ về các khoản đầu tư vào dự án dữ liệu lớn trong 5 năm

Hạng mục	*Năm 1 (triệu đô-la)*	*Năm 2 (triệu đô-la)*	*Năm 3 (triệu đô-la)*	*Năm 4 (triệu đô-la)*	*Năm 5 (triệu đô-la)*	*Năm 1-5 (triệu đô-la)*
						Tổng
Chi phí cơ sở hạ tầng/chuẩn bị cho dự án dữ liệu lớn (chỉ bỏ ra trong năm 1)	2	0	0	0	0	
Chi phí vận hành IT/kho dữ liệu	1	1,1	1,21	1,33	1,46	
Chi phí duy trì đội ngũ chuyên gia phân tích và nhân viên hỗ trợ trong nội bộ gồm 30 người	1	1,1	1,21	1,33	1,46	
Tổng	4	2,2	2,42	2,66	2,92	
Tổng vốn đầu tư trong 5 năm sẽ là 14,2 triệu đô-la						
Tăng doanh thu (triệu đô-la)	0	1,67	4,17	8,33	16,67	**30,84**
Biên lợi nhuận gộp (triệu đô-la)		1	2,5	5	10	
Đầu tư vào dữ liệu lớn (triệu đô-la)	4	2,2	2,42	2,66	2,92	**14,2**
Gia tăng thu nhập hoạt động (dòng tiền, triệu đô-la)	-4	-1,2	0,08	2,34	7,08	
NPV của khoản đầu tư (triệu đô-la)	1,43					

Lưu ý: Mục tiêu của dự án đầu tư này là gia tăng khả năng sinh lợi nhuận cho doanh nghiệp. Với chi phí

vốn 10% và chỉ số NPV là 1,43 triệu đô-la, khả năng sinh lợi nhuận của công ty (chứ không chỉ là lợi nhuận) đã gia tăng. Vì vậy, doanh thu sẽ lớn hơn 20,31 triệu đô-la.

BẢNG TÍNH C-ADAPT

Như chúng ta đã thấy ở phần trước, mô hình C-ADAPT là một phương pháp đơn giản và dễ dàng để lên cấu trúc cũng như triển khai các dự án dữ liệu lớn để tạo ra giá trị cho các tổ chức. Trong mọi giai đoạn của mô hình sáu giai đoạn này, các hoạt động hoặc nhiệm vụ đều được liên hệ với kết quả/đầu ra của giai đoạn trước đó. Ngoài ra, đây còn là một mô hình mang tính lặp lại. Sau khi bạn hoàn tất giai đoạn cuối cùng của mô hình – tức giai đoạn kiểm định giá trị được tạo ra, điều chỉnh, và lặp lại – và bạn hài lòng khi thấy mình đã đạt được hoặc đang tiến gần hơn đến mục tiêu đã vạch ra, và trong quá trình này bạn cũng đã tạo ra được giá trị tích cực cho công ty, khi đó bạn có thể khởi động lại toàn bộ quá trình này từ giai đoạn đầu tiên đến khi hoàn toàn đạt được mục tiêu hoặc hài lòng với kết quả về giá trị mà dự án mang lại.

Trong phần này, chúng tôi sẽ giới thiệu với bạn một công cụ rất thiết thực của mô hình C-ADAPT. Đây là một bảng tính khớp với sáu giai đoạn của mô hình C-ADAPT, các nhiệm vụ con và kết quả tại từng giai đoạn trong mô hình đều được liệt kê trong bảng

tính này. Như có thể thấy ở Hình 5.8, về cơ bản, bảng tính này là tài liệu hướng dẫn mẫu, cũng có thể gọi là một khung, để nhà quản lý sử dụng khi triển khai mô hình C-ADAPT cho các dự án phân tích dữ liệu lớn. Tất cả các phần trong bảng tính này đều có thể khớp vào sáu giai đoạn như đã mô tả về mô hình này ở phần trước.

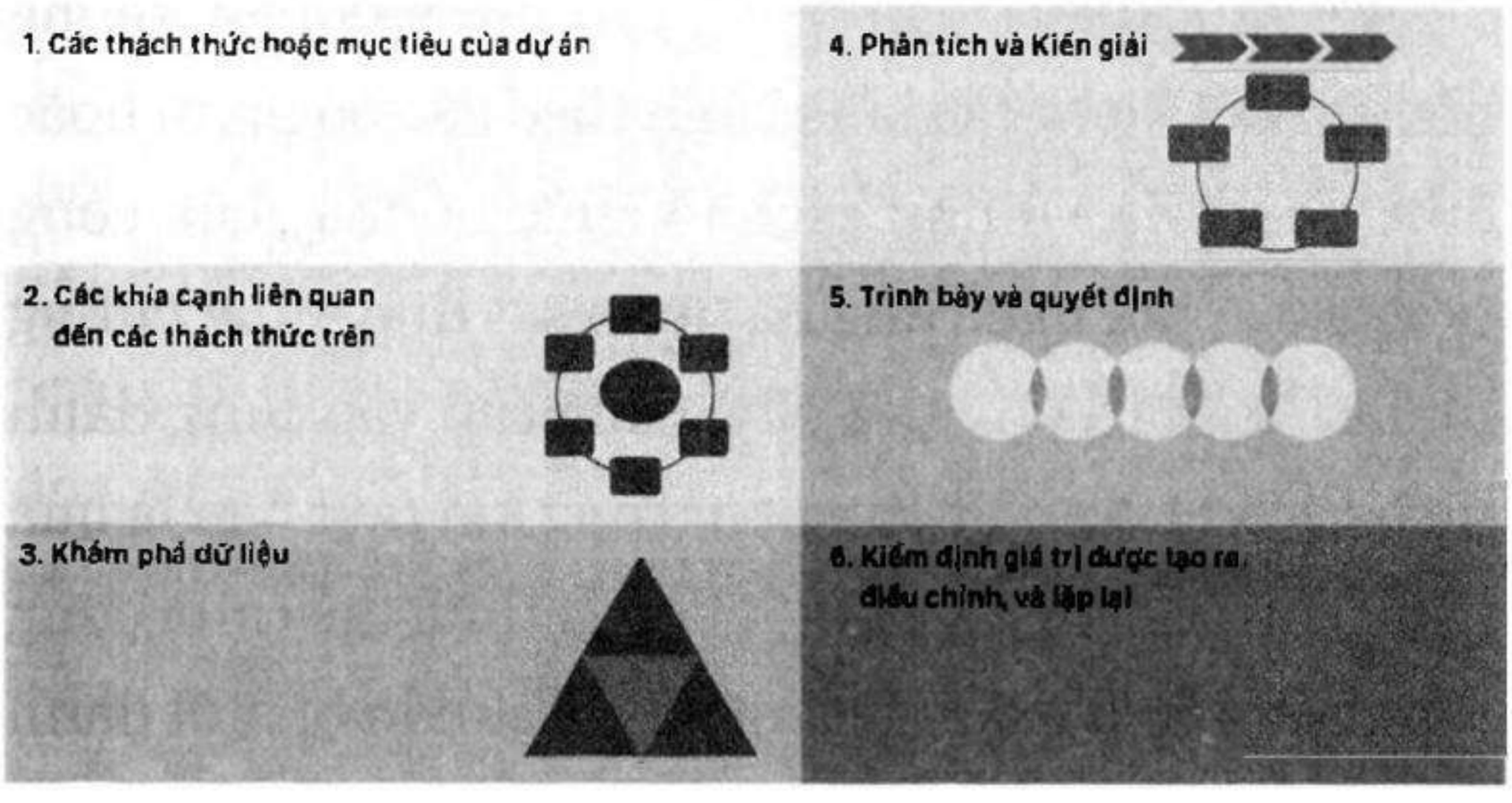
Phần đầu tiên của bảng tính dành chỗ cho nhà quản lý liệt kê tất cả các thách thức hoặc cơ hội và tự trình bày bằng đúng ngôn ngữ của mình. Ở đây, nhà quản lý có thể chỉ cần viết một tuyên bố mục tiêu dài một dòng, hoặc viết một đoạn với nhiều thông tin chi tiết hơn. Ở khâu này, bạn nên làm việc với nhóm của mình để cùng đưa ra ý tưởng và lập ra một danh sách các thách thức và mục tiêu cần nhắm tới, sau đó bạn có thể chia chúng thành các nhóm thách thức khác nhau. Bạn có thể dành thời gian để hoàn thiện quá trình này và cân nhắc đến một vài phương án trước khi chốt danh sách cuối cùng bao gồm các thách thức chính cần giải quyết trong dự án này.

Trong dự án ví dụ cho công ty giả tưởng E-COMM, mục tiêu chính là gia tăng khả năng sinh lợi nhuận cho công ty này – như vậy, ở giai đoạn này, bảng tính của bạn sẽ có một hạng mục ở phần đầu tiên (xem Hình 5.9).

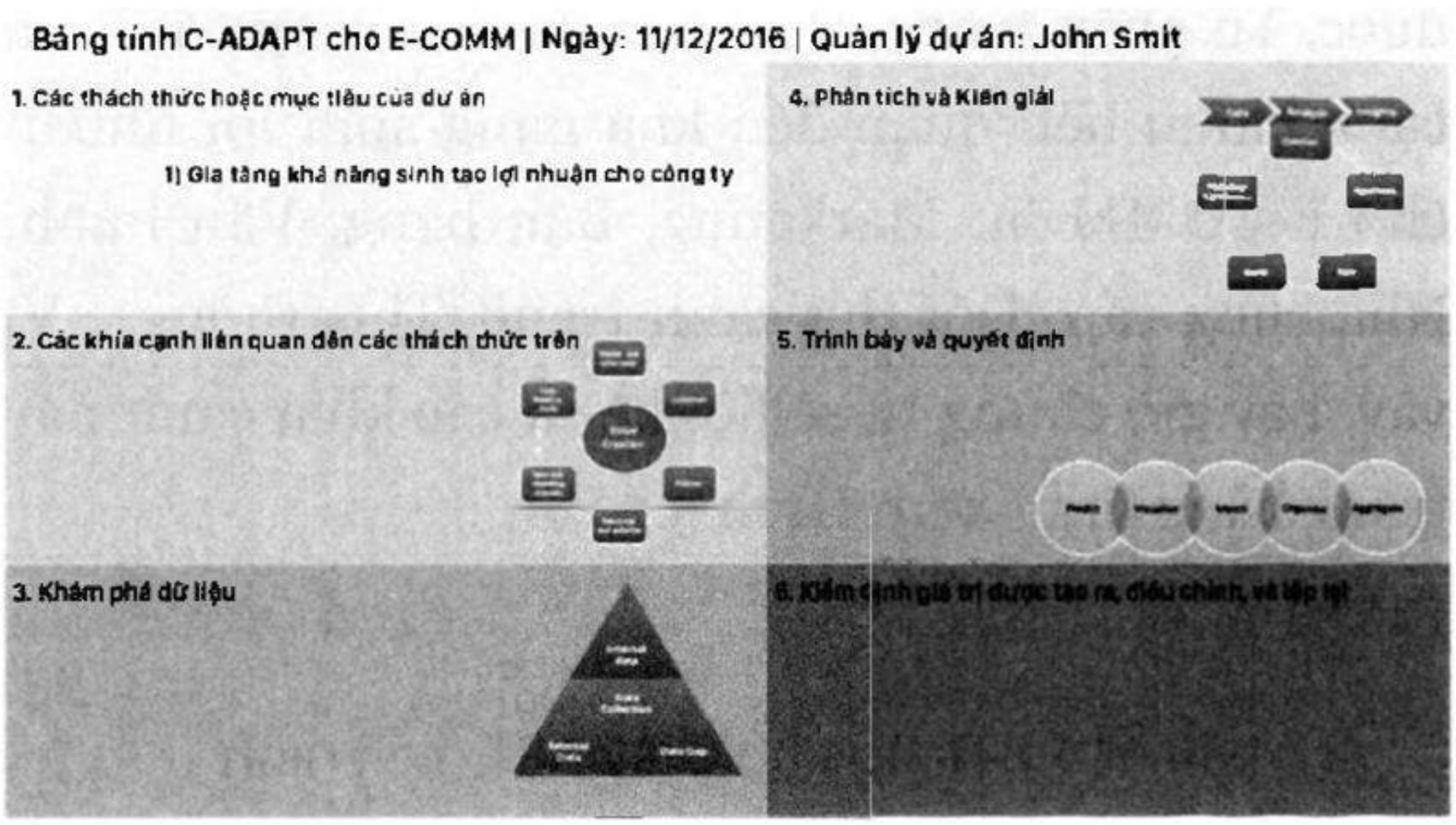
Sau khi đã liệt kê mục tiêu cần hướng đến vào bảng tính, chúng ta sẽ chuyển sang giai đoạn tiếp

theo của mô hình: xác định tất cả các khía cạnh có thể liên quan đến mục tiêu mà chúng ta đang cố gắng đạt được. Từ phần trước, chúng ta đã biết rằng các khía cạnh chính liên quan đến khả năng sinh lợi nhuận cho E-COMM là: Marketing, Bán hàng, Vận hành, Nhà cung cấp, Đối thủ cạnh tranh/Thị trường... Vì vậy, bây giờ chúng ta sẽ liệt kê ra các khía cạnh này trong bảng tính (xem Hình 5.10).

HÌNH 5.8 Bảng tính C-ADAPT điển hình



HÌNH 5.9 Bảng tính C-ADAPT điển hình
(đã hoàn thiện danh sách các thách thức)

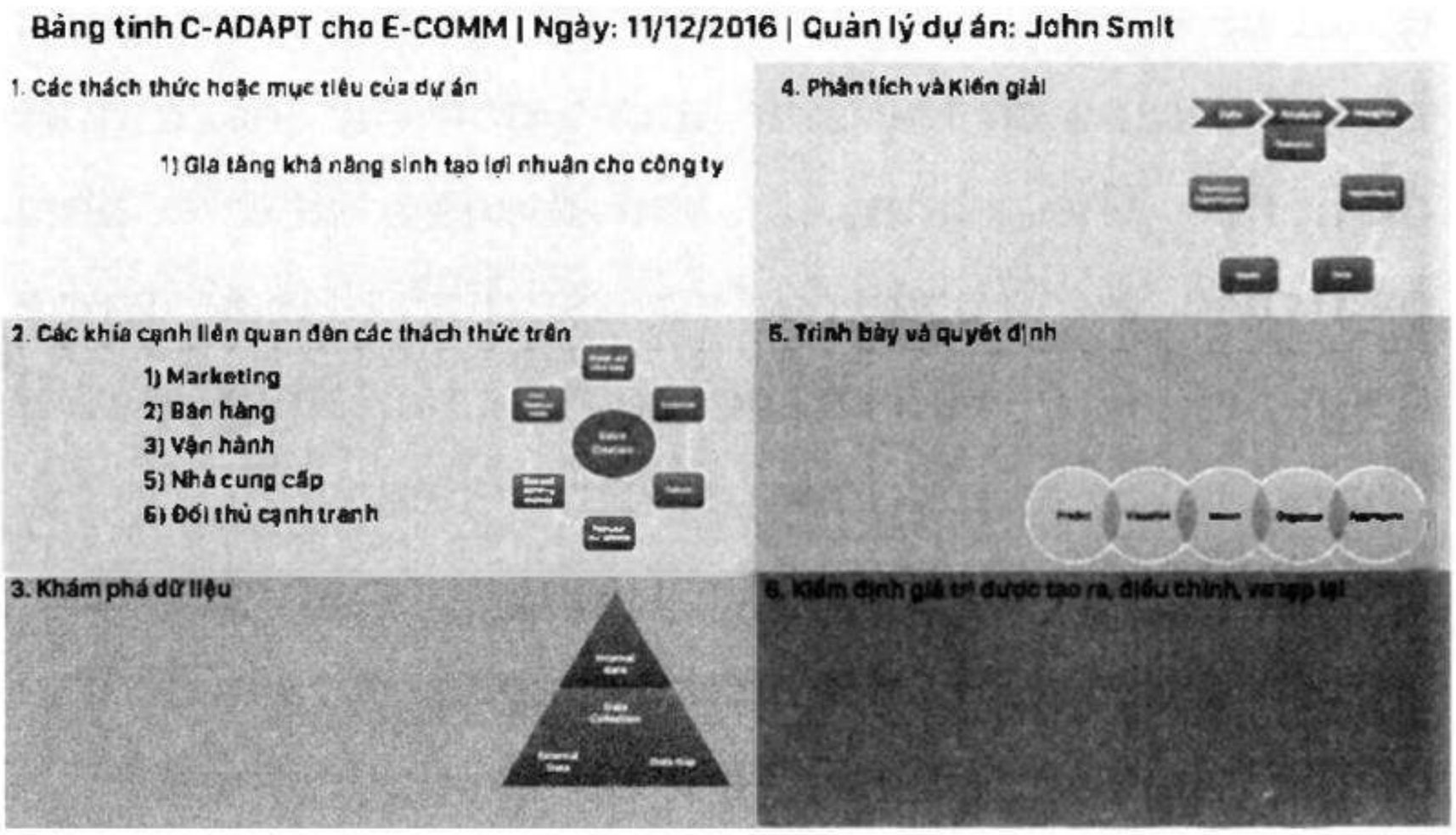


Có một điều quan trọng cần lưu ý là bạn không nên chỉ liệt kê ra các khía cạnh theo các bộ phận hoặc đơn vị nghiệp vụ như cách tổ chức bộ máy của công ty bạn. Có thể bạn cũng đã quan sát thấy rằng chúng tôi liệt kê Marketing và Nhà cung cấp vào cùng danh sách các khía cạnh liên quan trong khi một bên là quá trình (marketing), còn một bên là thực thể (nhà cung cấp). Mục đích của việc làm này là chuẩn bị một danh sách các khía cạnh có thể cho chúng ta biết manh mối về các nguồn dữ liệu cần thiết, để chúng ta có thể chuyển sang giai đoạn tiếp theo. Sau khi quyết định về các thách thức cần giải quyết trong dự án, bạn hãy liệt kê tất cả các khía cạnh/đơn vị/các khía cạnh theo chiều dọc có thể tác động đến những thách thức đã xác định từ đầu quá trình này. Các khía cạnh liệt kê

trong bảng tính của bạn có thể là đối tác, nguồn lực, các hoạt động, mối quan hệ với nhà cung cấp, mối quan hệ với nhân viên, hoặc bất kỳ điều gì khác mà bạn cho rằng có thể ảnh hưởng đến các thách thức đã liệt kê. Theo chúng tôi, bạn không nhất thiết phải bám theo các bộ phận đã được xác định từ trước trong doanh nghiệp – bạn có thể mở rộng tầm nhìn ra khỏi phạm vi của các nhà cung cấp, người mua, nhân viên, các điều kiện thị trường... Tương tự như giai đoạn trước, danh sách này càng dài càng tốt, và sau đó bạn có thể chia chúng thành một vài nhóm tiêu biểu.

Ở giai đoạn tiếp theo, chúng ta phải xác định các dữ liệu liên quan đến các khía cạnh vừa liệt kê; trong một số trường hợp, đây chính là những tập dữ liệu nội bộ có thể tìm thấy dễ dàng, nhưng trong một số trường hợp khác, chúng lại là những tập dữ liệu mà doanh nghiệp đang thiếu hoặc các tập dữ liệu bên ngoài. Sau khi xác định được chúng, chúng ta có thể bảo đảm rằng dữ liệu đã sẵn sàng cho khâu phân tích ở giai đoạn tiếp theo trong mô hình C-ADAPT này. Hãy xem bảng tính cho công ty E-COMM sau khi chúng ta hoàn thành giai đoạn khám phá dữ liệu (Hình 5.11).

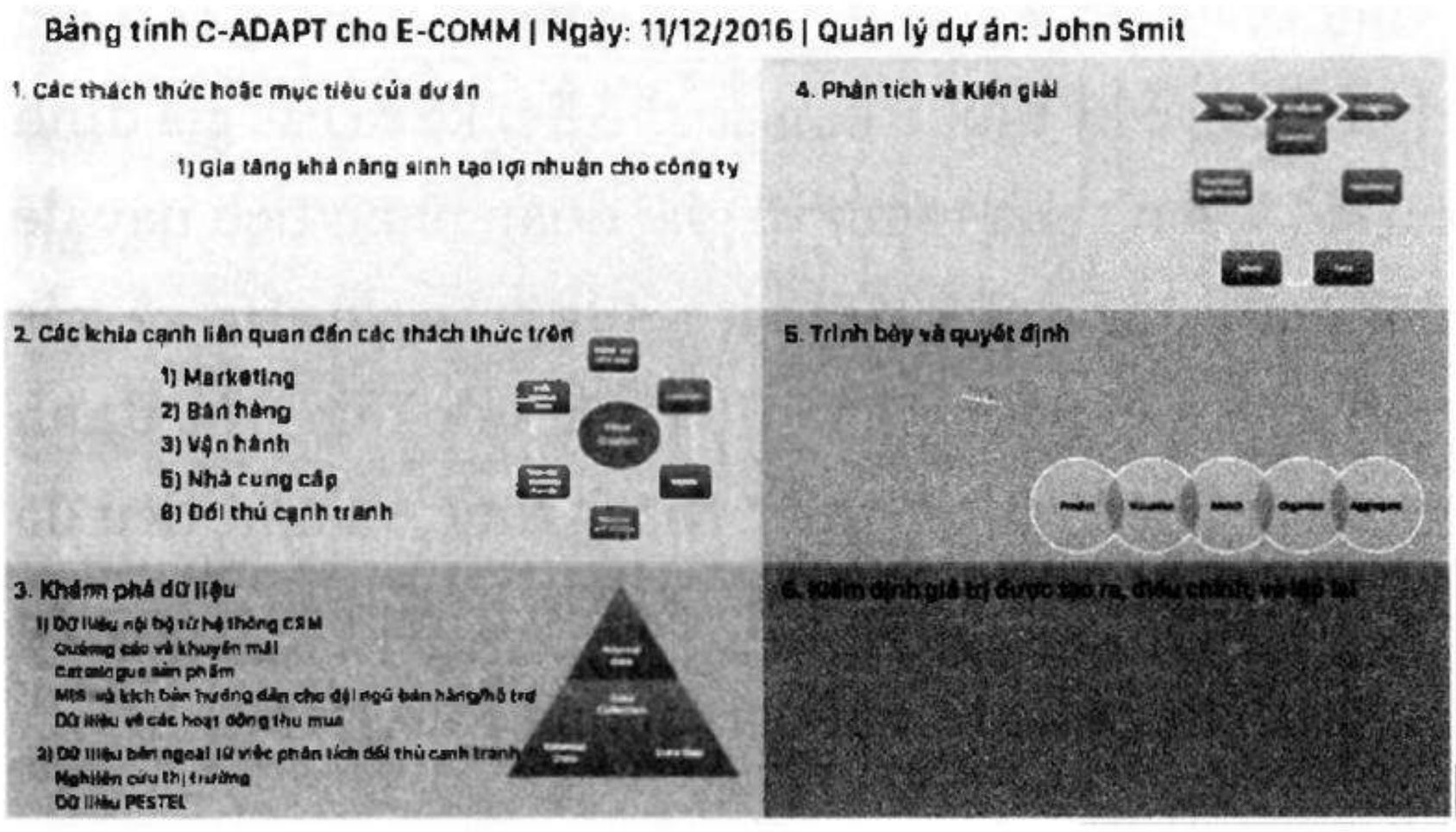
HÌNH 5.10 Bảng tính C-ADAPT điển hình (các khía cạnh đã được hoàn thành)



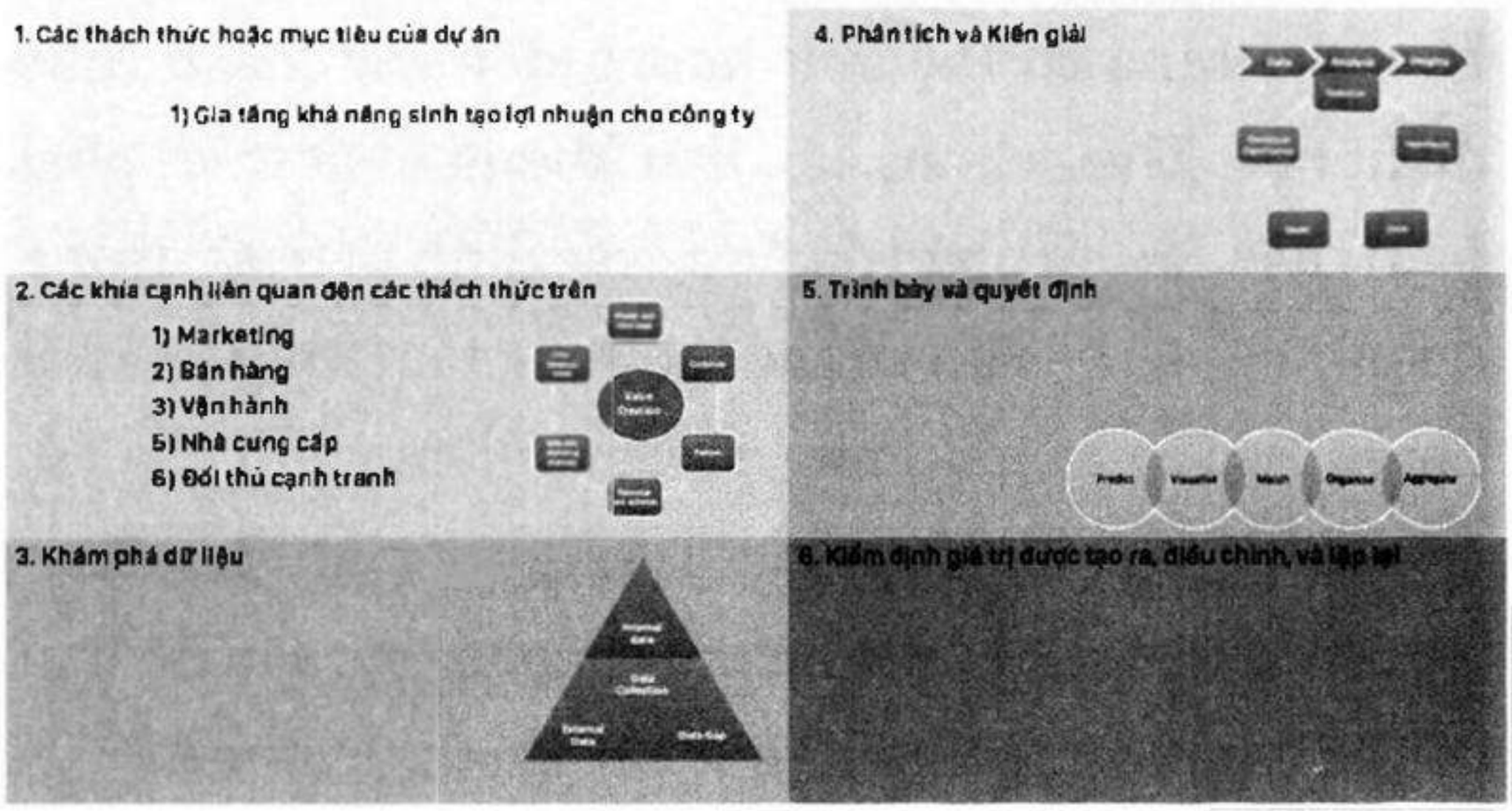
Bảng tính này dành chỗ để bạn liệt kê những dữ liệu liên quan đến các khía cạnh đã xác định ở giai đoạn trước. Xác định tập dữ liệu là một khâu hết sức quan trọng trong dự án bởi bạn phải nghĩ xem dữ liệu nào đã có sẵn trong hệ thống, dữ liệu nào có thể tải về từ công ty, và dữ liệu nào nằm bên ngoài công ty. Như đã nói ở trước, nguồn dữ liệu có thể là nội bộ, hoặc từ bên ngoài. Trên cương vị nhà quản lý, bạn hãy vận dụng tối đa khả năng sáng tạo và sự chủ động của mình để liệt kê ra những tập dữ liệu có thể thu thập và đưa vào sử dụng trong nghiên cứu này. Sự phụ thuộc hoàn toàn vào đội ngũ IT trong giai đoạn này có thể sẽ là điều bất lợi, bởi những thứ như dữ liệu marketing tồn tại bên ngoài công ty có thể đến từ các bảng tin trên mạng xã hội hoặc từ các website

đánh giá sản phẩm, và đội ngũ IT có thể không có thói quen tính đến các tập dữ liệu này trong công việc phân tích dữ liệu hằng ngày của họ. Chúng tôi cũng khuyên nhà quản lý nên viết ra thông tin càng chi tiết càng tốt về các tập dữ liệu trong phần này của bảng tính. Bằng cách làm này, nhà quản lý có thể thiết lập một cách hợp lý mối quan hệ giữa các tập dữ liệu mà không phải đi sâu vào các mô hình thống kê phức tạp hoặc phải tìm hiểu các phép tính thiên về toán học mà nhiều nhà quản lý vẫn tránh động đến. Đây cũng là thời điểm thích hợp để kết nối một vài điểm dữ liệu với nhau dựa trên cơ sở các giả định, và hỏi ý kiến đội ngũ IT xem những giả định đó là đúng hay sai.

HÌNH 5.11 Bảng tính C-ADAPT điển hình (đã hoàn thành khâu dữ liệu)



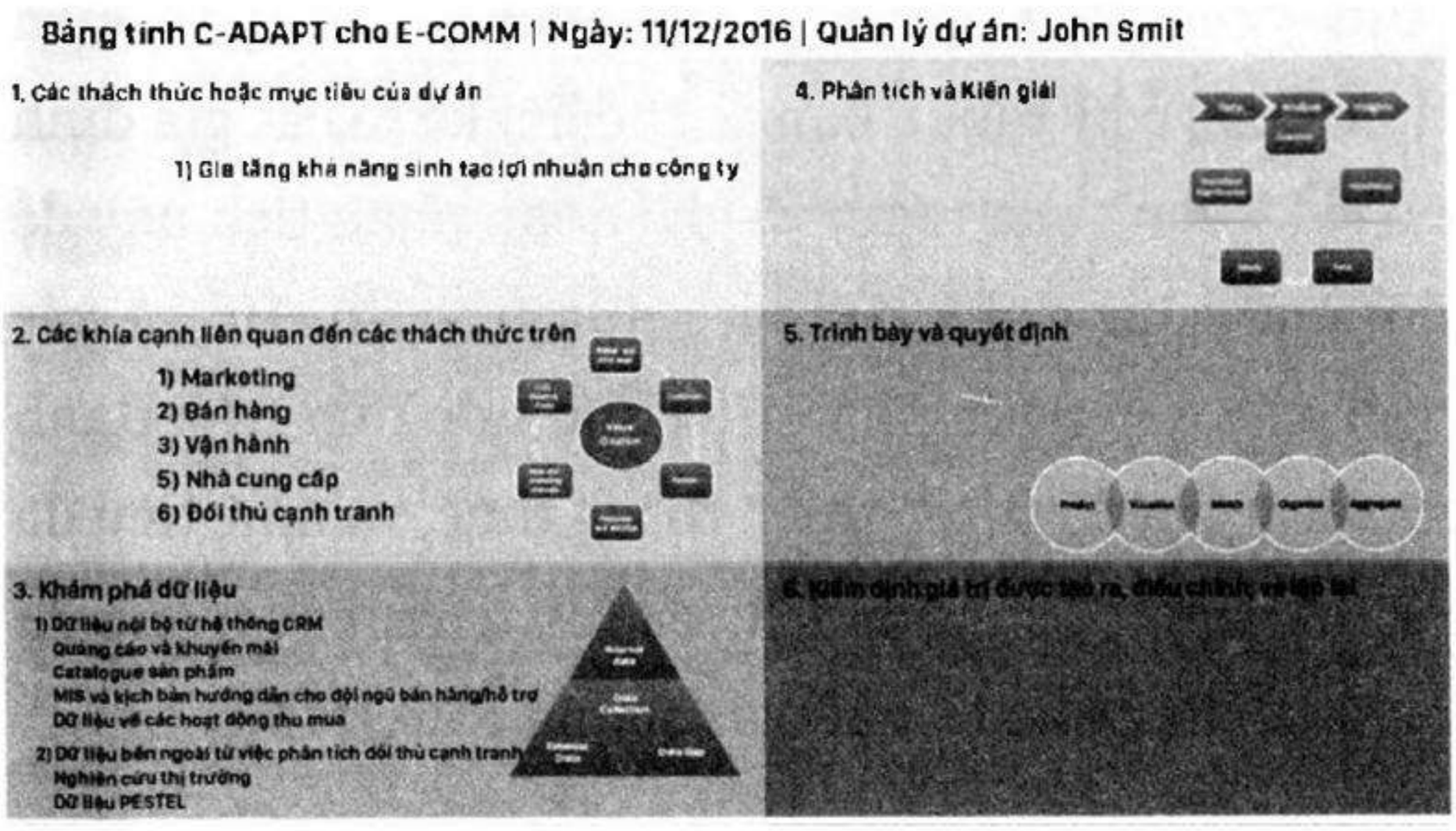
HÌNH 5.10 Bảng tính C-ADAPT điển hình (các khía cạnh đã được hoàn thành)



Bảng tính này dành chỗ để bạn liệt kê những dữ liệu liên quan đến các khía cạnh đã xác định ở giai đoạn trước. Xác định tập dữ liệu là một khâu hết sức quan trọng trong dự án bởi bạn phải nghĩ xem dữ liệu nào đã có sẵn trong hệ thống, dữ liệu nào có thể tải về từ công ty, và dữ liệu nào nằm bên ngoài công ty. Như đã nói ở trước, nguồn dữ liệu có thể là nội bộ, hoặc từ bên ngoài. Trên cương vị nhà quản lý, bạn hãy vận dụng tối đa khả năng sáng tạo và sự chủ động của mình để liệt kê ra những tập dữ liệu có thể thu thập và đưa vào sử dụng trong nghiên cứu này. Sự phụ thuộc hoàn toàn vào đội ngũ IT trong giai đoạn này có thể sẽ là điều bất lợi, bởi những thứ như dữ liệu marketing tồn tại bên ngoài công ty có thể đến từ các bảng tin trên mạng xã hội hoặc từ các website

đánh giá sản phẩm, và đội ngũ IT có thể không có thói quen tính đến các tập dữ liệu này trong công việc phân tích dữ liệu hằng ngày của họ. Chúng tôi cũng khuyên nhà quản lý nên viết ra thông tin càng chi tiết càng tốt về các tập dữ liệu trong phần này của bảng tính. Bằng cách làm này, nhà quản lý có thể thiết lập một cách hợp lý mối quan hệ giữa các tập dữ liệu mà không phải đi sâu vào các mô hình thống kê phức tạp hoặc phải tìm hiểu các phép tính thiên về toán học mà nhiều nhà quản lý vẫn tránh động đến. Đây cũng là thời điểm thích hợp để kết nối một vài điểm dữ liệu với nhau dựa trên cơ sở các giả định, và hỏi ý kiến đội ngũ IT xem những giả định đó là đúng hay sai.

HÌNH 5.11 Bảng tính C-ADAPT điển hình
(đã hoàn thành khâu dữ liệu)

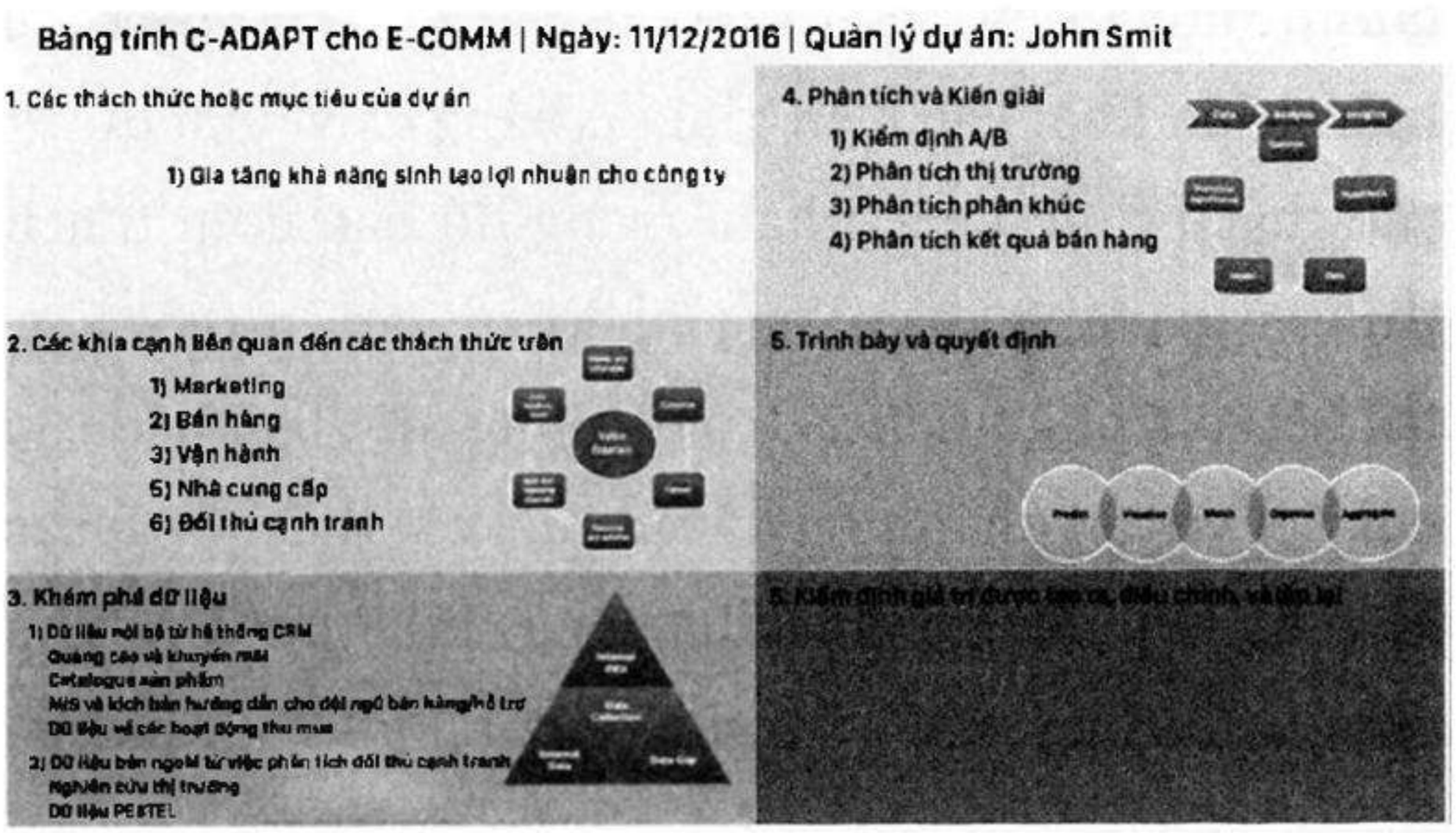


Sau giai đoạn thứ ba, các nguồn dữ liệu đã được xác định, việc thu thập dữ liệu và quy trình ETL đã được triển khai, và dữ liệu đã sẵn sàng bước vào khâu phân tích. Ở giai đoạn này, chúng ta phải liệt kê ra nội dung cần phân tích hoặc các giả định, như vậy việc phân tích mới có thể được thực hiện theo hướng đó. Ở giai đoạn này, các tập dữ liệu đã được xác định từ phần trước và các mối quan hệ giữa chúng có thể được đưa vào phân tích thông qua nhiều phương pháp thống kê khác nhau như đã trình bày trong các phần trước. Đây cũng là khía cạnh mà một số nhà quản lý có thể không cảm thấy tự tin, do đó chúng tôi khuyên bạn nên viện đến sự hỗ trợ từ đội ngũ phân tích công nghệ, bởi họ có kiến thức chuyên môn trong việc thực hiện các công thức thống kê và có khả năng trình bày kết quả theo dạng thức phù hợp nhất, đơn giản nhất, và dễ hiểu nhất. Đối với nhà quản lý không có kiến thức nền về kỹ thuật, bạn có thể liệt kê ra các giả định hoặc kỳ vọng của mình về giai đoạn phân tích này để đội ngũ công nghệ làm việc dựa trên đó. Họ sẽ xác định các quy trình phân tích phù hợp dựa trên danh sách các giả định, giả thiết, thách thức, và mục tiêu đã đề ra. Hình 5.12 minh họa về cách trình bày trong dự án thực hiện cho công ty giả tưởng E-COMM.

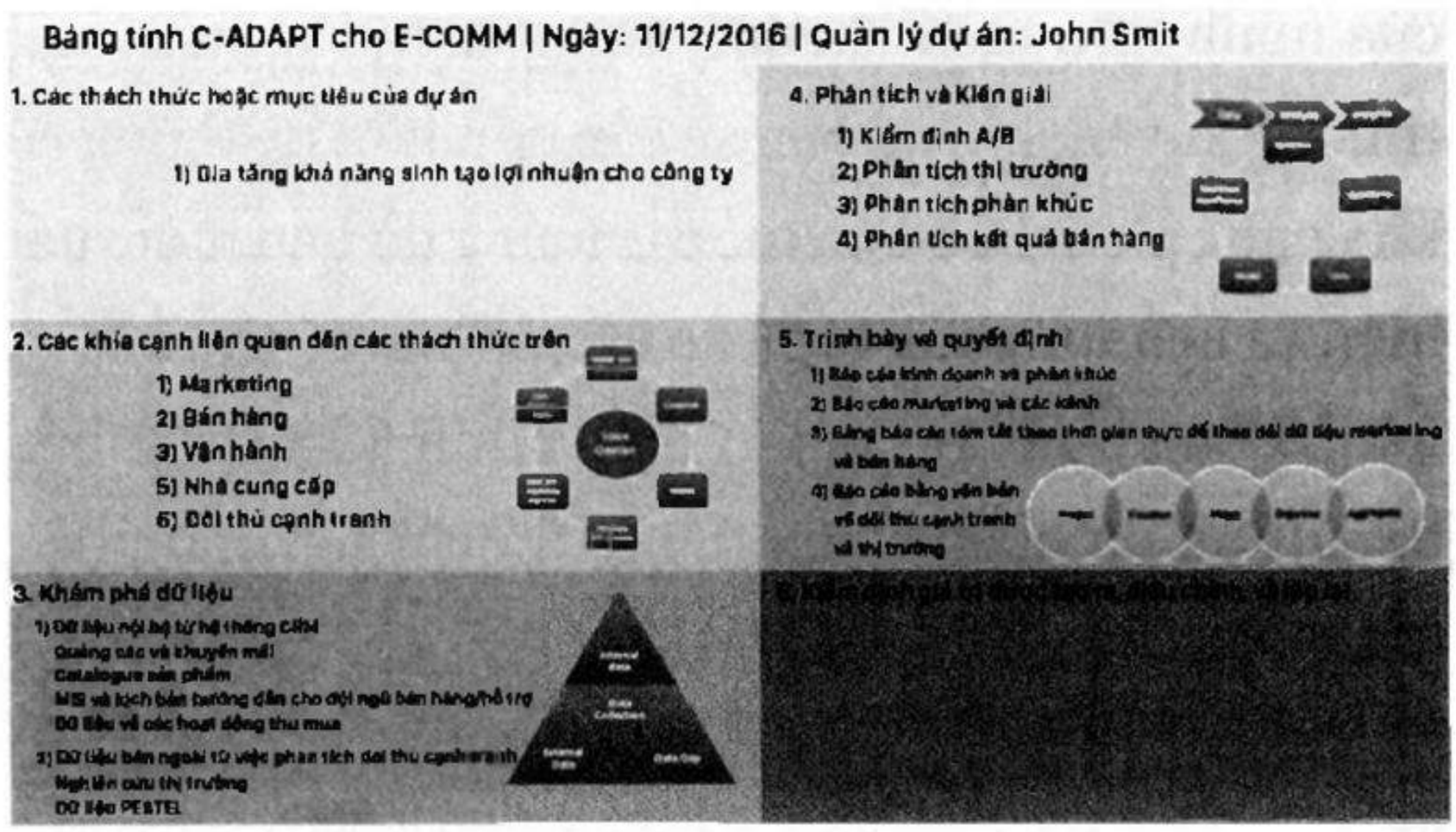
Ô tiếp theo trong bảng tính này dành cho phần trình bày – đây là nơi để bạn và đội ngũ của bạn liệt

kê ra các kết quả mong muốn, và minh họa cách trình bày kết quả đó. Nếu mục đích là xác định các kiến giải chính thông qua phân tích dữ liệu lớn, sau đó thực hiện triển khai các kiến giải đó bằng cách giải quyết các thách thức khác nhau, trong đó bao gồm thách thức về gia tăng khả năng sinh lợi nhuận cho công ty, thì kết quả phân tích sẽ trông giống một bản báo cáo chi tiết về các phát hiện đã rút ra, kèm theo đó là các kiến nghị nhằm gia tăng khả năng sinh lợi nhuận cho công ty. Trong một số trường hợp, bản trình bày có thể trông giống một công trình phân tích và báo cáo theo thời gian thực liên tục hoạt động. Chẳng hạn, nếu bạn muốn biết mọi người đang bàn tán những điều tích cực hay tiêu cực về công ty hoặc thương hiệu của mình trên các diễn đàn mạng xã hội với tần suất thường xuyên, khi đó bạn sẽ cần đến một hệ thống có khả năng phân tích liên tục các luồng dữ liệu mới xuất hiện và liên tục cung cấp báo cáo để bạn có thể đưa ra hành động phù hợp và kịp thời. Hình 5.13 minh họa bảng tính đã hoàn thiện thông tin chi tiết cho phần trình bày của dự án E-COMM.

HÌNH 5.12 Bảng tính C-ADAPT điển hình
(đã hoàn thành phần phân tích)



HÌNH 5.13 Bảng tính C-ADAPT điển hình
(đã hoàn thành phần trình bày và quyết định)



Giai đoạn cuối cùng của mô hình này là kiểm định và lặp lại – liệt kê ra tất cả các kịch bản kiểm định trên cơ sở kết quả phân tích đã thực hiện và

những thay đổi đề xuất thực hiện cho hệ thống hiện hành. Trong một số trường hợp, nếu trong giai đoạn trình bày bạn lập bảng báo cáo tóm tắt động, thì hầu hết các hoạt động kiểm định nội bộ trong hệ thống sẽ diễn ra theo cách thay đổi một hoặc nhiều biến trong bảng báo cáo động đó và xem thay đổi này ảnh hưởng thế nào đến các biến khác. Khi bản báo cáo được trình bày theo nhiều cách thức khác nhau, và các kiến nghị được đưa ra để triển khai thành chiến lược trong tương lai, thì ở giai đoạn kiểm định này, có thể bạn sẽ phải đợi một thời gian mới thu được kết quả của những sự thay đổi đã thực hiện. Chúng ta cũng phải kiểm định xem có phải dự án này đã tạo ra giá trị tuyệt đối cho công ty hay không. Bằng cách thiết lập một quy trình kiểm định vững chắc, bạn có thể đo lường tác động của dự án đối với mọi khía cạnh hoạt động trong doanh nghiệp, từ đó xúc tiến thực hiện các thay đổi cần thiết để cải thiện hiệu suất. Hình 5.14 minh họa bảng tính đã hoàn thiện cho ví dụ về công ty giả tưởng E-COMM, trong đó mục tiêu chính đã được xác định ngay trên đầu, và ở phần cuối là các kiến nghị cho việc thực hiện giai đoạn tiếp theo và thông tin chi tiết về bài kiểm định/vòng lặp.

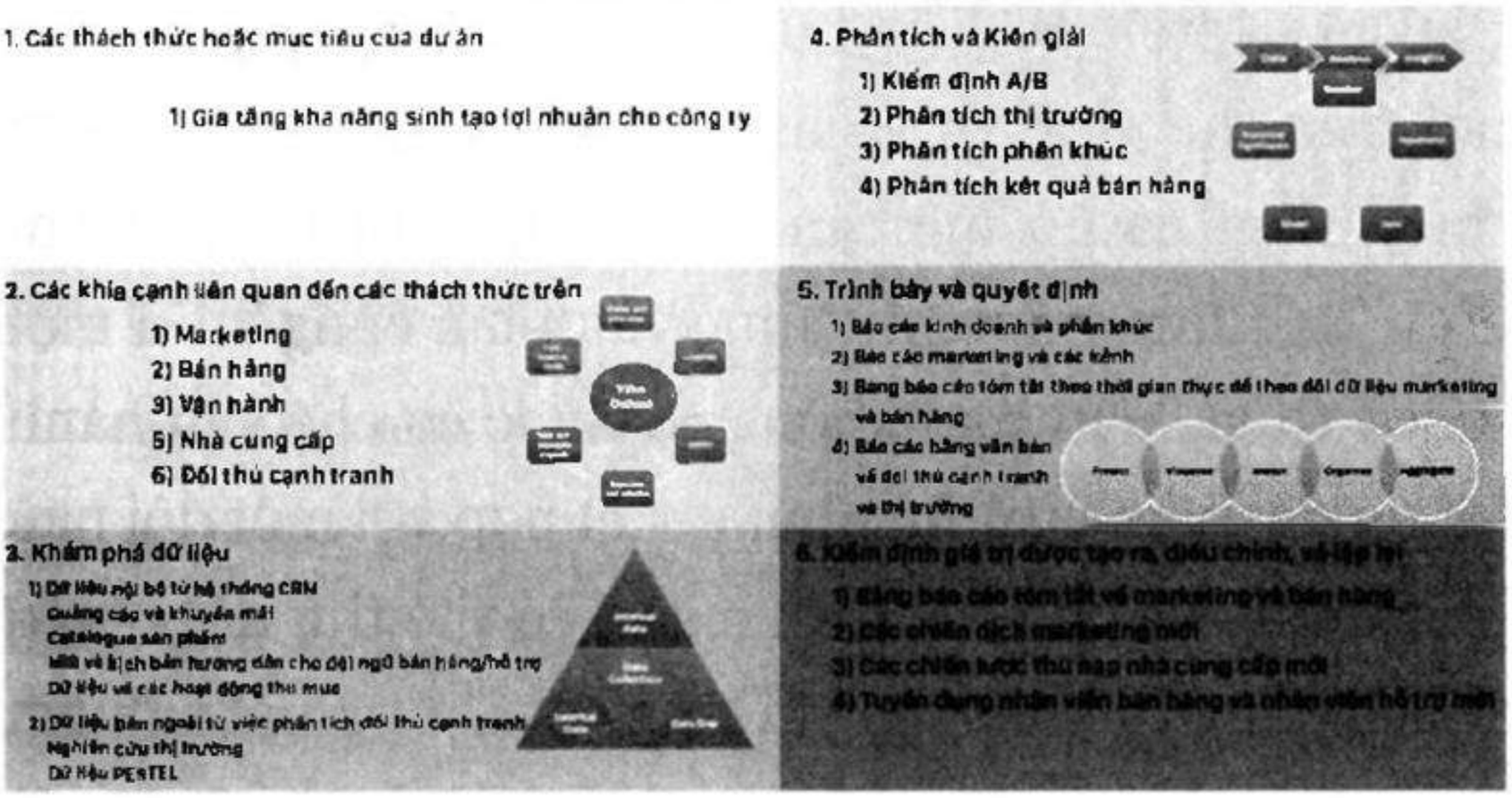
Như đã nói từ trước, C-ADAPT là một mô hình mang tính chất lặp lại, nên các quy trình và phân tích trong đó nên được lặp lại để bảo đảm thu được kết

quả tốt hơn, cải thiện hiệu suất thực hiện ngay tại thời điểm này và trong tương lai.

Bạn có thể tải phiên bản bảng tính C-ADAPT mẫu tại địa chỉ www.creatingvalue.org.uk.

Ví dụ về công ty giả tưởng E-COMM mà chúng tôi sử dụng là công ty hoạt động thiên về công nghệ và kỹ thuật, vì đó là một công ty thương mại điện tử, chuyên kinh doanh các sản phẩm điện tử – có thể một số nhà quản lý sẽ nhầm lẫn cho rằng nếu họ kinh doanh trong một lĩnh vực khác, ít đòi hỏi về công nghệ và kỹ thuật hơn, họ sẽ không thể dễ dàng tận dụng cơ hội vô song mà dữ liệu lớn đang mang lại cho chúng ta. Có một điều quan trọng cần hiểu là tất cả các tổ chức hiện nay đều thu thập rất nhiều dữ liệu và bất kể trong nội bộ tổ chức của bạn có những tập dữ liệu gì, lĩnh vực ngành nghề của bạn là gì, thị trường và nhiều yếu tố khác đều đang sản xuất ra những luồng dữ liệu nằm ngoài phạm vi kiểm soát của bạn. Nhà quản lý cũng nên hiểu rằng không phải dữ liệu nào cũng cần lưu lại và phân tích. Dĩ nhiên, càng nhiều dữ liệu càng tốt, nhưng không phải mọi dữ liệu đều quan trọng, vì thế hãy tiếp cận theo chiều từ trên xuống, trước tiên bạn xác định mục tiêu hoặc thách thức cần nhắm đến giải quyết, rồi thực hiện lần lượt từng bước. Chúng ta đã thấy quá trình hoàn thiện mô hình C-ADAPT từ trên xuống dưới, và sử dụng bảng tính C-ADAPT để thúc đẩy công việc và tạo giá trị từ dữ liệu lớn.

HÌNH 5.14 Bảng tính C-ADAPT tiêu biểu (đã hoàn thiện giai đoạn kiểm định)



Có hai điểm quan trọng cần nhắc lại: (1) Hợp phần chính của phân tích này là đưa vào phần phân tích chính thống nhiệm vụ phân tích các dữ liệu phi cấu trúc và đưa ra kết quả, và (2) Kết quả đưa ra phải đo lường được và mang lại tỉ suất ROI dương – nghĩa là chi phí tổng bao gồm tất cả, từ việc vận hành dự án đến đầu tư vào cơ sở hạ tầng cho dữ liệu lớn, các nguồn lực, triển khai... phải thấp hơn lợi nhuận ròng thu về, nếu không toàn bộ dự án này sẽ thua lỗ xét về lâu dài.

Hãy cùng nhau tìm hiểu một ví dụ nữa về một công ty không hoạt động trong lĩnh vực công nghệ và kỹ thuật như E-COMM, để thấy rằng mô hình và bảng tính C-ADAPT cũng có vai trò quan trọng và mang lại hiệu quả không thua kém gì.

Giả sử bạn là giám đốc vận hành của một chuỗi khách sạn lớn và bận rộn – chúng ta hãy tạm đặt tên cho chuỗi khách sạn giả tưởng này là BigDataHotels. BigDataHotels sở hữu hơn 100 bất động sản ở khắp nơi trên thế giới, và vì kinh doanh khách sạn là một thị trường có áp lực cạnh tranh lớn, nên việc kiếm được lợi nhuận dù chỉ ở mức vừa phải cũng là cả một thách thức. Nếu bạn không có được cơ chế vận hành tối ưu để kiểm soát lợi nhuận, kết hợp với một đội ngũ marketing và bán hàng xuất sắc, có thể thu hút nhiều khách hàng và khiến họ hài lòng với những trải nghiệm tuyệt vời, sẵn sàng quảng bá về thương hiệu của bạn ở nhiều nơi khác nhau, trong đó có mạng xã hội và các cổng website trực tuyến khác, thì bài toán tìm kiếm lợi nhuận sẽ rất nan giải. Ở ngoài đời, BigDataHotels là một chuỗi khách sạn lộng lẫy và nổi tiếng, và trong không gian số, họ cũng năng động và nổi tiếng hệt như vậy. BigDataHotels tiếp nhận rất nhiều dữ liệu từ nhiều nguồn khác nhau (dữ liệu có cấu trúc và phi cấu trúc) và họ liên tục đổi mới cách làm để luôn giữ vững vị thế dẫn đầu ngành theo cách riêng của mình. Tuy vậy, trong sáu tháng qua, một khách sạn trong chuỗi tọa lạc ở Phố Bond, trung tâm thủ đô London đột nhiên sụt giảm số lượng khách đặt phòng, và không ai biết vì sao lại có sự đột biến này. Đây không phải hiệu ứng theo mùa, bởi các khách sạn khác của chuỗi ở khắp nơi trên thế giới, kể cả ở chính London, cũng đều đang hoạt động tốt theo đúng kỳ vọng, chỉ riêng khách

sạn ở Phố Bond lại không đạt được thành tích như mong đợi suốt sáu tháng qua. Tất cả các bộ phận, từ marketing, bán hàng, đến bộ phận dọn dẹp vẫn làm việc như trước đây, và cũng không có sự kiện hay quá trình bên ngoài nào có thể giải thích cho hiện tượng sụt giảm số lượng khách đặt phòng, hay doanh thu bán hàng, đến như vậy.

Bây giờ, trên cương vị giám đốc vận hành, bạn và đội ngũ của bạn chịu trách nhiệm tìm hiểu nguyên nhân sâu xa cho sự sụt giảm doanh số bán hàng và doanh thu này.

Như đã đề cập từ trước, không có bất kỳ dấu hiệu nào cho thấy có sự thay đổi diễn ra trong sáu tháng qua, vì thế, chúng ta hãy sử dụng mô hình C-ADAPT để xử lý thách thức này, xác định nguyên nhân sâu xa của vấn đề, và đưa doanh thu trở lại mức trước kia.

Chúng ta sẽ cùng nhau sử dụng mô hình và bảng tính C-ADAPT để rà soát toàn bộ quy trình, sau đó chúng ta sẽ hoàn thiện phần dữ liệu ở bảng tính cho tất cả các giai đoạn trong mô hình C-ADAPT.

C – Challenges and goals (thách thức và mục tiêu)

Trong ví dụ này đề cập đến hai thách thức chính:

1. Số lượng khách hàng đặt phòng và doanh thu bán hàng nói chung của khách sạn ở Phố Bond,

London thuộc chuỗi khách sạn BigDataHotels bị sụt giảm trong sáu tháng qua, trong khi tất cả các khách sạn khác trong chuỗi đều hoạt động tốt đúng như kỳ vọng. Chúng ta phải tìm ra nguyên nhân sâu xa cho hiện tượng này.

2. Sau khi đã xác định nguyên nhân sâu xa đằng sau xu hướng sụt giảm doanh thu của khách sạn ở Phố Bond, London, chúng ta phải đưa ra các kiến nghị để khôi phục doanh thu và vạch ra chiến lược cho việc đó.

A – Areas related (khoanh vùng)

Dưới đây là các khía cạnh chính ảnh hưởng đến doanh thu và doanh số tổng trong lĩnh vực kinh doanh khách sạn; chúng ta có thể nghĩ rằng sẽ tìm ra được nguyên nhân bằng cách phân tích dữ liệu liên quan đến các khía cạnh này.

1. Doanh số bán hàng: Chúng ta cần tìm hiểu về các quy trình trong lĩnh vực kinh doanh khách sạn, từ marketing, bán hàng, đến hỗ trợ khách hàng sau bán hàng. Tất cả các dữ liệu liên quan đến bán hàng này đều có thể tìm thấy ở bộ phận phụ trách kinh doanh. Chúng ta cũng phải nắm rõ số lượng các kênh bán hàng mà khách sạn đang có và hiệu quả hoạt động của chúng, từ các đối tác và đại

lý liên kết, đến kênh marketing trực tuyến và các công ty lữ hành.

2. Marketing: Từ phân khúc khách hàng đến định vị thương hiệu, marketing đóng một vai trò lớn trong việc thu hút khách hàng đến với chuỗi BigDataHotels, do đó dữ liệu marketing cũng sẽ có những đóng góp lớn trong dự án này. Bộ phận phụ trách marketing có thể cung cấp cho chúng ta dữ liệu liên quan đến các chiến dịch marketing, các sản phẩm dành cho khách hàng, và các cung đường tiếp cận thị trường mà chuỗi này đã và đang thiết lập cũng như sử dụng trong suốt những năm qua, trong đó bao gồm cả hoạt động marketing trong không gian số và trên mạng xã hội.

3. Dịch vụ khách hàng: Doanh thu của ngành du lịch và lữ hành chịu tác động lớn từ các trải nghiệm của khách hàng, từ thời điểm họ lựa chọn khách sạn để đặt phòng đến thời điểm họ trả phòng và sau đó quyết định chia sẻ về những trải nghiệm trong thời gian ở khách sạn với bạn bè, gia đình, hoặc cộng đồng lớn xung quanh. Vì thế, chúng ta cũng phải đưa vào phần phân tích của dự án này các dữ liệu về hành trình khách hàng và dịch vụ khách hàng. Chúng ta có thể lấy dữ liệu về hỗ trợ bán hàng và dịch vụ từ bộ phận dịch vụ khách hàng.

4. Vận hành: Trong phần lớn các trường hợp, vận hành ảnh hưởng đến lợi nhuận của doanh nghiệp và khía cạnh này thường bị bỏ qua trong những phân tích về lợi nhuận hoặc doanh số gộp – nhưng trong một số lĩnh vực ngành nghề như du lịch, những gì xảy ra sau bán hàng có thể tác động đến doanh thu trong tương lai. Đặc biệt trong lĩnh vực du lịch, quy trình nhịp nhàng và tối ưu có thể khiến khách hàng hài lòng, từ đó khiến họ để lại những ý kiến phản hồi tích cực, đóng vai trò quan trọng trong doanh số bán hàng tương lai của doanh nghiệp. Vì vậy, chúng ta cũng sẽ tính đến các dữ liệu về vận hành trong phần phân tích.

5. Tình hình cạnh tranh và thị trường: Doanh thu bán hàng và doanh số cũng là một đặc điểm chịu ảnh hưởng bởi hiệu quả hoạt động của đối thủ cạnh tranh. Nếu đối thủ cạnh tranh gần nhất của bạn cung cấp chương trình giảm giá mạnh hoặc phục vụ các phòng có chất lượng tốt hơn, khi đó bạn sẽ thấy doanh số của mình chịu ảnh hưởng bởi động thái này, trừ phi bạn đã và đang có những biện pháp bảo vệ tốt cho vị thế của mình, hoặc không ngừng cạnh tranh so kè từng đồng với đối thủ. Ngoài đối thủ truyền thống như các khách sạn khác, bạn cũng phải cân nhắc đến các công ty cung cấp giải pháp và dịch vụ thay

thế như Airbnb, bởi họ có thể thu hút một phần khách hàng của bạn. Các nghiên cứu về thị trường và tình hình cạnh tranh có thể mang lại rất nhiều thông tin có ý nghĩa ở khía cạnh này cho dự án của chúng ta.

6. Các khía cạnh mở khác: Tất cả các khía cạnh liệt kê ở trên có thể có mối quan hệ trực tiếp với hiệu quả hoạt động của doanh nghiệp và trong hầu hết các trường hợp, bạn sẽ phát hiện ra rằng những dữ liệu này đều đang nằm trong tổ chức. Tình hình chính trị, kinh tế, và môi trường có thể tác động đến hiệu quả kinh doanh và bạn cũng nên đưa những điểm này vào danh sách các khía cạnh cần thu thập dữ liệu.

D – Data discovery (khám phá dữ liệu)

Như vậy, chúng ta đã xác định được danh sách các khía cạnh có thể có liên quan và tác động đến doanh số bán hàng của khách sạn tọa lạc trên Phố Bond, London thuộc chuỗi khách sạn BigDataHotels – đây là một thách thức mà chúng ta đã liệt kê ở phần đầu tiên của mô hình này. Bây giờ, chúng ta hãy chuẩn bị một danh sách các tập dữ liệu và nguồn dữ liệu có thể sử dụng cho các khía cạnh vừa được khoanh vùng và có thể đưa vào trong khâu phân tích ở các phần tiếp theo. Một điểm quan trọng nữa cần lưu ý là: Nếu hiệu

suất kinh doanh bị ảnh hưởng trong sáu tháng qua, vậy chúng ta nên sử dụng bao nhiêu dữ liệu là đủ để phân tích? Do có thể có độ trễ nhất định trong việc cảm nhận được tác động thực sự của bất kỳ sự kiện nào nên chúng ta hãy tính đến dữ liệu trong vòng 12 tháng của tất cả các nguồn như liệt kê dưới đây:

1. Dữ liệu nội bộ: Đây là danh sách các tập dữ liệu có thể lấy được từ bên trong tổ chức và có thể liên quan đến các phòng ban nội bộ hoặc các sự kiện. Dữ liệu loại này tương đối dễ tiếp cận, và thực ra, biết đâu đội ngũ kỹ thuật trong công ty bạn cũng rất quen thuộc với các dữ liệu này, thậm chí có thể còn đang sử dụng một số tập dữ liệu trong đó.
    a. Dữ liệu CRM (dữ liệu có cấu trúc và phi cấu trúc): Dữ liệu đến từ hệ thống CRM sẽ cung cấp cho bạn thông tin chi tiết về khách hàng, tình hình kinh doanh, doanh số, đặc điểm về vị trí địa lý của khách hàng... Nhìn chung, bạn có thể thu thập dữ liệu đầy đủ về hồ sơ khách hàng từ hệ thống CRM. Ngoài ra, chúng ta cũng có thể truy cập kho dữ liệu chứa đựng các thông tin chi tiết về tình hình kinh doanh, hiệu quả hoạt động của từng kênh bán hàng, và tác động của các chiến dịch marketing đối với doanh số.
    b. Dữ liệu về các chiến dịch marketing (dữ liệu phi cấu trúc): Bạn cần nắm được những chiến

dịch nào đang mang lại nhiều lợi ích cho công ty, và mỗi chiến dịch mang về bao nhiêu doanh thu. Dữ liệu đến từ các cổng quảng cáo và các hãng quảng cáo có thể giúp bạn thu được những kiến giải thấu đáo hơn. Ngày nay, khách hàng có thói quen đặt phòng khách sạn qua mạng với máy tính xách tay hoặc điện thoại di động – doanh nghiệp có thể thu thập loại hình dữ liệu này bằng cách sử dụng các công cụ đo lường web như Google Analytics và Google Ads.

c. Dữ liệu về hoạt động chăm sóc khách hàng (dữ liệu có cấu trúc và phi cấu trúc): Hầu hết các khách sạn đều áp dụng những quy trình rõ ràng về hỗ trợ khách hàng và dịch vụ khách hàng, theo đó các cuộc trao đổi giữa nhân viên khách sạn với khách hàng đều được ghi âm lại và đưa vào lưu trữ trong cơ sở dữ liệu. Bạn có thể tiếp cận những thông tin này, chẳng hạn số lượng các cuộc gọi, thường là với nội dung than phiền, thông tin báo cáo về cách giải quyết vấn đề của nhân viên... Qua đó, bạn có thể diễn giải các dữ liệu này để tìm hiểu xem liệu có bất kỳ vấn đề gì đang tồn tại trong chuỗi khách sạn này hay không, và liệu đội ngũ nhân viên phục vụ có làm việc tốt đúng

như kỳ vọng hay không. Bạn cũng có thể tiếp cận được kho dữ liệu lưu trữ các ý kiến phản hồi của khách hàng vốn được cung cấp thường xuyên và lưu trữ trong hệ thống nội bộ.

d. Dữ liệu về nhà cung cấp và các hoạt động thu mua (dữ liệu có cấu trúc và phi cấu trúc): Loại dữ liệu này sẽ được lưu ở các hệ thống ERP (hoạch định chiến lược doanh nghiệp) trong các tổ chức; một phần lớn dữ liệu ở đây liên quan trực tiếp tới các quy trình vận hành và mối quan hệ với các nhà cung cấp của nhiều loại sản phẩm và dịch vụ mà doanh nghiệp sử dụng. Chúng ta có thể biết quy trình chào đón khách hàng là gì, chúng ta cung cấp những món đồ hoặc vật dụng nào trong phòng của họ, những món đồ đó xuất phát từ đâu, và chi phí chúng ta bỏ ra để có chúng là bao nhiêu? Các nhà cung cấp này đã làm việc với chúng ta được bao lâu rồi... Loại thông tin này có thể được lấy từ kho dữ liệu của bộ phận thu mua.

2. Dữ liệu bên ngoài: Có nhiều nguồn dữ liệu tồn tại bên ngoài tổ chức và chúng cũng tạo ra những dữ liệu phù hợp với công ty và lĩnh vực hoạt động của bạn. Một số dữ liệu thuộc nhóm này do các bộ phận phòng ban trong chính công ty bạn thu thập và lưu trữ, và một số khác lại cần được thu thập

riêng cho dự án mà chúng ta đang thực hiện. Dưới đây là danh sách các tập dữ liệu bên ngoài mà chúng ta nên cân nhắc đưa vào khâu phân tích:

a. Dữ liệu mạng xã hội (dữ liệu phi cấu trúc): Có thể chuỗi BigDataHotels hiện đã tiến hành theo dõi và lưu trữ dữ liệu trên mạng xã hội cũng như dữ liệu liên quan đến hoạt động marketing trong không gian số thông qua các công cụ như Hootsuite hoặc Odimax. Trong trường hợp chưa có sẵn các dữ liệu này trong tổ chức, chúng ta cần thu thập những dữ liệu cần thiết từ các nguồn mạng xã hội và lưu trữ chúng trong hệ thống dữ liệu nội bộ để thực hiện phân tích trong các bước sau. Chúng ta có thể theo dõi số lượt nhắc đến tên thương hiệu BigDataHotels, đến các đối thủ cạnh tranh, và các từ khóa liên quan khác. Chúng ta có thể tiếp cận dữ liệu về nội dung mà khách hàng bình luận về dịch vụ của chúng ta trên các cổng mạng xã hội như Facebook và Tripadvisor sau khi họ rời khách sạn. Trong phần lớn các trường hợp, những ý kiến phản hồi như vậy lại minh bạch và có giá trị cao hơn so với phiếu đánh giá dịch vụ mà nhân viên khách sạn thu thập từ khách hàng, bởi trên mạng xã hội hoặc trên các cổng web khác

như Tripadvisor, khách hàng được tự do viết và chia sẻ về trải nghiệm của mình mà không cảm thấy họ có nghĩa vụ gì với nhân viên đang trực tiếp lấy ý kiến của họ ở quầy lễ tân khi họ làm thủ tục trả phòng.

b. Nghiên cứu thị trường (dữ liệu có cấu trúc và phi cấu trúc): Với sự hỗ trợ của các báo cáo nghiên cứu thị trường độc lập, và thông tin chi tiết về các đối thủ cạnh tranh và các doanh nghiệp kinh doanh sản phẩm/dịch vụ hỗ trợ hoặc bổ sung cho sản phẩm/dịch vụ của bạn, bạn có thể hiểu được tác động kinh doanh từ góc nhìn của người ngoài cuộc. Dữ liệu mạng xã hội thu thập về đối thủ cạnh tranh và các công ty khác trên thị trường cũng có thể bổ sung thêm giá trị vào khâu phân tích này. Bạn có thể so sánh ý kiến đánh giá/phản hồi của mọi người về công ty bạn và về công ty đối thủ để có thêm nhiều kiến giải sâu sắc hơn.

c. Dữ liệu PESTEL (dữ liệu có cấu trúc và phi cấu trúc): Chúng ta đã cùng nhau bàn về phân tích PESTEL ở phần trước. Phải tiếp cận được các dữ liệu, chi tiết, và thông tin về những thay đổi trong môi trường chính trị, kinh tế, xã hội, công nghệ, và pháp luật, đồng thời phải nắm được tác động của chúng đối với

công việc kinh doanh của mình. Loại dữ liệu này cũng có thể được thu thập thông qua các công cụ theo dõi mạng xã hội, các báo cáo thị trường được cung cấp miễn phí trên mạng, và nhiều nguồn khác.

A – Analysis and insights (phân tích và kiến giải)

Chúng ta vừa liệt kê ra tất cả các khía cạnh liên quan đến thách thức và mục tiêu đặt ra cho dự án này; chúng ta cũng đã xác định được tất cả các nguồn dữ liệu có thể cần đến từ các nguồn nội bộ và bên ngoài. Bây giờ, chúng ta nên bắt tay vào chuẩn bị cho giai đoạn phân tích.

Ở giai đoạn này, có thể bạn sẽ phải làm việc với đội ngũ công nghệ để họ thực hiện khâu ETL, làm sạch, và lưu trữ tất cả các loại hình dữ liệu có cấu trúc, bán cấu trúc, và phi cấu trúc trong một dạng thức phù hợp để tiến hành các kỹ thuật phân tích. Sau khi khâu chuẩn bị đã hoàn tất và dữ liệu đã sẵn sàng để sử dụng và phân tích, lúc này chúng ta sẽ phải xác định loại phân tích cần thực hiện và các câu hỏi mà từng loại phân tích có thể trả lời. Như cách làm ở ví dụ trước (về công ty giả tưởng E-COMM), nếu không quen làm việc về mặt chuyên môn kỹ thuật, bạn có thể lập ra một danh sách các câu hỏi mà bạn muốn có

được câu trả lời, để dựa vào đó đội ngũ kỹ thuật có thể tiến hành cài đặt để chạy phân tích và trình bày kết quả phân tích theo dạng thức phù hợp.

Dưới đây là các câu hỏi chính mà chúng ta cần tìm kiếm câu trả lời từ phân tích này:

1) Tất cả các kênh bán hàng hiện đang hoạt động ra sao?
2) Các phân khúc khác nhau (địa lý, nguồn dữ liệu, phương tiện...) hiện đang hoạt động ra sao trong mối tương quan lẫn nhau?
3) Hiệu quả của tất các các chiến dịch marketing diễn ra trong vòng 12 tháng qua như thế nào?
4) Quy trình thu thập ý kiến phản hồi của khách hàng diễn ra như thế nào, và ý kiến phản hồi của khách hàng trong 12 tháng qua là gì?
5) Hãy liệt kê tất cả các nhà cung cấp của chúng ta, họ đã làm việc với chúng ta được bao lâu rồi, và hiệu quả công việc của họ ra sao?
6) Ý kiến phản hồi của khách hàng ở các nguồn bên ngoài là gì, ví dụ, ý kiến khách hàng chia sẻ trên mạng xã hội và các website chuyên về du lịch như Tripadvisor...?
7) Có phải các đối thủ đang tranh giành công việc kinh doanh của chúng ta? Vì sao khách hàng lại

tìm đến họ thay vì chúng ta, và họ tiếp cận đối thủ của chúng ta như thế nào?

Để tìm câu trả lời cho các câu hỏi này, đội ngũ kỹ thuật của bạn có thể sử dụng nhiều kỹ thuật và công cụ như chúng tôi đã đề cập đến ở Chương 4. Có thể họ sẽ phải thực hiện cài đặt nhiều lần và sử dụng nhiều công cụ khác nhau, và sau khi bạn đã có trong tay bản báo cáo hoặc kết quả phân tích, chúng tôi hy vọng rằng bạn sẽ nhìn ra được những kiến giải có thể giúp bạn xử lý thách thức như đã xác định ở phần đầu tiên của dự án giả tưởng này.

Dưới đây là một số phân tích có thể thực hiện trên các tập dữ liệu đã thu thập được để qua đó tìm ra những phát hiện chính và mối tương thuộc giữa chúng:

1) Phân tích so sánh đối thủ cạnh tranh và thị trường, sử dụng tất cả các dữ liệu đã thu thập được, bao gồm cả dữ liệu liên quan đến mạng xã hội.
2) Phân tích doanh số bán hàng.
3) Phân tích phân khúc khách hàng.
4) Phân tích về marketing.
5) Thu thập ý kiến phản hồi của khách hàng và tiến hành phân tích dựa trên kết quả thu được – từ các nguồn nội bộ và nguồn bên ngoài thông qua kỹ thuật phân tích cảm xúc.

6) Phân tích về hoạt động và hiệu quả hoạt động của nhà cung cấp.

Có thể thực hiện một số phân tích khác dựa trên những phát hiện và các quyết định đưa ra trong các vòng lặp của mô hình này. Trong phần tiếp theo, chúng ta sẽ bàn về việc trình bày kết quả phân tích sao cho có thể phát huy tối đa lợi ích của quy trình này.

P – Presentation and visualization (trình bày và trực quan hóa)

Như đã đề cập ở phần trước, sau khi chạy nhiều loại phân tích khác nhau trên dữ liệu thu thập được từ các nguồn nội bộ và nguồn bên ngoài, đội ngũ thực hiện sẽ tiến hành trình bày kết quả phân tích theo các hình dạng khác nhau tùy theo mục đích sử dụng.

1) Báo cáo doanh số: Có thể trình bày báo cáo doanh số thành một bảng tóm tắt động để bạn, trên cương vị một giám đốc vận hành, đưa ra được những thay đổi về ngày tháng và các yếu tố khác để xem những thay đổi đó tác động ra sao đến doanh số và doanh thu.

   Giả sử qua báo cáo doanh số bạn thấy rằng doanh thu từ bán hàng và các số liệu bán hàng có xu hướng đi xuống trong sáu tháng qua – đây vốn là

một thực tế chúng ta đều đã biết ngay từ khi mới bắt đầu dự án. Tuy nhiên, bạn cũng thấy rằng đây không phải sự sụt giảm đột ngột, mà là xu hướng giảm từ từ, bắt đầu từ khoảng chín tháng trước đó nhưng mới chỉ thể hiện rõ nét hơn trong sáu tháng vừa qua. Điều này nghĩa là không tồn tại một sự kiện duy nhất nào tác động đến sự suy giảm này, mà có điều gì đó đã diễn ra trong các tháng trước, và hiệu ứng của chúng chỉ nổi rõ trong vòng sáu tháng qua.

Tiếp theo, bạn thấy rằng số liệu bán hàng đến từ các kênh trực tuyến như các nền tảng số, mạng xã hội, và website chuyên về du lịch như Tripadvisor... chịu tác động nhiều hơn so với số liệu bán hàng đến từ các nguồn phi trực tuyến. Qua đó, chúng ta rút ra được hai phát hiện quan trọng:

a) *Số liệu bán hàng có xu hướng giảm từ từ, và*

b) *Doanh số trực tuyến chịu ảnh hưởng nhiều hơn so với doanh số phi trực tuyến.*

2) Báo cáo phân khúc: Báo cáo phân khúc bán hàng cho thấy rằng không có khu vực địa lý hay địa điểm cụ thể nào có bất kỳ sự thay đổi lớn nào trong các số liệu bán hàng. Tất cả mọi nơi đều thu được doanh số ít hơn. Như vậy, chúng ta có thể kết luận rằng:

c) *Sự suy giảm về doanh số bán hàng không liên quan gì đến bất kỳ khu vực địa lý nào hay không phụ thuộc vào các cách phân chia nhân khẩu học khác.*

3) Báo cáo phân tích thị trường và so sánh với đối thủ cạnh tranh: Báo cáo này sẽ là một phần trong bảng tóm tắt, trong đó bạn có thể thay đổi thời gian và các yếu tố khác để xem tác động của những thay đổi đó đối với hiệu quả kinh doanh của đối thủ. Dĩ nhiên, bạn không lấy được dữ liệu về tình hình doanh số của công ty đối thủ, nhưng bạn vẫn có thể biết về mức độ nổi tiếng của họ trong không gian trực tuyến và thực hiện các so sánh về hiệu quả từ khóa tìm kiếm giữa hai bên thông qua các công cụ như Google Ads và Google Analytics. Tần suất và tỉ lệ chuyển đổi từ dữ liệu đó không cung cấp thông tin giải thích vì sao một đối thủ cụ thể lại không tận dụng khoảng trống cơ hội từ việc chuỗi BigDataHotels sụt giảm doanh số. Phát hiện rút ra là:

d) *Lợi ích từ sự sụt giảm doanh số của BigDataHotels được chia đều cho tất cả các công ty đối thủ.*

4) Báo cáo về ý kiến phản hồi của khách hàng: Báo cáo này đúc kết lại các dữ liệu về ý kiến phản hồi của khách hàng và dịch vụ khách hàng trong 12

tháng qua. Bạn có thể thay đổi ngày giờ và nguồn cung cấp ý kiến phản hồi để xem liệu có tồn tại bất kỳ xu hướng nào trong luồng thông tin phản hồi của khách hàng hay không. Qua đây, chúng ta phát hiện ra rằng có một lượng lớn ý kiến phản hồi tiêu cực đến từ các nguồn trực tuyến và mạng xã hội, nhưng không có nhiều sự khác biệt trong các ý kiến phản hồi trực tiếp tại khách sạn hoặc qua các kênh phi trực tuyến. Khi tiến hành lọc thông tin phản hồi trên mạng, chúng ta thấy rằng xu hướng gia tăng luồng ý kiến phản hồi tiêu cực bắt đầu từ khoảng 10-11 tháng trước đó.

Bây giờ, chúng ta có thể tiếp tục phân tích nội dung các ý kiến phản hồi tiêu cực để xem liệu có phải tất cả các khách hàng đều than phiền về một vấn đề hoặc về các vấn đề tương tự nhau hay không. Sử dụng kỹ thuật phân tích văn bản, chúng ta phát hiện ra rằng có một số khách hàng phàn nàn về việc gối trong phòng khách sạn khiến họ đau cổ và đau lưng – đây là một phát hiện hay, nhưng chúng ta vẫn cần xem xét chi tiết hơn. Như vậy, sau khi phân tích dữ liệu về ý kiến phản hồi của khách hàng, chúng ta rút ra các phát hiện chính sau đây:

e) *Khách hàng dường như hay nêu ý kiến phản hồi về các trải nghiệm tiêu cực trên mạng*

thay vì nêu vấn đề trực tiếp, ngay trước mặt nhân viên khách sạn.

f) *Mọi người dường như hay đưa lên mạng các ý kiến phản hồi tiêu cực thay vì ý kiến phản hồi tích cực.*

g) *Có sự gia tăng về ý kiến phản hồi của độc giả ở tất cả các nền tảng web, và phần nhiều trong số đó là những lời phàn nàn về việc gối dùng trong khách sạn gây chứng đau cổ và lưng.*

5) Báo cáo về dịch vụ khách hàng: Báo cáo này trình bày thông tin chi tiết về các khách hàng đã liên hệ với nhân viên của khách sạn thuộc chuỗi BigDataHotels tọa lạc tại Phố Bond, London trước, trong, và sau khi họ ở đây. Khi tách riêng để tính số lượng các cuộc gọi từ khách hàng tới nhân viên khách sạn theo thời điểm trước bán hàng, sau bán hàng, và trong thời gian lưu trú tại khách sạn, chúng ta phát hiện ra rằng số lượng các cuộc gọi trong thời gian khách hàng lưu trú tại khách sạn có xu hướng gia tăng trong 11 tháng trở lại đây.

Chúng ta thực hiện phân tích văn bản sâu thêm đối với nội dung các cuộc gọi đến bộ phận dịch vụ khách sạn đã được lưu trữ lại; hóa ra nhiều

khách hàng phàn nàn về chất lượng gối và/hoặc đệm, và một số người thậm chí còn yêu cầu nhân viên đổi chiếc khác. Như vậy, phát hiện chính từ báo cáo này là:

h) *Nhiều khách hàng thông báo về chất lượng gối và đệm trong thời gian họ lưu trú tại khách sạn.*

6) Báo cáo về nhà cung cấp: Tiếp theo chúng ta nghiên cứu báo cáo về nhà cung cấp và những mặt hàng họ đang cung cấp cho khách sạn này. Kết quả, không khó để nhận ra rằng khoảng 11 tháng trước, khách sạn đổi sang một nhà cung cấp mới, và họ đã tiến hành thay thế hầu hết các loại gối và đệm trong khách sạn. Lúc này, mọi chuyện đã bắt đầu sáng tỏ, nhưng có một điều chúng ta còn chưa rõ là tại sao thông tin này lại không được báo cáo trong 12 tháng qua. Cần điều tra một chút về sự bất thường này – và chúng ta sẽ bàn đến ở phần sau – nhưng trước mắt, từ báo cáo này chúng ta rút ra phát hiện sau:

i) *Một nhà cung cấp mới bắt đầu trang bị các loại gối và đệm mới cho khách sạn, khiến nhiều khách hàng cảm thấy không thoải mái và bị đau khi sử dụng chúng.*

Tóm lại, dưới đây là các phát hiện mà chúng ta rút ra sau quá trình phân tích này:

a. *Số liệu bán hàng có xu hướng giảm từ từ, và*

b. *Doanh số trực tuyến chịu ảnh hưởng nhiều hơn so với doanh số phi trực tuyến.*

c. *Sự suy giảm về doanh số bán hàng không liên quan gì đến bất kỳ khu vực địa lý nào hay không phụ thuộc vào các cách phân chia nhân khẩu học khác.*

d. *Lợi ích từ sự sụt giảm doanh số của BigDataHotels được chia đều cho tất cả các công ty đối thủ.*

e. *Khách hàng dường như hay nêu ý kiến phản hồi về các trải nghiệm tiêu cực trên mạng thay vì nêu vấn đề trực tiếp, ngay trước mặt nhân viên khách sạn.*

f. *Mọi người dường như hay đưa lên mạng các ý kiến phản hồi tiêu cực thay vì ý kiến phản hồi tích cực.*

g. *Có sự gia tăng về ý kiến phản hồi của độc giả ở tất cả các nền tảng web, và phần nhiều trong số đó là những lời phàn nàn về việc gối dùng trong khách sạn gây chứng đau cổ và lưng.*

h. *Nhiều khách hàng thông báo về chất lượng gối và đệm trong thời gian họ lưu trú tại khách sạn.*

i. *Một nhà cung cấp mới bắt đầu trang bị các loại gối và đệm mới cho khách sạn, khiến nhiều khách hàng cảm thấy không thoải mái và bị đau khi sử dụng.*

Dựa trên các phát hiện này, đây là một kết luận hợp logic mà chúng ta có thể rút ra: Khoảng 11 tháng trước, một nhà cung cấp mới đã bắt đầu cung cấp các loại gối và đệm mới, và các sản phẩm này đã khiến khách hàng lưu trú tại khách sạn bị đau lưng và cổ. Nhiều người đã lên tiếng phàn nàn về chuyện này, và yêu cầu nhân viên khách sạn đổi gối/đệm cho họ. Tuy nhiên, cũng có nhiều người không để lại ý kiến phản hồi gì khi trả phòng, có lẽ họ thấy không thoải mái khi phải phàn nàn chuyện tiêu cực trực tiếp ngay trước mặt nhân viên. Do đó, nhiều người trong số họ đưa những ý kiến phản hồi tiêu cực lên mạng và mạng xã hội. Trong suốt 12 tháng qua, những thông tin tiêu cực này đã tích tụ ngày càng nhiều trong không gian trực tuyến và bắt đầu xuất hiện trên các website bán hàng trực tuyến cũng như các website đối tác của khách sạn như Tripadvisor. Vì phần lớn người dùng các thiết bị kết nối internet đều phụ thuộc rất nhiều vào luồng ý kiến phản hồi đăng tải trên mạng, nên luồng thông tin tiêu cực trên đã ảnh hưởng lớn đến số liệu bán hàng trực tuyến của khách sạn, và cuối cùng doanh số bán hàng nói chung giảm xuống.

Có thể coi như chúng ta đã phát hiện ra nguyên nhân khiến doanh số của khách sạn giảm, nhưng chúng ta vẫn còn một vài câu hỏi chưa có lời giải đáp và một thách thức nữa cần giải quyết:

Vì sao không có ai phát hiện ra vấn đề này?

Hầu hết các khách hàng đều dùng đến gối và đệm trong lúc họ ngủ vào buổi tối, sau đó họ gọi điện thông báo về chất lượng gối/đệm, và vấn đề đã được khắc phục. Dĩ nhiên, thông tin về các cuộc gọi này đều được lưu trữ lại, chỉ có điều các báo cáo hằng ngày chỉ nêu số lượng các cuộc gọi chứ không cung cấp nội dung cuộc gọi. Có bằng chứng rõ ràng cho thấy rằng các cuộc gọi này phần nhiều được thực hiện vào giờ khuya, nhưng vì không có người gõ lại nội dung cuộc trao đổi đã được ghi âm lại để tiến hành phân tích trên kho dữ liệu phi cấu trúc đó, nên vấn đề này không được phát hiện và giải quyết.

Trong phân tích của mình, chúng ta đã xác định được một vài dấu chấm, và trong quá trình kết nối các dấu chấm đó lại với nhau, chúng ta tìm hiểu và biết được nội dung của các cuộc trao đổi đó, nên cuối cùng đã tìm ra được nguyên nhân thực sự đằng sau hiện tượng doanh số của khách sạn sụt giảm.

Quyết định: Bây giờ, chúng ta đã xác định được nguyên nhân sâu xa của vấn đề mà BigDataHotels

đang phải đối mặt, từ đó họ có thể đưa ra được những hành động phù hợp nhằm thay thế các sản phẩm do nhà cung cấp kia đưa đến, vốn là yếu tố khiến khách hàng cảm thấy khó chịu khi sử dụng.

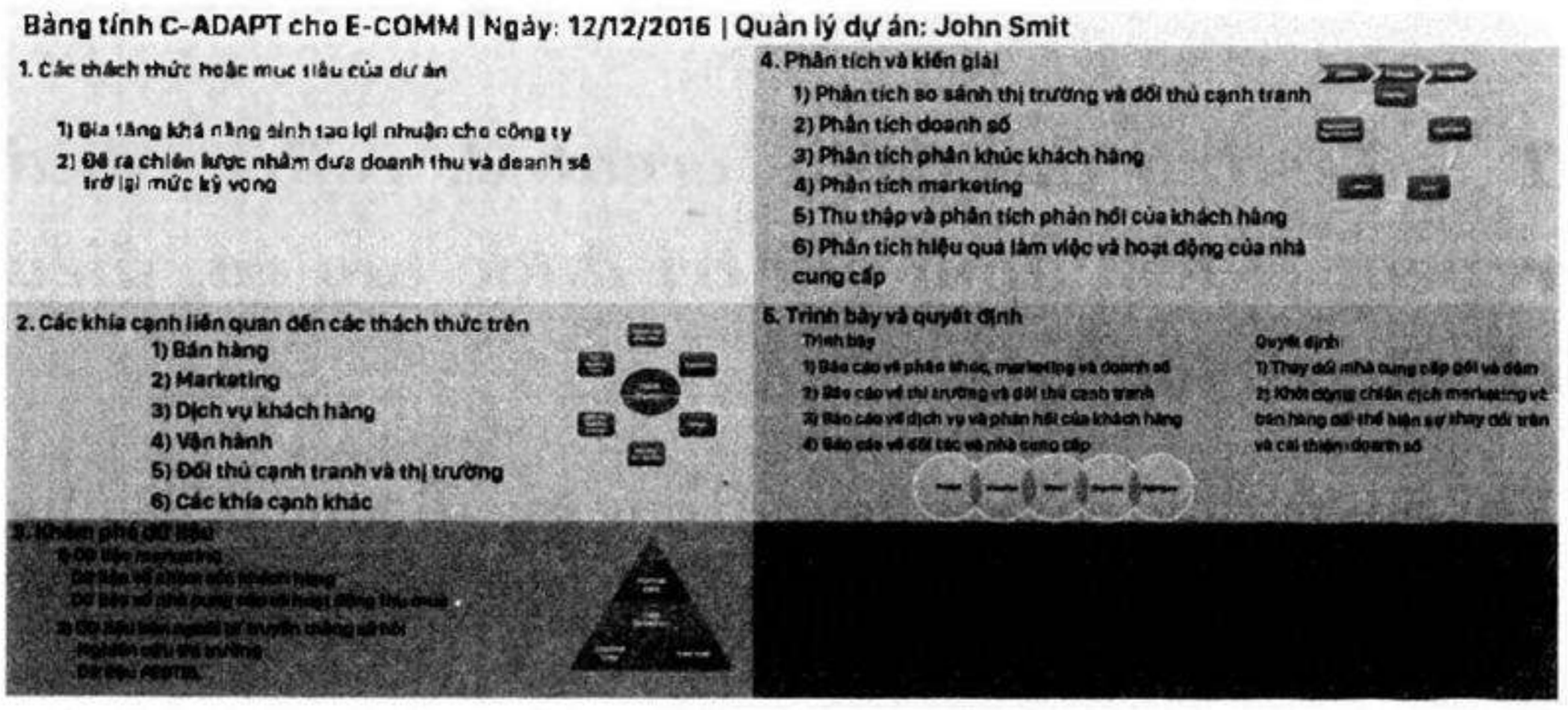
Mục tiêu tiếp theo là khôi phục lại doanh số cho khách sạn như lúc đầu. Sau khi đã khắc phục vấn đề về nhà cung cấp, nên giao lại dự án này cho đội ngũ marketing và bán hàng để họ triển khai các chiến dịch nhằm thông báo về các thay đổi này và tìm cách tác động đến hành vi khách hàng để họ quay lại sử dụng khách sạn như trước kia. Nên thực hiện các dự án quản lý danh tiếng trực tuyến nhằm khôi phục lại vị thế thương hiệu và củng cố thêm cho nó.

T – Testing the value created, refine, and repeat (Kiểm định giá trị được tạo ra, điều chỉnh, và lặp lại)

Khi quyết định được đưa ra trong giai đoạn cuối cùng của mô hình C-ADAPT, nhà quản lý cần phải liên tục theo dõi và kiểm định hiệu suất bán hàng trực tuyến, đồng thời quan sát kỹ lưỡng các ý kiến phản hồi mới... Bảng báo cáo tóm tắt động về hiệu suất hoạt động của các bộ phận bán hàng, marketing, nhà cung cấp, và dịch vụ khách hàng sẽ giúp đội ngũ thực hiện dự án gia tăng doanh thu tổng, nhưng đồng thời cũng mang lại sự hài lòng cho khách hàng.

Quan trọng hơn, dự án này không chỉ nhằm xác định nguyên nhân chính hay tìm cách khôi phục lại doanh số. Điều cốt lõi chính là quá trình kiến tạo giá trị từ các phân tích mà chúng tôi thực hiện. Như đã bàn đến ở Chương 3, giá trị chung phải được tính toán trên cơ sở giá trị được tạo ra và khoản đầu tư đã bỏ ra để vận hành dự án dữ liệu lớn – nếu sự chênh lệch giữa hai con số này là dương, có thể khẳng định rằng dự án này đã tạo ra giá trị, vì thế nên tiếp tục lặp lại để hoàn thiện nó.

HÌNH 5.15 Bảng tính C-ADAPT hoàn chỉnh điển hình



Đây là bảng tính C-ADAPT hoàn chỉnh (xem Hình 5.15) phản ánh toàn bộ dự án trên – nên làm tuần tự từng phần nhưng với dự án này, chúng tôi giới thiệu luôn bảng tính cuối cùng. Bảng tính này đã được hoàn thiện dần ở phần trước, khi chúng ta cùng

nhau rà soát một lần mô hình C-ADAPT và qua đó bàn về nhiều vấn đề khác nhau khi làm việc với dự án này.

TÓM TẮT

Đây là chương quan trọng nhất, cũng là bài học chính có thể rút ra từ cuốn sách này. Sau khi học về dữ liệu lớn, các khái niệm liên quan đến kiến tạo giá trị, các công cụ và kỹ thuật xung quanh dữ liệu lớn, thì điều quan trọng nhất đối với nhà quản lý là mô hình và khung có tính hệ thống để đưa những kiến thức họ vừa học được vào thực tế. Dù bạn là nhà quản lý mới hoặc đã dày dạn kinh nghiệm dẫn dắt nhiều dự án dữ liệu lớn trước đây, bạn vẫn sẽ tìm thấy nhiều giá trị to lớn trong mô hình C-ADAPT và bảng tính đi kèm mà chúng tôi vừa giới thiệu.

Như bạn đã thấy trong chương này, mô hình và bảng tính C-ADAPT là hai công cụ đắc lực mang lại cho bạn những kiến giải và cách thức tiếp cận bất kỳ dự án dữ liệu lớn nào cho công việc của bạn. Chúng tôi cũng đã trình bày hai trường hợp ví dụ và cùng bạn đi một lượt các quy trình để bạn có thể dễ dàng liên tưởng tới các dự án của mình và tự tin triển khai chúng.

Có hai điểm chúng ta phải luôn cân nhắc đến khi làm việc với mô hình C-ADAPT:

- Phân tích dữ liệu lớn nên bao gồm cả *phân tích phi cấu trúc*, nếu không dự án của bạn sẽ không có gì khác biệt so với các dự án tình báo doanh nghiệp (Business Intelligence – BI) truyền thống.
- Nhà quản lý phải đo lường được *giá trị chung tạo ra* từ dự án dữ liệu lớn, tính đến tất cả các loại chi phí bỏ ra cho dự án này, từ chi phí chuẩn bị, xây dựng cơ sở hạ tầng, đến chi phí nguồn lực... Nếu giá trị bạn tạo ra không lớn hơn khoản đầu tư đã bỏ vào đó, bất kể bạn có tạo ra bao nhiêu doanh thu sau dự án này hoặc đã tiết kiệm được bao nhiêu chi phí, thì giá trị chung tạo ra vẫn là âm, và dự án này không thể tiếp tục thêm nữa.

Nếu bắt buộc phải chọn ra hai bài học chính của cuốn sách này, thì theo chúng tôi, đó sẽ là mô hình và bảng tính C-ADAPT.

6

CÁC TRƯỜNG HỢP VÍ DỤ ĐIỂN HÌNH VỀ DỮ LIỆU LỚN

Ở các chương trước, chúng ta đã thấy tầm quan trọng của dữ liệu lớn, ý nghĩa của việc kiến tạo giá trị, cũng như cách sử dụng dữ liệu lớn để tạo giá trị tài chính và phi tài chính cho doanh nghiệp. Ngoài ra, chúng tôi cũng lưu ý rằng hiện nay, một số vị lãnh đạo và những người sớm tiếp nhận công nghệ này đã bắt đầu sử dụng dữ liệu lớn để tạo giá trị. Chúng tôi dành chương này để nói về các lãnh đạo và những người nhanh chân lẹ bước đó cùng với những câu chuyện thành công và trải nghiệm của họ trong lĩnh vực này.

Ở chương này, chúng tôi giới thiệu các trường hợp ví dụ về hoạt động phân tích dữ liệu lớn và cách các tổ chức áp dụng phát kiến công nghệ tiên tiến nhất

này vào quá trình kiến tạo giá trị. Để bao quát được nhiều trường hợp, và cũng là để độc giả có được cái nhìn toàn diện về khái niệm và phạm vi ứng dụng của dữ liệu lớn, chúng tôi đưa vào đây nhiều ví dụ khác nhau, từ các công ty công nghệ đang phát triển mạnh mẽ như Ooredoo (Qtel) đến các công ty hoạt động trong lĩnh vực nông nghiệp tuy không quá nổi tiếng nhưng đóng vai trò rất quan trọng và giá trị như John Deere. Rất dễ nhận thấy rằng nếu bạn có bất kỳ bản mô tả vấn đề nào như chúng tôi sẽ trình bày trong các trường hợp ví dụ sau đây, bạn có thể tìm ra một giải pháp ấn tượng nhờ sử dụng mô hình C-ADAPT mà chúng tôi đã đề xuất. Chúng ta có thể thấy rằng hầu hết các giai đoạn của mô hình C-ADAPT đều có thể trực tiếp áp dụng vào đa số các trường hợp ví dụ đó, nhưng vì không có đầy đủ thông tin chi tiết về tình hình đầu tư thực tế và lợi nhuận thu về, nên chúng tôi không thể thực hiện các phép tính ROIC cuối cùng về giá trị ròng tạo được như minh họa ở Chương 3. Tuy vậy, trong những trường hợp đó chúng tôi vẫn có thể yên trí giả định rằng các tổ chức này đã đạt được những mục tiêu mà họ nhắm đến, lợi ích chung mà họ thu được lớn hơn so với các khoản đầu tư đã bỏ ra, các yêu cầu về việc duy trì dự án và nguồn lực... phải đáp ứng để cuối cùng dẫn đến giá trị đạt được dương.

Các tác giả của cuốn sách này đã làm việc trực tiếp trong một số trường hợp ví dụ mà chúng tôi nêu

ra trong chương này, có thể trong công ty cũ hoặc các dự án tư vấn riêng mà chúng tôi đảm nhận. Chúng tôi cũng giới thiệu một số trường hợp ví dụ do những nhà quản lý làm việc trong các công ty khác ở nhiều lĩnh vực ngành nghề đóng góp. Mục đích của các ví dụ này là giúp bạn vận dụng được kiến thức đã lĩnh hội từ các chương trước và liên hệ chúng với những thách thức và mục tiêu của tổ chức của bạn, qua đó tìm hiểu xem bạn có thể triển khai các dự án như vậy theo cách nào để tạo giá trị bền vững trong tổ chức.

OOREDOO (TÊN CŨ LÀ QTEL)

Tái xây dựng thương hiệu và khuếch tán thông điệp

Thông tin chung

Qatar Telecom (Qtel) là một nhà cung cấp dịch vụ viễn thông hàng đầu trong khu vực Trung Đông, đồng thời từng là "công ty viễn thông có tốc độ phát triển nhanh nhất thế giới". Ooredoo cung cấp các dịch vụ di động, không dây, có dây, và dịch vụ nội dung với thị phần ở cả các thị trường viễn thông trong nước và quốc tế, trên cả thị trường doanh nghiệp lẫn dân sinh. Đây là một trong những công ty viễn thông di động lớn nhất thế giới; tại thời điểm tháng 9 năm 2015, công ty này có hơn 114 triệu

khách hàng trên khắp thế giới. Cũng theo số liệu vào tháng 9 năm 2015, giá trị vốn hóa thị trường của Ooredoo là 301,2 tỉ QAR. Tại lễ trao giải của chương trình World Communication Awards (Giải thưởng Truyền thông Thế giới) năm 2013, công ty này được vinh danh là "Nhà mạng di động tốt nhất năm".

Ooredoo là một công ty viễn thông quốc tế hàng đầu thế giới chuyên cung cấp các dịch vụ dành cho điện thoại di động, điện thoại cố định, dịch vụ internet băng thông rộng, và các dịch vụ quản lý doanh nghiệp đáp ứng nhu cầu của người tiêu dùng và các doanh nghiệp hoạt động trên khắp các thị trường Trung Đông, Bắc Phi, và Đông Nam Á. Ooredoo có mặt ở các thị trường như Qatar, Kuwait, Oman, Algeria, Tunisia, Iraq, Palestine, Maldives, Myanmar, và Indonesia. Năm 2014, công ty này báo cáo doanh thu lên tới 9,1 tỉ đô-la, và có một cơ sở khách hàng toàn cầu vững chắc gồm hơn 107 triệu người (theo số liệu tính đến ngày 31 tháng 12 năm 2014). Cổ phiếu của Ooredoo được niêm yết trên Sàn Giao dịch Chứng khoán Qatar và Sàn Giao dịch Chứng khoán Abu Dhabi.

Các cơ hội/thách thức

Theo thông báo của Qtel tại Hội nghị Di động Thế giới năm 2013, tập đoàn này dự định triển khai một chiến dịch làm lại thương hiệu trên toàn cầu nhằm hợp nhất tất cả các thương hiệu của họ ở từng quốc gia

dưới một cái tên mới: Ooredoo. Trong tiếng Ả-rập, từ *Ooredoo* nghĩa là *tôi muốn*. Thương hiệu mới mang ý nghĩa toàn cầu nhưng vẫn đậm màu sắc cá nhân nhằm làm hài lòng khách hàng hiện tại và tương lai của họ cũng như các bên liên quan khác.

Nhà mạng xuất thân từ Trung Đông này thuê hãng tư vấn và quan hệ công chúng toàn cầu FTI Consulting London để hướng dẫn họ phát triển và triển khai thành công chiến lược xây dựng thương hiệu mới. Gã khổng lồ viễn thông muốn các thông điệp của thương hiệu mới tiếp cận một lượng khán giả mục tiêu rộng lớn hơn để tạo ra những ấn tượng sâu đậm và lâu dài.

Trong các cộng đồng gắn bó khăng khít với nhau của vùng Trung Đông, đại đa số các cuộc trao đổi trực tuyến đều diễn ra trong không gian riêng tư hoặc giữa những người dùng đã quen biết nhau ngoài đời thực. Qtel muốn tìm kiếm những cách thức hiệu quả để đưa các thông điệp của họ tới những người cần nghe chúng, đặc biệt trong giai đoạn tái xây dựng thương hiệu.

Phương pháp

Nhiệm vụ lớn nhất của FTI là tìm kiếm một phương pháp nhanh chóng và hiệu quả để cùng lúc đưa thông điệp về thương hiệu mới tới đại bộ phận khán giả

trong nước vốn không quen nói tiếng Anh và một bộ phận khán giả toàn cầu nói tiếng Anh. Nhìn rộng ra, đó là vấn đề liên quan đến marketing và xây dựng thương hiệu, và nhiệm vụ đầu tiên là khoanh vùng các nhóm cộng đồng cần đẩy mạnh truyền thông về thông điệp của thương hiệu mới thông qua một chiến dịch gây nhiều ảnh hưởng lớn.

FTI bắt đầu cộng tác với một công ty khởi nghiệp công nghệ có trụ sở tại London, làm việc trong lĩnh vực phân tích truyền thông xã hội và marketing. Họ thực hiện các thuật toán SMA (Social Media Analysis – Phân tích mạng xã hội) để xác định ra đề tài được ưa chuộng nhất và có liên quan nhất trong thị trường mục tiêu, và sau đó sử dụng dữ liệu thu thập được để tìm ra những nút mạng (tức các tài khoản mạng xã hội) nổi tiếng nhất đang hoạt động hoặc đang nói về các đề tài này. FTI bắt đầu tìm kiếm những người có ảnh hưởng trong thế giới trực tuyến trong nước và trên phạm vi toàn cầu ở nhiều kênh khác nhau và tiếp cận để mời họ chung tay lan truyền các thông điệp truyền thông của Qtel. Công ty này sử dụng hệ thống phân tích mạng xã hội và phân tích văn bản thông minh nhằm xác định các mạng xã hội, mạng lưới, và cá nhân có tầm ảnh hưởng lớn.

Toàn bộ chiến lược quan hệ công chúng và marketing được lên kế hoạch và triển khai trên cơ sở

là các kiến giải do đội ngũ của FTI phát hiện ra. Nhờ thực hiện phân tích nội dung và các cuộc trao đổi, họ biết được đâu là những nút mạng chính trong các mạng xã hội, chủ đề bàn luận và các kênh ưa thích của họ là gì, và có thể làm gì để mời gọi họ trở thành người ủng hộ thương hiệu, sẵn sàng lan truyền thông điệp của thương hiệu. Tiếp đến mỗi người có ảnh hưởng đó lại được phân tích và đưa vào các nhóm phù hợp để hiểu rõ hơn về sở trường và mức độ ảnh hưởng của từng người cũng như các khán giả mà họ tương tác. Những dữ liệu marketing và truyền thông trong nội bộ công ty cũng như bên ngoài được sử dụng để tìm ra xu hướng giao tiếp và các mối quan tâm của những người có ảnh hưởng.

Ngoài ra, trong Hội nghị Di động Thế giới (World Mobile Congress) năm 2013, toàn bộ hoạt động tái xây dựng thương hiệu diễn ra chóng vánh: Vào ngày đầu tiên của hội nghị, trọng tâm dồn vào Qtel và cả sự kiện ngập trong tông màu xanh của thương hiệu cũ; ngày thứ hai, toàn bộ không gian lại được trang trí bằng màu đỏ sống động – màu chủ đạo của thương hiệu mới! Cùng lúc đó, những thông điệp mới cũng tiếp cận những nhân vật có ảnh hưởng đã được xác định ở khâu trước và các mạng lưới của họ. Jack Morton, một hãng xây dựng thương hiệu toàn cầu, được Qtel thuê để thực hiện cuộc lột xác chóng vánh này.

Kết quả

Sau khi đã xác định những người có ảnh hưởng trên nhiều kênh liên lạc số khác nhau với những chủ đề ưa thích khác nhau, FTI và Qtel trực tiếp tiếp cận các nhân vật đó để mời gọi họ trở thành người ủng hộ thương hiệu và lan truyền thông điệp về thương hiệu. Kết quả, những nội dung truyền thông về thương hiệu của công ty viễn thông trên đã tiếp cận được một lượng khán giả rộng khắp, nhiệt tình hơn trên một khu vực rộng lớn gồm Bắc Phi và Trung Đông, khiến màn ra mắt của thương hiệu Ooredoo mới có sức ảnh hưởng đặc biệt.

Với hệ thống quản lý và phân tích mạng xã hội, Qtel có thể tiếp tục xác định và lôi kéo sự tham gia của những người có ảnh hưởng trong không gian trực tuyến, qua đó bảo đảm việc tiếp tục khuếch tán các tin tức và nội dung marketing của công ty.

DOMINO'S PIZZA

Tìm hiểu hành vi mua hàng

Thông tin chung

Domino's Pizza là một chuỗi nhà hàng pizza của Mỹ, đồng thời là thương hiệu nhượng quyền mô hình nhà hàng pizza trên toàn thế giới có trụ sở tại trang trại Domino ở thị trấn Ann Arbor, Michigan, Mỹ. Được thành lập năm 1960, Domino's hiện nay đang hoạt

động ở 81 quốc gia, là chuỗi pizza nhượng quyền lớn thứ hai thế giới sau Pizza Hut.

Domino's Pizza được coi là công ty giao pizza hàng đầu thế giới. Với cơ sở khách hàng toàn cầu gồm toàn những người tiêu dùng tích cực tham gia hoạt động trong không gian số, Domino's rất quan tâm đến việc tận dụng tối đa kho dữ liệu số đồ sộ đang có trong tay để tối ưu hóa cách thức vận hành của họ.

Các cơ hội

Domino's Pizza nổi tiếng với việc phát triển các quy trình và không ngừng sáng chế ra những cách thức mới mẻ để áp dụng công nghệ vào việc phát triển kinh doanh. Bằng cách đẩy lùi các giới hạn và đổi mới công nghệ và kỹ thuật để tiếp cận khách hàng, chuỗi này tạo điều kiện cho khách hàng đặt bánh trên Twitter, Facebook, SmartWatch, ti-vi, hay thậm chí từ hệ thống giải trí gắn trên ô tô. Phương châm đưa nút đặt hàng tới mọi nơi cho khách hàng thuận tiện sử dụng được gọi là "Domino's Anyware".

Domino's Anyware bắt đầu thu thập dữ liệu khách hàng từ rất nhiều hướng khác nhau, và đã mang lại một cơ hội lớn để công ty này có thể phân tích và lập hồ sơ khách hàng trên cơ sở hành vi mua sắm của họ, qua đó khuyến khích họ tích cực đặt hàng nhiều hơn nữa đồng thời cải thiện không ngừng chất lượng phục vụ khách hàng.

Phương pháp

Đội ngũ của Domino's bắt đầu thu thập dữ liệu từ tất cả các kênh có thể như tin nhắn văn bản, Twitter, Pebble, Android, Amazon Echo... và sau đó chuyển tất cả các dữ liệu này tới hệ thống quản lý thông tin. Dữ liệu này sau đó được kết hợp với các nguồn dữ liệu khác như địa điểm, thông tin nhân khẩu học, và dữ liệu về hoạt động trên các website... Nguồn dữ liệu kết hợp phong phú mới này mở ra những kiến giải quan trọng và có ý nghĩa đối với tất cả các bộ phận trong công ty, từ vận hành đến kinh doanh. Thông tin thu thập thông qua các điểm hệ thống bán hàng của công ty cùng với kho dữ liệu phong phú trên tạo thành tới 85.000 nguồn dữ liệu, cả có cấu trúc và phi cấu trúc, liên tục đổ về hệ thống mỗi ngày. Một đội ngũ gồm các chuyên gia cố vấn phân tích dữ liệu nghiên cứu các dữ liệu này và đưa ra những thông tin sâu sắc ẩn chứa trong biển dữ liệu mênh mông, xác định những điểm cần cải thiện và liên hệ nguồn dữ liệu đó với các khía cạnh khác nhau trong hoạt động của Domino's.

Kết quả

Sau khi thực hiện khâu phân tích dữ liệu lớn này, Domino's không những có thể nhìn nhận mỗi người tiêu dùng như một cá nhân riêng biệt và đánh giá xu hướng mua sắm của họ, mà còn xác định được những yếu tố gây ảnh hưởng trong quá

trình mua sắm. Domino's có thể tiếp cận các dữ liệu về vị trí, xu hướng mua sắm, và thói quen sử dụng các phiếu giảm giá của khách hàng... Nhờ vậy, họ biết kênh bán hàng nào hoạt động tốt hơn, và làm thế nào để cải thiện doanh số cho các kênh. Hiện nay Domino's tiếp nhận 55% yêu cầu đặt hàng từ các kênh trực tuyến hoặc kỹ thuật số, và điều này mang lại cho họ nhiều lợi ích lớn so với đối thủ. Với tất cả những ý tưởng đổi mới sáng tạo trên, Domino's Pizza đang chuyển mình trở thành một cổng thương mại trực tuyến cho các mặt hàng đồ ăn rất cụ thể. Giờ đây họ đã thoát khỏi giai đoạn chỉ xử lý các dữ liệu liên quan đến doanh số và các chỉ số vận hành trong nhà kho; mỗi đơn vị nghiệp vụ trong Domino's hiện đều đang tích cực tìm cách tận dụng nguồn dữ liệu dồi dào để thúc đẩy quá trình vận hành của mình nhanh hơn nữa, hiệu quả hơn nữa về mặt chi phí.

CÔNG TY KINH DOANH PHẦN MỀM CHỐNG VIRUS HÀNG ĐẦU

Gia tăng sự tham gia của khách hàng và đẩy mạnh doanh thu

Thông tin chung

Một công ty chuyên cung cấp dịch vụ internet và an ninh di động thuộc hàng phát triển nhanh bậc nhất nước Mỹ muốn thúc đẩy hơn nữa các quy trình quản lý marketing và truyền thông của mình, trong đó

nhấn mạnh đến nhiệm vụ nâng cao nhận thức, sự tham gia, và tài sản thương hiệu của công ty nhằm tác động tích cực đến sự tăng trưởng doanh thu.

Thị trường phần mềm chống virus hiện đã trở nên quá đông đúc, trong đó mười công ty lớn nhất chiếm giữ khoảng 88% thị phần. Công ty của chúng ta đã thiết lập được một cơ sở khách hàng phù hợp, nhưng lúc này họ đang phải đối mặt với thách thức của việc tiếp tục phát triển hơn nữa dựa trên thành công ban đầu đó.

Cơ hội/thách thức

Qua tìm hiểu, họ xác nhận được rằng mạng xã hội đóng vai trò thiết yếu trong các nỗ lực marketing của công ty để vượt qua những thách thức trên. Với nguồn lực hạn chế, họ phải đẩy mạnh lưu lượng truy cập tới website của mình, gia tăng lượng tải phần mềm, cải thiện doanh số, duy trì lượng khách hàng hiện tại và thúc đẩy doanh thu nói chung.

Họ thuê một công ty phân tích dữ liệu để giúp họ phân tích và sử dụng mạng xã hội nhằm đạt được các mục tiêu đã đề ra.

Phương pháp

Công ty phân tích dữ liệu được lựa chọn đã thực hiện một phân tích sâu về các hoạt động trên mạng xã hội của công ty phần mềm này và công ty đối thủ.

Họ sử dụng một công cụ phân loại văn bản thông minh nhằm xác định và phân tích những chủ đề chính trong các cuộc trao đổi trực tuyến của công ty này, đối thủ cạnh tranh của họ, và người tiêu dùng.

Cách tiếp cận này cho phép họ xác định được:

- Các nội dung có khả năng khơi dậy sự hào hứng và thu hút sự tham gia của người dùng trực tuyến, mang lại nhiều lượt thích, chia sẻ, bình luận... trên mạng.
- Những lỗ hổng giữa nội dung các cuộc trao đổi của khách hàng và nội dung truyền thông của thương hiệu.
- Sự khác biệt trong thông điệp và nội dung giữa các nhóm nhân khẩu học khác nhau, chẳng hạn vị trí, giới tính, và thời gian.

Họ sử dụng một hệ thống phân tích mạng xã hội để thực hiện phân tích ngữ nghĩa nhằm xác định những mô típ và chủ đề tạo ra những ý kiến phản hồi tích cực và tiêu cực. Hệ thống này cũng được sử dụng để tìm ra những người có ảnh hưởng chính, vì nếu được họ ủng hộ thì phạm vi tiếp cận nói chung của thương hiệu sẽ còn được mở rộng hơn rất nhiều nữa.

Những phân tích này được kết hợp các dữ liệu về marketing và bán hàng từ Google Analytics, từ đó đưa

ra được một bức tranh toàn thể về vai trò của mạng xã hội trong việc thúc đẩy lưu lượng truy cập website và cuối cùng là gia tăng doanh thu. Sau khi thực hiện các hoạt động phân tích sâu này, các chuyên gia của chúng ta có thể đưa ra nhiều kiến nghị giá trị cho khách hàng về việc nên sử dụng các tài sản mạng xã hội như thế nào để tăng cường lưu lượng truy cập website và doanh thu.

Kết quả

Công ty này bắt tay vào thực hiện các kiến nghị được đưa ra sau khi phân tích dữ liệu. Sử dụng một hệ thống quản lý và giám sát được thiết kế dành riêng cho họ, công ty phần mềm này đã gia tăng sự hiện diện của họ trên mạng xã hội, theo dõi được tác động của các chiến dịch họ thực hiện, tìm kiếm khách hàng mới và thường xuyên theo dõi đường đi của khách hàng từ lần tương tác đầu tiên đến khâu cuối cùng là mua sắm. Hệ thống phân tích này giúp đội ngũ marketing của công ty phần mềm tạo ra các hoạt động marketing phù hợp và đúng mục tiêu để khuyến khích sự tham gia của các khách hàng hiện tại cũng như khách hàng mới. Kết quả, họ giảm được chi phí marketing, gia tăng tỉ lệ chuyển đổi thành khách hàng, và cải thiện doanh thu.

GATE GOURMET

Sử dụng dữ liệu cho hoạt động phân tích đối thủ cạnh tranh

Thông tin chung

Gate Gourmet là một công ty con của tập đoàn Gategroup có trụ sở ở Zurich, Thụy Sĩ và các trung tâm khác đặt tại Reston thuộc bang Virginia của Mỹ và London ở Vương quốc Anh. Công ty này chuẩn bị hơn 250 triệu bữa ăn mỗi năm tại 120 phòng bếp phục vụ cho các chuyến bay ở khắp nơi trên thế giới. Mỗi ngày công ty này thực hiện tới 1.200 cuộc kiểm tra an toàn thực phẩm, góp phần vào việc duy trì một thành tích ấn tượng về an toàn thực phẩm và được coi là một trong những công ty tốt nhất trong lĩnh vực cung cấp thực phẩm.

Cơ hội/thách thức

Thay mặt cho Gate Gourmet, một nhóm cố vấn tìm cách xác định vị thế hiện tại của công ty này và tìm ra những cơ hội mới để họ mở rộng hợp tác với các hãng hàng không và các công ty tàu hỏa trong mảng cung cấp thực phẩm độc lập.

Gate Gourmet muốn tìm ra những cách thức mới để khai thác thị trường và phát triển thương hiệu, để thực hiện các phân tích về đối thủ, và để hiểu biết sâu sắc hơn về vị thế của mình trên thị trường.

Phương pháp

Sử dụng một hệ thống phân tích dữ liệu, nhóm cố vấn trên thực hiện một đánh giá toàn diện về tình hình thị trường, so sánh hoạt động và sự hiện diện trong không gian trực tuyến cũng như tài sản thương hiệu của Gate Gourmet với các đối thủ lớn của họ trong ngành.

Họ theo dõi và phân tích những lượt đề cập đến thương hiệu Gate Gourmet và đối thủ trên mạng nhằm xác định ra những kênh đang hoạt động tốt và quan trọng hơn là các kênh đang chưa được khai thác đầy đủ, bao gồm mạng xã hội, website tin tức, diễn đàn, và blog.

Kết quả, họ phát hiện ra rằng thương hiệu Gate Gourmet được nhắc đến nhiều nhất trên mạng xã hội, nhưng các đối thủ của công ty này lại chiếm phần lớn lưu lượng cho các tin tức và bài báo lớn, có ý nghĩa quan trọng hơn.

Họ phân tích các lượt đề cập đến thương hiệu và nội dung các cuộc trao đổi này để tìm ra những đề tài lặp đi lặp lại, những cơ hội tiềm năng mới công ty này có thể khai thác. Ngoài ra, họ cũng tiến hành phân tích các nội dung phổ biến trong toàn ngành để định hình nên con đường sáng tạo và tối ưu hóa nội dung khả dĩ trong tương lai nhằm phát huy tối đa cơ hội cho các nội dung của họ được nhiều người dùng quan tâm và chia sẻ.

Họ sử dụng một hệ thống phân tích dữ liệu để tìm ra những người có ảnh hưởng lớn trong ngành và những người đang ủng hộ cho thương hiệu Gate Gourmet cũng như thương hiệu đối thủ. Có thể nhắm đến tiếp cận những người có ảnh hưởng này để củng cố thêm độ lan tỏa của các chiến dịch và thông điệp của công ty.

Sau khi thực hiện phân tích để tìm ra nguồn tạo nội dung, họ phát hiện ra rằng đại đa số các bài đăng đều được thực hiện qua web, nghĩa là từ máy tính xách tay hoặc máy tính để bàn. Để gia tăng tỉ lệ chia sẻ trên mạng xã hội, phải tối ưu hóa nội dung để người dùng có thể dễ dàng xem và đọc trên các thiết bị di động có kết nối với internet, trong đó ưu tiên việc tối ưu hóa nội dung trên các thiết bị được hệ thống phân tích dữ liệu xác định là đang được nhóm khách hàng mục tiêu sử dụng.

Kết quả phân tích từ khóa cho thấy mọi người cũng bàn luận về sự nghiệp và cơ hội việc làm ở Gate Gourmet trên các mạng xã hội và website. Điều này chỉ ra một cơ hội tiềm năng để công ty này tận dụng các kênh mạng xã hội nhằm tìm kiếm nguồn nhân tài mới.

Giá trị/kết quả

Kết quả phân tích cho thấy có một số kênh trực tuyến hiện đang chưa được khai thác đầy đủ, qua đó cho thấy những cơ hội tiềm năng để thu hút độc giả, thúc đẩy hoạt động trong không gian trực tuyến, và gia tăng mức độ phủ sóng trên mạng internet.

Họ xác định ra các phương thức để nội dung tối ưu hóa được chia sẻ và phân phối nhằm khích lệ sự tham gia của độc giả và thiết lập vị thế dẫn đầu cho Gate Gourmet trên tất cả các kênh thông tin. Cần tối ưu hóa nội dung để có thể xem được trên cả máy tính và các thiết bị di động nhằm tối đa hóa sự tương tác và hoạt động chia sẻ nội dung.

Dựa trên kết quả phân tích về những nhân vật có ảnh hưởng trong ngành và cơ sở khán giả rộng lớn, nhóm cố vấn cũng đưa ra những kiến nghị về cách hành văn, đề tài, và phương tiện truyền tải nội dung để khích lệ sự tham gia của độc giả, khiến họ tích cực chia sẻ nội dung để gia tăng thiện cảm dành cho thương hiệu cũng như tài sản thương hiệu với mục tiêu cuối cùng là gia tăng sức khỏe thương hiệu nói chung cho Gate Gourmet.

Kết quả phân tích về các cuộc trao đổi liên quan đến thương hiệu và ngành dịch vụ cung cấp thực phẩm nói chung cho thấy nhiều người rất quan tâm

đến các cơ hội việc làm ở Gate Gourmet, vì vậy nhóm cố vấn kiến nghị họ tận dụng mạng xã hội và các phương tiện truyền thông trực tuyến khác để chiêu mộ các nhân tài có sẵn mối quan tâm và lòng nhiệt tình dành cho công ty.

TESCO

Hiểu khách hàng nhờ dữ liệu

Thông tin chung

Tesco là một trong những nhà bán lẻ lớn nhất thế giới, được thành lập năm 1919 tại Anh. Hiện nay, tập đoàn này có hệ thống cửa hàng ở 12 quốc gia và hoạt động dưới nhiều tên thương hiệu khác nhau. Trong năm tài chính 2013/2014, toàn tập đoàn đạt được doanh số bán hàng là 110 tỉ đô-la, và 3,6 tỉ đô-la lợi nhuận trước thuế. Họ có hơn 500.000 nhân viên làm việc tại 7.599 cửa hàng ở khắp nơi trên thế giới, trong đó có cả các cửa hàng hoạt động theo mô hình nhượng quyền. Thị trường lớn nhất của tập đoàn này nằm ở Anh, với hơn 3.500 cửa hàng và trên 310.000 nhân viên. Ngoài hệ thống siêu thị, tập đoàn này còn hoạt động kinh doanh ở nhiều mảng khác, chẳng hạn họ vận hành hệ thống trạm bán xăng, một ngân hàng, và các mảng kinh doanh điện thoại di động, điện thoại để bàn, và dịch vụ băng thông rộng. Dĩ nhiên,

họ cũng triển khai một chương trình khách hàng trung thành. Tất cả các hoạt động này nghĩa là các bộ phận nghiệp vụ của họ đang ngày ngày tạo ra nhiều luồng dữ liệu khổng lồ. Họ bắt đầu sử dụng kho dữ liệu đó từ thập niên 1990 và tới nay đã mở rộng kho dữ liệu lên rất nhiều lần.

Cơ hội

Tesco là một trong những nhà bán lẻ đầu tiên thực hiện thu thập dữ liệu từ khách hàng. Thẻ câu lạc bộ Tesco là một loại thẻ dành cho khách hàng trung thành ra đời vào năm 1995 – và kể từ đó tới nay, thông qua thẻ này, Tesco đã và đang chủ động thu thập dữ liệu về thói quen chi tiêu và hành vi của khách hàng. Ngoài việc sử dụng dữ liệu thu thập qua thẻ câu lạc bộ thông qua các chương trình khuyến mãi và mã giảm giá, Tesco còn có một cơ hội lớn để tìm hiểu hành vi của khách hàng trong bối cảnh các dữ liệu và nhân tố khác, sau đó sử dụng những kiến giải rút ra được để tối ưu hóa các quy trình nghiệp vụ và nâng cao lợi nhuận cũng như doanh thu tổng.

Phương pháp

Hiện nay 2/3 khách hàng của Tesco sử dụng thẻ câu lạc bộ, nghĩa là họ để lại cho Tesco một kho dữ liệu khổng lồ về các hoạt động mua sắm của mình. Dữ liệu

về thói quen chi tiêu và hành vi mua sắm của khách hàng được đưa vào phân tích cùng với nhiều tập dữ liệu và nhân tố khác, qua đó tìm ra những điểm cần cải thiện trong nhiều khía cạnh khác nhau của công ty. Tesco tiến hành phân tích dữ liệu có cân nhắc đến nhiều tập dữ liệu nội bộ khác, chẳng hạn hồ sơ khách hàng, mức độ tiêu thụ điện của tủ lạnh trong cửa hàng, và các tập dữ liệu bên ngoài như các bản tin dự báo thời tiết và xu hướng mua sắm trong điều kiện thời tiết thay đổi/khi trời mưa.

Loại phân tích này đã giúp Tesco dự đoán được nhu cầu và doanh số của các cửa hàng, từ đó tiết kiệm được lượng thực phẩm bị lãng phí do tồn kho.

Kết quả

Với sự hỗ trợ của các bản tin dự báo thời tiết và những dữ liệu bên ngoài, Tesco có thể dự đoán nhu cầu cho hầu hết các sản phẩm mà họ bán. Điều này cũng giúp họ giảm bớt lãng phí thực phẩm, đồng thời tiết kiệm được nhiều khoản tiền lớn. Bằng cách kết hợp dữ liệu thời tiết với dữ liệu về doanh số bán hàng, họ biết điều gì có thể sẽ diễn ra, và trong những năm qua, nhờ khả năng này họ đã giảm được 9 triệu đô-la lượng thực phẩm bị lãng phí vào mùa hè, giảm lãng phí 47 triệu đô-la nhờ tối ưu hóa cách thức vận hành của hệ thống cửa hàng, và giảm 78 triệu đô-la lượng hàng tồn kho.

Một lợi ích nữa là Tesco tối ưu hóa được nhiệt độ của tủ lạnh trong cửa hàng, nhờ đó tiết kiệm được chi phí sử dụng điện. Tesco đã thực hiện dự án phân tích này kết hợp với phòng thí nghiệm nghiên cứu của IBM, và nhờ tối ưu hóa hoạt động của tủ lạnh dựa trên các kết quả phân tích, Tesco đã giảm hóa đơn tiền điện tới gần 25 triệu đô-la mỗi năm.

Như vậy, nhờ biết sử dụng các kỹ thuật phân tích dữ liệu, Tesco không những gia tăng được doanh thu do hiểu rõ khách hàng hơn, mà còn có thể tiết kiệm một khoản tiền lớn thông qua việc giảm bớt lượng thực phẩm bị lãng phí và hóa đơn tiền điện.

HÃNG HÀNG KHÔNG DELTA AIRLINES

Dữ liệu giúp họ tìm kiếm hành lý bị thất lạc

Thông tin chung

Delta Airlines là một hãng hàng không lớn của Mỹ; trụ sở và cũng là trung tâm lớn nhất của hãng này đặt tại sân bay quốc tế Hartsfield-Jackson Atlanta ở Atlanta, bang Georgia. Delta Airlines phục vụ gần 180 triệu hành khách mỗi năm. Năm 2016, Delta được vinh danh trong danh sách 50 Công ty được ngưỡng mộ nhất do tạp chí *Fortune* bình chọn; ngoài ra, đây cũng là lần thứ năm trong sáu năm liên tiếp họ có mặt trong danh sách này. Không chỉ có vậy, Delta còn

xếp thứ nhất trong năm năm liên tục trong khảo sát thường niên của ngành hàng không do hãng thông tấn Business Travel News thực hiện.

Thách thức

Ngành hàng không là một môi trường hoạt động vô cùng khó khăn bởi cường độ cạnh tranh gay gắt và biên lợi nhuận vô cùng mỏng. Nhưng mặc cho vô vàn những thách thức mà ngành hàng không phải đối mặt, Delta vẫn giữ vững vai trò người dẫn đầu ngành với hiệu suất tài chính và vận hành vượt trội.

Tất cả các hãng hàng không đều biết một mối lo ngại lớn đối với hành khách là nguy cơ bị thất lạc hành lý, nhất là khi chuyến bay của họ bị hoãn và bị mất kết nối nhiều lần. Delta nghiên cứu thật kỹ kho dữ liệu của họ và đưa ra một giải pháp nhằm loại bỏ sự không chắc chắn trong thông tin về vị trí hành lý của hành khách.

Delta nổi tiếng với cách tiếp cận mang tinh thần đổi mới và lấy khách hàng làm trung tâm. Họ quyết định bắt tay vào giải quyết một trong những vấn đề phổ biến nhất mà hành khách hay gặp phải: tìm kiếm hành lý bị thất lạc. Theo Paul Skrbec, một phát ngôn viên cho hãng hàng không Delta Airlines, "hầu như hành khách nào cũng đã từng gặp cảnh sau khi gửi hành lý và lên máy bay, họ lại ngồi thấc thỏm không biết hành lý có ở đúng chỗ hay không".

Phương pháp

Delta triển khai công nghệ nhận dạng qua tần số vô tuyến (Radio Frequency Identification – RFID) để theo dõi hành lý của hành khách. Đây là lần đầu tiên một hãng vận tải ở Mỹ áp dụng công nghệ này, cho phép hành khách theo dõi hành lý của họ theo thời gian thực trong suốt hành trình di chuyển. Quyết định này đánh dấu một bước chuyển mang tính lịch sử cho Delta và 120 triệu hành lý mà họ phải xử lý mỗi năm.

Delta đã đầu tư hơn 100 triệu đô-la vào các hệ thống theo dõi hành lý tại sân bay nhằm theo dõi và cải thiện dần khả năng xử lý hành lý của họ. Hiện nay họ đang tích cực sử dụng những công cụ tiên tiến để thu thập và phân tích dữ liệu về hành lý, và đội ngũ vận hành của Delta ở các sân bay cũng như tại trụ sở chính của họ hiện đã có thể xác định chính xác hơn các nguyên nhân cùng xu hướng chính trong các trường hợp xử lý sai nguyên tắc đối với hành lý của hành khách để từ đó triển khai những giải pháp hiệu quả nhằm khắc phục chúng. Bằng cách tích hợp dữ liệu chuyến bay theo thời gian thực vào hệ thống quản lý hành lý, giờ đây Delta có thể tự động phát tín hiệu cảnh báo cho nhân viên phụ trách hành lý khi nào hành lý chuyển tiếp cần phải được chuyển thẳng sang một máy bay khác thay vì gửi tới hệ thống phân

loại hành lý của sân bay. Delta cũng bắt đầu chia sẻ dữ liệu theo dõi đường đi của hành lý với hành khách, nên hành khách bây giờ có thể theo dõi quá trình di chuyển của hành lý từng phút một trên ứng dụng Delta và trên website của họ.

Hành khách có thể chụp ảnh thẻ hành lý của họ thông qua tính năng "Track My Bag"(theo dõi hành lý của tôi) trên ứng dụng Delta, và sau đó có thể theo dõi đường đi của hành lý tới đích đến cuối cùng của nó. Ngay cả trong trường hợp hành lý không được chuyển đến chuyến bay chuyển tiếp như lịch trình ban đầu, hành khách vẫn tiết kiệm được nhiều thời gian vì không phải mất công truy tìm nó nữa. Nhờ tìm ra cách mới để tận dụng dữ liệu lớn nhằm mang lại lợi ích cho hành khách, Delta có được vị trí nổi bật trong một thị trường cạnh tranh.

Kết quả

Bằng cách đưa dữ liệu vào lĩnh vực vận hành vốn trước đây vẫn có hàm lượng công sức lao động của con người cao, từ năm 2007 tới nay, Delta đã giảm tỉ lệ xử lý sai sót hành lý tới 71% – đây là con số cao nhất trong lĩnh vực hàng không – và cải thiện trải nghiệm khách hàng. Hiện nay Delta là một hãng hàng không toàn cầu dẫn đầu thế giới về hiệu suất xử lý hành lý. Năm 2015, Delta là hãng hàng không xếp thứ nhất về

hiệu suất xử lý hành lý theo danh sách bình chọn của Bộ Giao thông Hoa Kỳ sau khi liên tiếp lập kỷ lục mới hằng tháng và tiến tới lập kỷ lục cả năm trong danh sách của bộ này.

Delta đã sử dụng những dữ liệu cao cấp hơn nhằm thu hút khách hàng hiệu quả hơn và gây dựng sự trung thành của khách hàng. Hãng hàng không này đã kết hợp dữ liệu khách hàng từ các giao dịch mua vé máy bay, các tuyến đường họ đã đi và hoạt động chi tiêu qua thẻ tín dụng của họ để từ đó tạo nên chân dung khách hàng cùng với thói quen đi lại, năng lực chi tiêu, và thậm chí là tên công ty mà họ đang làm việc. Sau đó, công ty sử dụng dữ liệu này để cẩn thận điều chỉnh các chương trình khuyến mãi và nhắm mục tiêu tiếp cận những đối tượng mà họ cho là có thể mang lại nhiều cơ hội hơn.

INTEL

Tiết kiệm chi phí sản xuất với dữ liệu lớn

Thông tin chung

Tập đoàn Intel là nhà sản xuất chip bán dẫn lớn nhất thế giới có trụ sở tại Mỹ. Được thành lập năm 1968, Intel là nhà phát minh của chuỗi kiến trúc tập lệnh x86 trong các bộ vi xử lý của hầu hết mọi chiếc máy vi tính cá nhân. Intel cung cấp bộ xử lý cho các nhà sản

xuất hệ điều hành máy tính như Apple, Lenovo (trước đây là IBM), Hewlett Packard, và Dell. Intel còn sản xuất các bộ bảng mạch, bộ điều khiển giao diện mạng lưới, và mạch tích hợp, bộ nhớ flash, chip đồ họa, các bộ xử lý nhúng, và nhiều thiết bị khác liên quan đến viễn thông và điện toán.

Intel có 11 nhà máy sản xuất cùng sáu cơ sở lắp ráp và kiểm định đặt rải rác ở khắp nơi trên thế giới. Kể từ sau khi thành lập năm 1968, công ty này đã làm thay đổi đáng kể thị trường toàn cầu; họ đã phát minh ra bộ vi xử lý, vốn được mệnh danh là "chiếc máy tính trên một con chip", nhờ vậy những chiếc máy tính cầm tay và máy vi tính cá nhân đầu tiên mới có cơ hội ra đời.

Cơ hội/thách thức

Vài năm trước, Intel bắt tay vào khám phá các cơ hội để có thể tận dụng kho dữ liệu khổng lồ mà họ đã tích lũy được trong suốt nhiều năm qua. Không lâu sau, họ nhận ra rằng với sự hỗ trợ của dữ liệu lớn và công nghệ phân tích dự đoán, họ hoàn toàn có thể giảm bớt thời gian sản xuất các con chip.

Phương pháp

Intel phát hiện ra rằng có một kho giá trị lớn ẩn giấu trong dữ liệu lớn. Trong hai năm trở lại đây, công ty

này đã phát triển hơn mười dự án liên quan đến dữ liệu và nhờ đó gia tăng được hiệu suất hoạt động cũng như lợi nhuận.

Intel chạy rất nhiều bài kiểm định phức tạp đối với từng con chip trong quá trình sản xuất. Và trong khi phát triển các loại chip mới, quy trình này đã giúp họ phát hiện ra vô số các lỗi và tìm cách khắc phục chúng.

Mỗi con chip Intel làm ra đều phải trải qua một cuộc kiểm tra chất lượng bao gồm một loạt các bài kiểm định toàn diện. Bằng cách phân tích dữ liệu lịch sử thu thập trong quá trình sản xuất, Intel có thể giảm số lượng các bài kiểm định cần thực hiện. Những bài kiểm định này bao gồm một số thông tin đến từ quá trình sản xuất cho các con chip trước khi ra mắt thị trường, và nhìn vào dữ liệu này, thay vì đưa từng con chip qua 19.000 bài kiểm định, Intel chỉ cần tập trung vào những con chip cụ thể để giảm bớt thời gian kiểm định.

Các quy trình vận hành dựa vào dữ liệu cũng giúp Intel phát hiện được những sai sót trong dây chuyền sản xuất, vốn là một môi trường vận hành trên cơ sở các khâu việc đều được tự động hóa. Sử dụng các file nhật ký lấy từ máy móc sản xuất và kiểm định trong toàn bộ mạng lưới nhà máy, Intel có thể xử lý khoảng 5 terabyte mỗi giờ.

Kết quả

Bằng cách thu thập và phân tích các thông tin này, Intel biết được khi nào thì một khâu cụ thể trong những quy trình sản xuất của họ bắt đầu hoạt động bất thường. Chính nhờ quy trình phân tích dự đoán này, vốn được triển khai trên một dòng xử lý lõi Intel Core năm 2012, Intel đã tiết kiệm được 3 triệu đô-la chi phí sản xuất. Trong giai đoạn 2013-2014, Intel mở rộng quy trình này áp dụng cho nhiều dòng chip khác và tiết kiệm được thêm 30 triệu đô-la nữa.

Dữ liệu lớn cũng mang lại nhiều lợi ích cho khía cạnh an ninh của Intel. Công ty này cho biết nền tảng dữ liệu lớn của họ có thể xử lý 200 tỉ sự kiện do máy chủ gửi tới, và có thể đưa ra những cảnh báo sớm về các mối đe dọa an ninh trong vòng 30 phút.

TXU ENERGY

Tiết kiệm chi phí nhờ sử dụng các thiết bị đo điện thông minh

Thông tin chung

Tên chính thức trước đây của TXU Energy là Texas Utilities. Kể từ năm 2002, khi thị trường năng lượng ở bang Texas được gỡ bỏ các hàng rào quy định, TXU đã trở thành nhà cung cấp năng lượng hàng đầu khu vực. Công ty này có lịch sử hoạt động kéo

dài hơn 100 năm. Công ty Chiếu sáng Điện Dallas thành lập năm 1882 chính là một tiền thân gián tiếp của TXU Energy.

TXU Energy là một nhà cung cấp lẻ điện tiêu dùng hàng đầu thị trường, đây là nhà bán lẻ điện lớn nhất cho người dân sống ở bang Texas. TXU Energy đưa ra nhiều sản phẩm và giải pháp mang tinh thần đổi mới sáng tạo cho cả khách hàng cá nhân lẫn khách hàng doanh nghiệp để họ tự do lựa chọn sản phẩm đáp ứng tốt nhất các nhu cầu của mình, bao gồm dịch vụ khách hàng xuất sắc, các gói dịch vụ điện với mức giá cạnh tranh, các phương án tiết kiệm năng lượng đầy sáng tạo, các chương trình năng lượng tái chế, cùng nhiều sản phẩm và dịch vụ khác liên quan đến điện.

Cơ hội/thách thức

Thị trường năng lượng vốn có áp lực cạnh tranh cao độ, và với lỗ hổng cung-cầu như vậy, chi phí là một trong những yếu tố then chốt để khách hàng quyết định tiếp tục sử dụng hay chuyển sang nhà cung cấp năng lượng khác. Khi các luật mới ra đời gỡ bỏ chiếc vòng kim cô ra khỏi nhiều thị trường ở khắp nơi trên thế giới, ngành dịch vụ tiện ích công cũng phải đối mặt với một thách thức mới. Họ phải cung cấp được dịch vụ tuyệt hảo với giá thành thấp, nếu không khách hàng có thể dễ dàng chuyển sang một nhà cung cấp khác đang hoạt động trên thị trường.

Trước đây, các nhà cung cấp năng lượng ghi lại mức độ sử dụng năng lượng của người tiêu dùng hằng tháng, và như vậy, họ chỉ có được thông tin chung chung về việc tiêu thụ năng lượng của các hộ gia đình. Tuy nhiên, mức độ tiêu thụ điện vào những thời điểm khác nhau trong ngày rất đa dạng. Điều này mang lại cơ hội cho TXU giới thiệu thiết bị đo thông minh để người tiêu dùng có thể tìm hiểu kỹ hơn về thói quen sử dụng năng lượng trong suốt cả ngày của chính họ và sau đó đưa ra những lời khuyên dựa trên cơ sở vững chãi của dữ liệu để họ biết cách tiết kiệm năng lượng và các chi phí khác đi kèm.

Phương pháp

Thiết bị đo thông minh của TXU gửi báo cáo gần theo thời gian thực về tình hình sử dụng điện cho các máy chủ của TXU hoạt động trên đám mây. Gần như cứ sau mỗi 15 phút lại có dữ liệu mới về tình hình tiêu thụ điện được lưu trữ trong máy chủ, và một công cụ phân tích sẽ sử dụng dữ liệu này để tính ra chi phí. Khi áp dụng phân tích chi tiết vào luồng dữ liệu đến từ các bộ đo này, người ta có thể nhận về rất nhiều kết quả. Chẳng hạn, phân tích dữ liệu có thể xác định được tải lượng điện của từng thiết bị riêng lẻ để làm cơ sở cho những lời khuyên cụ thể về cách tiết kiệm năng lượng.

Người tiêu dùng có thể truy cập vào cổng website dành riêng cho họ và mở ứng dụng trên di động để xem thông tin chi tiết về tình hình tiêu thụ năng lượng của mình, trong đó hiển thị các đồ thị được trình bày rất đơn giản và dễ hiểu để họ làm theo những lời khuyên dành riêng cho mình. Để trình bày được kết quả phân tích phù hợp theo dạng thức dễ hiểu, TXU đã thuê Bidgely, một công ty khởi nghiệp chuyên cung cấp dịch vụ giúp các công ty cung cấp dịch vụ tiện ích công đáp ứng được các mục tiêu năng lượng đối với phía khách hàng bằng cách thu hút sự tham gia và làm hài lòng khách hàng của họ thông qua các giải pháp phân tích dữ liệu.

Vì mục đích này, TXU đã thiết lập mối quan hệ đối tác với Bidgely vào năm 2014.

Hình 6.1 cung cấp một số bức ảnh chụp màn hình ứng dụng di động mà khách hàng sử dụng điện do TXU cung cấp đang dùng để tìm hiểu những kiến giải sâu sắc đằng sau thói quen tiêu thụ năng lượng của họ và tiết kiệm chi phí tiêu thụ năng lượng.

HÌNH 6.1 Ví dụ mô phỏng về bảng báo cáo trên thiết bị di động

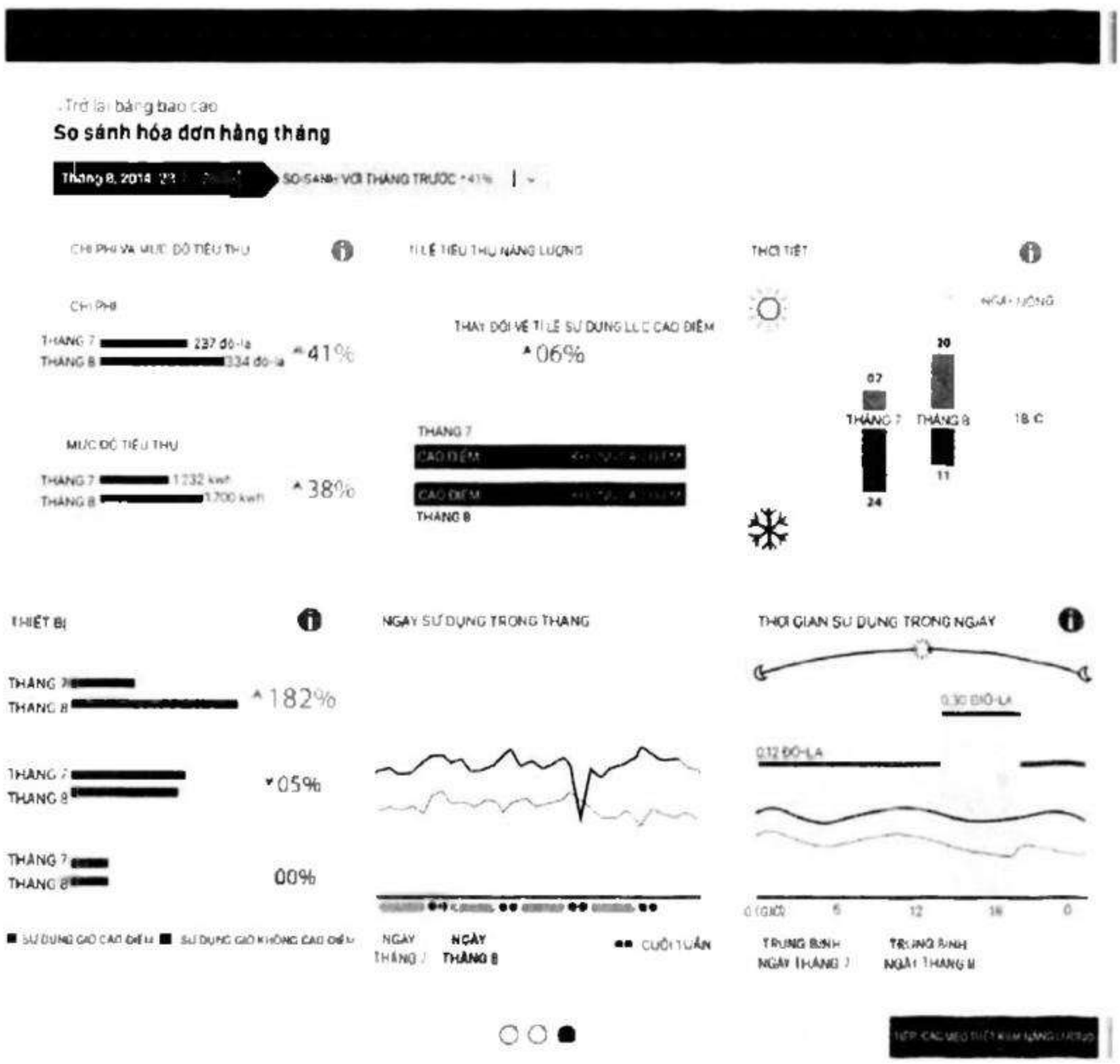


Chỉ cần một vài động tác lướt màn hình, khách hàng đã có thể biết:

- Họ đang tiêu thụ bao nhiêu năng lượng cho cả gia đình
- Mỗi thiết bị lớn trong gia đình tiêu thụ hết bao nhiêu năng lượng
- Số điện hiện tại của họ là bao nhiêu
- Số lượng điện đã sử dụng, hoặc số tiền đã chi ra cho điện năng, tính theo thời điểm trong ngày, và
- Chi phí tiền điện trong tháng này khoảng bao nhiêu

Ngoài nội dung như phần ảnh chụp màn hình ở trên, khách hàng còn có thể truy cập vào trang chủ Home Agent để thực hiện so sánh hóa đơn tiền điện và mức độ tiêu thụ năng lượng theo tháng, qua đó có thể biết rõ họ đã tiết kiệm được bao nhiêu chi phí dành cho năng lượng.

HÌNH 6.2 Ví dụ mô phỏng về bảng báo cáo tóm tắt trên trang chủ



Kết quả

Thông qua việc triển khai các thiết bị đo thông minh và bảng phân tích đi kèm (xem Hình 6.2), TXU đã trao quyền kiểm soát việc tiêu thụ năng lượng cho khách hàng của mình. Khách hàng cảm thấy họ được cung cấp thông tin đầy đủ hơn, tiết kiệm được chi phí tiêu thụ điện năng, và gắn bó với TXU trong thời gian lâu dài hơn thay vì nghĩ đến chuyện chuyển sang nhà cung cấp khác.

Với quyết định triển khai chương trình trên, TXU đã đi trước một bước và giới thiệu cho khách hàng một số ưu đãi nhằm định hình lại đường cong nhu cầu cho khách hàng. Một trong những ưu đãi như vậy là "Miễn phí điện năng vào ban đêm – Suốt đêm. Mọi đêm. Kéo dài cả năm".

Trên thực tế, trong nội dung truyền thông của mình, TXU thuyết phục khách hàng "Hãy giặt giũ hoặc chạy máy rửa bát vào ban đêm, bạn sẽ được miễn phí tiền điện." Mục đích TXU Energy nhắm đến là dùng chính sách định giá để định hình lại nhu cầu năng lượng nhằm quản lý tốt hơn nhu cầu năng lượng vào giờ cao điểm, giúp cả TXU và khách hàng đều tiết kiệm được tiền.

Dĩ nhiên, tất cả những chuyện này sẽ không thể diễn ra nếu không có các thiết bị đo điện thông minh và các kỹ thuật phân tích dữ liệu.

OMEDARX

Dữ liệu lớn củng cố sự tuân thủ các hướng dẫn điều trị y khoa

Thông tin chung

OmedaRx là công ty quản lý lợi ích dược phẩm của Cambia Health Solutions, công ty mẹ của Regence Blue Cross Blue Shield, và quản lý các chương trình thuốc kê đơn cho những thành viên thuộc gói chăm sóc sức khỏe của Regence ở Oregon, Washington, Utah, và Idaho.

Sự phát triển của nhiều công ty dược phẩm phụ thuộc vào năng lực phân tích dữ liệu sức khỏe để qua đó bảo đảm rằng người bệnh có được loại thuốc phù hợp nhất và mang lại hiệu quả tốt cho họ khi sử dụng. OmedaRx cũng là một công ty dược phẩm như vậy. Tuy nhiên, việc đào sâu nghiên cứu kho dữ liệu thay vì chỉ hành động dựa trên những thông tin chung chung ở mức tổng quát là một nhiệm vụ quan trọng.

Cơ hội/thách thức

Theo Cơ quan đặc trách Nghiên cứu và Chất lượng Y tế của Bộ Y tế và Dịch vụ Nhân sinh Hoa Kỳ, không tuân thủ đúng hướng dẫn điều trị là một hiện tượng tương đối phổ biến. Các nghiên cứu đã chỉ ra rằng 20-30% đơn thuốc không được mua đầy đủ, và tính

trung bình, 50% đơn thuốc điều trị các bệnh mãn tính không được tuân thủ đúng như hướng dẫn. Việc không tuân thủ hướng dẫn điều trị khiến người Mỹ phải chi thêm một khoản chi phí chăm sóc sức khỏe không đáng có vào khoảng 290 tỉ đô-la, chiếm 13% trong tổng số các khoản chi phí chăm sóc sức khỏe của họ – đây là số liệu của Viện Chăm sóc Sức khỏe New England, một tổ chức phi lợi nhuận hoạt động độc lập.

Các phương pháp truyền thống giúp xác định bệnh nhân cần tiếp cận không nhắm đến những người có nguy cơ cao gặp phải những hệ quả tiêu cực về sức khỏe. Với dữ liệu lớn và các công nghệ phân tích tiên tiến hiện đại, có thể xác định rằng chỉ có 700 người thực sự đang gặp rủi ro thay vì phải thực hiện các biện pháp can thiệp với 5.000 người như cách làm trước đây của OmedaRx.

Phương pháp

Sự phát triển và mức độ sẵn có của dữ liệu là hai yếu tố chủ đạo mở ra cơ hội cho sự ra đời của các ứng dụng phân tích giúp điều chỉnh và hoàn thiện các chương trình tuân thủ hướng dẫn điều trị y khoa. Một chìa khóa quan trọng nữa là công nghệ máy học, đưa ra được những câu hỏi có thể hỏi và trả lời.

Một thời gian trước đây, OmedaRx bắt tay vào thử nghiệm chương trình Max for Medication Adherence

(tối đa hóa sự tuân thủ hướng dẫn điều trị y khoa) của GNS Healthcare, một nhà cung cấp dịch vụ phân tích mà Cambia Health Solutions là một nhà đầu tư lớn trong đó. Hệ thống này sử dụng dữ liệu lớn để xác định ra những cá nhân có nguy cơ phải chi tiêu nhiều khoản đất đỏ để mua thuốc bởi vì họ không tuân thủ đúng hướng dẫn điều trị về thời điểm và liều lượng dùng thuốc.

Chương trình này bắt đầu với việc OmedaRx chia sẻ với GNS Healthcare bảng tin điện tử từ dữ liệu về đơn thuốc của họ, sau đó GNS Healthcare lại kết hợp dữ liệu này với bảng tin điện tử từ dữ liệu về đơn thuốc và tình hình bệnh nhân từ Cambia (và các công ty bảo hiểm của họ như Regence BCBS) cùng với thông tin nhân khẩu học của bệnh nhân từ các nhà cung cấp bên thứ ba, trong đó có cả các trung tâm xử lý thông tin. Phần lớn dữ liệu OmedaRx gửi cho GNS Healthcare đều là dữ liệu có cấu trúc, nhưng GNS Healthcare cũng có thể làm việc với dữ liệu phi cấu trúc.

Dữ liệu này sau đó được đưa vào Nền tảng Giải pháp Max của GNS Healthcare có sử dụng một số ứng dụng lập trình có khả năng xử lý các điểm đặc thù của từng tập dữ liệu. Từ đó, dữ liệu kết hợp này lại được dùng để phát triển các mô hình điện toán.

Các tập dữ liệu và nền tảng phân tích trên được lưu trong kho dữ liệu của GNS Healthcare tên là MeasureBase. Các chuyên gia phân tích có thể truy vấn kho dữ liệu thông qua ngôn ngữ lập trình Measure Language của chính GNS Healthcare. Công ty này sử dụng thứ ngôn ngữ rõ ràng và rành mạch để làm việc bởi theo họ, ngôn ngữ này giúp họ dễ dàng nêu chi tiết về những khía cạnh mà họ muốn đo lường, thực hiện đo lường với những người nào, và trong những khoảng thời gian nào. Nền tảng của GNS Healthcare có thể tổng hợp hàng nghìn tỉ điểm dữ liệu đến từ kho dữ liệu lịch sử về các đơn thuốc, các hồ sơ y khoa điện tử, dữ liệu về kinh tế, xã hội, và địa lý, dữ liệu về hành vi của người tiêu dùng, dữ liệu về bộ gen, dữ liệu về thông tin sinh học, và vô số các đề tài khác.

Dữ liệu được cho chạy qua công cụ máy học và mô phỏng mang tên Reverse Engineering and Forward Simulation (REFS – kỹ thuật đảo ngược và mô phỏng chuyển tiếp) cài đặt trong Nền tảng Giải pháp Max của GNS Healthcare. Ứng dụng này làm nhiệm vụ phân tích và lập mô hình cho các tập dữ liệu như các quan sát đa chiều về bệnh nhân theo thời gian. Công cụ REFS học bằng cách sử dụng kỹ thuật đảo ngược đối với các tập mô hình, và sau đó mô phỏng các phép biểu diễn tương tự để từ đó đưa ra những dự đoán, bao gồm dự đoán về mức độ rủi ro khi gặp phải các kết quả tiêu cực.

GNS Healthcare thường tạo ra các báo cáo và mô hình phân tích cho OmedaRx. Dữ liệu này có thể được trình bày trong các bảng báo cáo tóm tắt, hoặc theo danh sách phân tầng các cá nhân; dữ liệu cũng bao gồm cả thông tin về rủi ro và các yếu tố khác xuất phát từ những thay đổi về hành vi và hành vi trong điều trị bệnh của bệnh nhân.

Kết quả

Đội ngũ quản lý chăm sóc sức khỏe của OmedaRx sử dụng dữ liệu này để tạo ra các chương trình tiếp cận bệnh nhân hằng ngày, hằng tuần, và hằng tháng, bao gồm các phương thức liên lạc như gọi điện thoại, tư vấn trực tiếp tận nơi, gửi email... Qua đó, họ có cơ hội để trò chuyện với từng người bệnh, tìm hiểu rõ hơn về các mối lo lắng của họ, và đưa ra những đề xuất về cách sử dụng đơn thuốc nhằm củng cố hơn nữa mức độ tuân thủ hướng dẫn điều trị.

Sử dụng kỹ thuật phân tích dữ liệu lớn, OmedaRx có thể xem xét các dữ liệu về đơn thuốc và tình hình bệnh, lịch sử các lần bệnh nhân đến gặp bác sĩ, lịch sử các lần phải vào phòng cấp cứu, các lần nhập viện, các loại thuốc đã kê đơn, tình hình tuân thủ hướng dẫn điều trị, và dữ liệu của chính họ đối với từng bệnh nhân, và dự đoán nguy cơ phải nhập viện hoặc cấp cứu của bệnh nhân. Từ những người này, OmedaRx

có thể dễ dàng xác định những người hiện đang không tuân thủ hướng dẫn dùng thuốc.

JOHN DEERE

Cách mạng hóa nông nghiệp với dữ liệu lớn

Thông tin chung

Deere and Company (tên thương hiệu là John Deere) là một tập đoàn của Mỹ chuyên sản xuất các loại máy móc phục vụ nông nghiệp, lâm nghiệp, và xây dựng; các loại máy diesel; các hệ thống truyền động (trục, bộ truyền động, hộp số) sử dụng trong các trang thiết bị nặng và các trang thiết bị cắt cỏ. Năm 2016, John Deere đứng thứ 97 trong danh sách Fortune 500 xếp hạng các công ty lớn nhất nước Mỹ, và đứng thứ 364 trong danh sách Fortune Global 500 xếp hạng các công ty lớn nhất toàn cầu. Kể từ khi thành lập năm 1837, John Deere đã trải qua nhiều thay đổi trong hoạt động kinh doanh cũng như trong các sản phẩm và dịch vụ của họ.

Ngành nông nghiệp hiện đang trải qua một cuộc cách mạng số. Chẳng hạn, ngay cả các nông dân hoạt động ở quy mô nhỏ bây giờ cũng đang lấy thông tin được thu thập thụ động bởi các trang thiết bị nông nghiệp chính xác, và nhiều nông dân sử dụng thông tin từ các tập dữ liệu lớn và kỹ thuật phân tích chính xác để đưa ra những quyết định tức thời ở nông trại.

Cơ hội

Trong suốt hơn một thế kỷ qua, ngành nông nghiệp chủ yếu hoạt động dựa trên kinh nghiệm, nhưng dữ liệu thu về lại không ở dạng kỹ thuật số. Các trung tâm nghiên cứu thuộc Bộ Nông nghiệp Canada (khoảng thập niên 1920) đã tỉ mẩn tính toán sản lượng lúa mỳ ở các trang trại và xu hướng thời tiết nhằm gia tăng hiệu quả sản xuất. Điểm khác biệt của dữ liệu lớn so với cách thu thập thông tin truyền thống này nằm ở khối lượng dữ liệu và tiềm năng phân tích có sẵn trong các công nghệ kỹ thuật số đương đại. Những người ủng hộ dữ liệu lớn khẳng định rằng dữ liệu lớn là một nấc thang phát triển mới về sự chính xác, khả năng lưu trữ thông tin, năng lực xử lý và phân tích chưa từng có trước đây do những hạn chế về công nghệ.

Nổi tiếng với danh mục sản phẩm mang đậm tinh thần đổi mới sáng tạo, John Deere nhận ra rằng nếu các thiết bị của họ có thể thu thập dữ liệu thời gian thực, khi đó luồng dữ liệu ấy sẽ được sử dụng cho nhiều mục đích khác nhau chứ không chỉ dừng lại ở việc cung cấp thông tin để họ thực hiện những thay đổi về thiết kế và chức năng của bản thân thiết bị.

Phương pháp

John Deere sử dụng dữ liệu lớn để bước chân vào tương lai của ngành nông nghiệp nhằm giúp nông

dân đạt được mục tiêu tham vọng này. Năm 2012, họ giới thiệu một số sản phẩm có thể kết nối các thiết bị của John Deere với nhau và với chủ sở hữu thiết bị, người vận hành thiết bị, đại lý bán thiết bị, và các chuyên gia cố vấn nông nghiệp. Sự liên kết này giúp nông dân gia tăng năng suất và hiệu suất mùa màng. John Deere sử dụng các bộ cảm biến gắn trên sản phẩm xe kéo để giúp nông dân quản lý đoàn xe của họ tốt hơn, giảm thời gian máy nghỉ, đồng thời tiết kiệm nhiên liệu. Thông tin này được kết hợp với các dữ liệu thời tiết (cả dữ liệu lịch sử và dữ liệu thời gian thực), dữ liệu về tình trạng đất canh tác, dữ liệu về các đặc điểm của mùa màng, và nhiều tập dữ liệu khác.

Nền tảng MyJohnDeere hoạt động trên điện thoại di động, có thể dùng với iPad; ứng dụng này giúp nông dân xác định các xu hướng trồng trọt để đưa ra quyết định về loại hoa màu sẽ trồng trong năm, vòng đời của nó, và các sự kiện liên quan đến nó. Sử dụng ngôn ngữ lập trình R, họ có thể dự đoán nhu cầu theo từng mùa, sản lượng, và ước tính được thu nhập cho nông dân.

Kết quả

John Deere hiện đang mang lại những giá trị vô cùng lớn lao cho tất cả những nông dân sử dụng nền tảng MyJohnDeere – nền tảng này cung cấp rất nhiều lợi

ích, chẳng hạn nó giúp gia tăng hiệu suất vận hành của các loại máy móc đang sử dụng, cải thiện chu kỳ sản xuất, và cung cấp thông tin tài chính liên quan đến nông nghiệp.

Hệ thống quản lý trang trại di động cung cấp thông tin lịch sử và thông tin theo thời gian thực về các cánh đồng, thực hiện lấy mẫu đất, và người dùng có thể chia sẻ trực tiếp thông tin này với các chuyên gia cố vấn để được họ hỗ trợ từ xa ngay trong lúc đang làm việc trên cánh đồng.

AIRBNB

Gợi ý giá bằng dữ liệu lớn

Thông tin chung

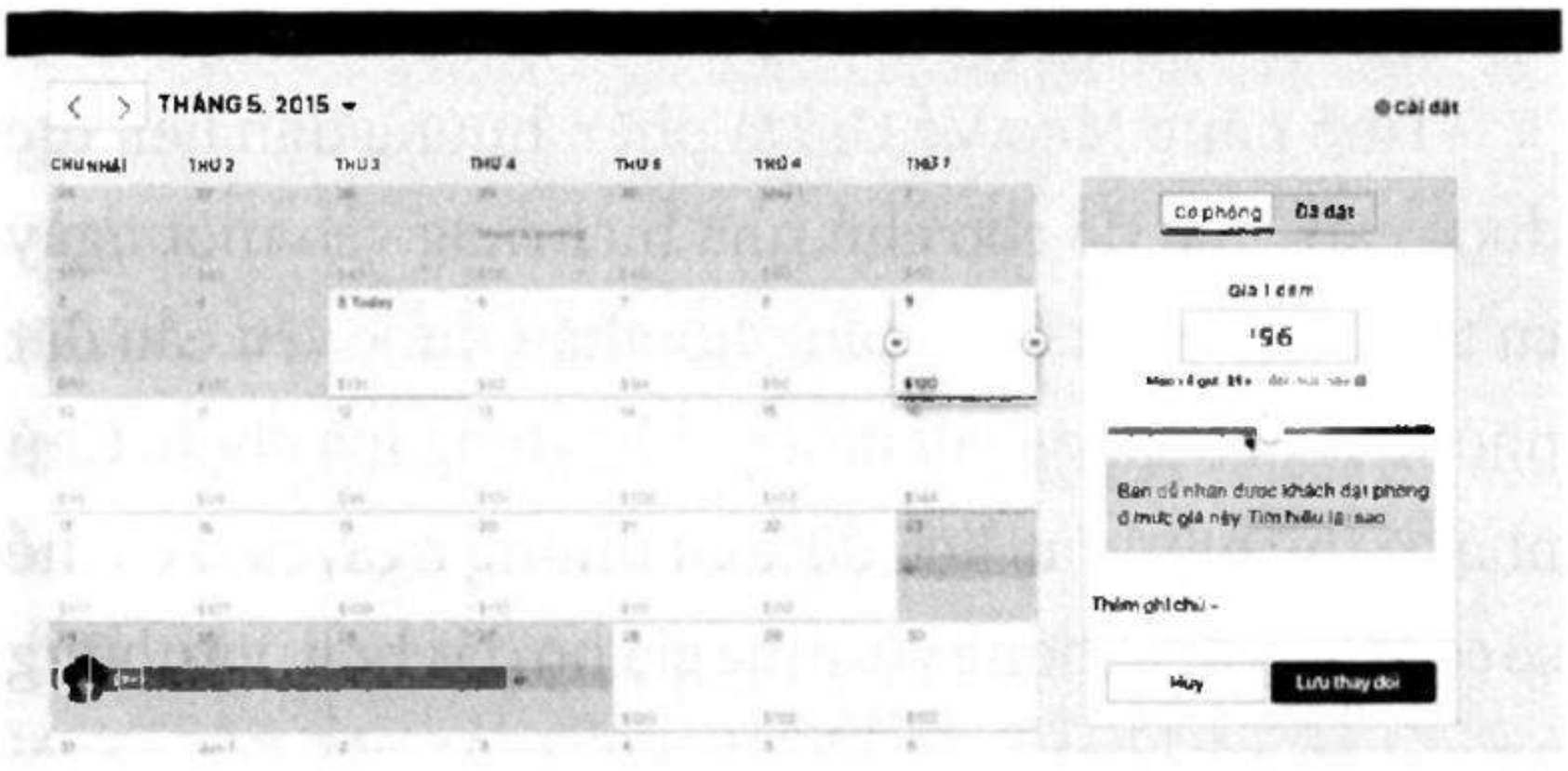
Airbnb được thành lập vào tháng 8 năm 2008, có trụ sở tại San Francisco, California. Airbnb là thị trường trực tuyến ngang hàng và là mạng lưới lưu trú cho phép người dùng đăng ký cho thuê hoặc thuê ngắn hạn phòng trống trong nhà, với chi phí do chủ nhà đặt ra (còn gọi là hình thức lưu trú homestay). Airbnb nhận phí dịch vụ tính theo phần trăm giao dịch từ cả hai phía khách thuê và chủ nhà. Hiện nay, trên nền tảng của công ty này có tới 2.000.000 hạng mục cho thuê đang rao mời khách hàng ở 34.000 thành phố và 191 quốc gia.

Trong lĩnh vực lữ hành, định giá chịu ảnh hưởng của nhiều yếu tố khác nhau – cả yếu tố nội bộ và bên ngoài – và việc tiếp cận nguồn dữ liệu đó có thể mang lại lợi ích lớn cho các công ty như Airbnb, vốn tạo ra doanh thu từ các nhà cung cấp dịch vụ và khách hàng. Airbnb còn được hưởng lợi ích lớn nếu các chủ nhà luôn có khách đến thuê.

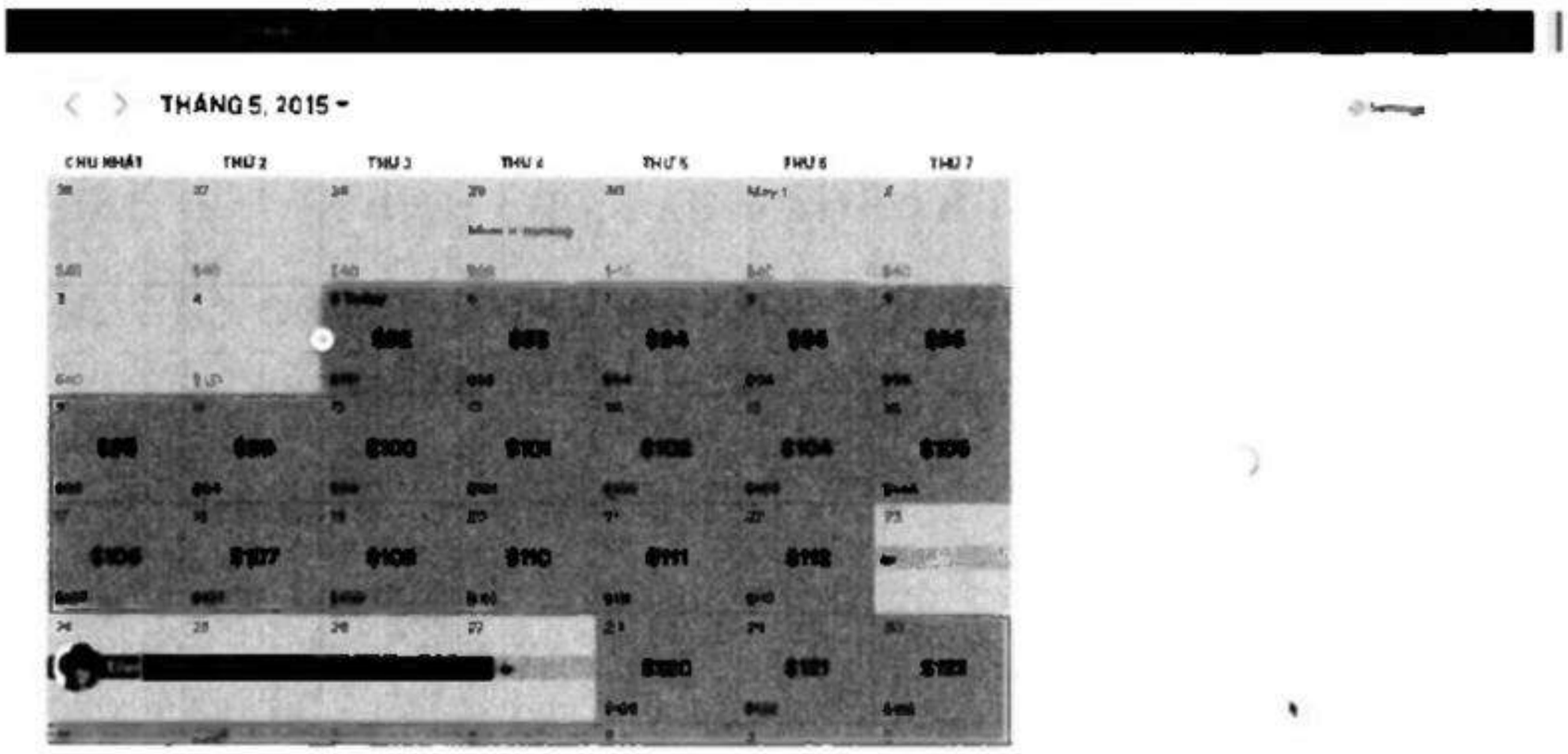
Cơ hội

Có lẽ nếu Airbnb có thể sử dụng công nghệ phân tích dự đoán để đưa ra lời khuyên thiết thực cho chủ nhà về điểm giá phù hợp, khi đó chủ nhà sẽ dễ dàng thu hút khách tìm đến hơn, và Airbnb cũng kiếm được thêm doanh thu mà không cần mở rộng để tìm thêm chủ nhà và nguồn lực mới.

HÌNH 6.3 Ví dụ về màn hình đăng ký của Airbnb



HÌNH 6.4 Ví dụ về màn hình đăng ký – chức năng xem lịch của Airbnb



Phương pháp

Airbnb đã dành nhiều tháng trời để phát triển công
cụ gợi ý mức giá. Công cụ này xử lý 5 tỉ điểm dữ liệu
huấn luyện và có hai hợp phần chính: lập mô hình và
máy học. Mô hình này tổ chức và sắp xếp những phát
hiện rút ra được từ tập dữ liệu khổng lồ của Airbnb
về mức giá tối ưu dựa trên những yếu tố như khu vực
xung quanh phòng cho thuê và diện tích phòng thuê.

Tính năng Mẹo Về Giá là phần hướng dẫn liên tục
được cập nhật để cho chủ nhà biết rằng vào một ngày
cụ thể trong năm, khả năng họ nhận được yêu cầu đặt
phòng là bao nhiêu với mức giá họ đang lựa chọn. Chủ
nhà có thể nhìn vào lịch để xem những ngày nào có thể
sẽ có khách đặt phòng với mức giá họ đã chọn, và những
ngày nào sẽ không có khách với mức giá đó; ngoài
ra, họ cũng nhận được nhiều gợi ý về cách định giá.

Khi chủ nhà đặt giá du di trong khoảng 5% so với giá gợi ý, khả năng nhận được yêu cầu đặt phòng của họ sẽ tăng lên "gần bốn lần". Phần mẹo về giá được trình bày trên một giao diện dễ sử dụng, nhưng đằng sau nó là cả một quy trình phức tạp, tính toán tất cả mọi thứ, từ ngày trong tuần tới khu vực xung quanh nơi có phòng cho thuê, đặc điểm địa hình giữa các vĩ độ và kinh độ, hay những từ khóa như "bãi biển".

Giá trị của các tiện ích trong phòng cho thuê cũng có nhiều sự khác biệt đáng kể. Phòng có trang bị Wi-Fi ở San Francisco sẽ nhanh chóng có khách đặt thuê, trong khi đó, ở những nơi có nền khí hậu nóng hơn, số lượng khách đặt thuê phòng sẽ giảm mạnh nếu trong phòng không lắp điều hòa – tuy nhiên, điều này cũng phụ thuộc vào mùa trong năm nữa. Chủ Nhật là ngày ít nhận được yêu cầu đặt phòng nhất trong tuần, và giá cả có xu hướng tăng dần từ đầu tuần cho đến thứ Bảy, ngày được ưa thích nhất.

Ngoài những yếu tố phổ biến mà Airbnb biết đến để lập mô hình, nhằm tìm kiếm thêm những mối quan hệ giữa phòng cho thuê và mức giá, Airbnb phát triển Aerosolve, một chương trình máy học ra mắt vào tháng 5 năm 2016.

Kết quả

Với Aerosolve, Airbnb có thể phát hiện ra nhiều xu hướng mới, sau đó họ sử dụng các thông tin này để

tìm hiểu kỹ hơn xem đâu là những yếu tố khiến chủ nhà có thể đặt giá thuê phòng ở một mức giá nhất định. Chẳng hạn, mô hình này phát hiện ra rằng các phòng cho thuê tọa lạc ở một vị trí kinh độ và vĩ độ nhất định đang yêu cầu mức giá tương đối cao và các chủ nhà ở đó thường xuyên sử dụng từ "sabbia" trong nội dung quảng cáo phòng. Hóa ra nơi đó chính là Playa del Carmen, một thị trấn ven biển rất nổi tiếng ở Mexico, và từ "sabbia" trong tiếng Italy có nghĩa là "cát" – vậy là Airbnb có thêm một chút thông tin gợi ý về giá để chuyển cho các chủ nhà sống ở khu vực đó.

Với sự hỗ trợ của Aerosolve, Airbnb đã giúp các chủ nhà của họ gia tăng thu nhập, nghĩa là doanh thu của công ty này cũng tăng trưởng theo.

WALMART

Tìm kiếm thông minh với kỹ thuật phân tích ngữ nghĩa

Thông tin chung

Walmart là một tập đoàn bán lẻ đa quốc gia của Mỹ có trụ sở tại Bentonville, Arkansas. Tập đoàn này sở hữu một chuỗi các đại siêu thị, cửa hàng giảm giá, và cửa hàng bán đồ tạp hóa. Walmart do Sam Walton thành lập năm 1962 và chính thức đăng ký kinh doanh vào ngày 31 tháng 10 năm 1969.

Walmart là một công ty gia đình, được niêm yết trên Sàn giao dịch Chứng khoán New York năm 1972. Vào giai đoạn cuối thập niên 1980 đầu thập niên 1990, Walmart từ địa vị một công ty hoạt động trong vùng vươn lên trở thành một tập đoàn lớn với quy mô hoạt động trên toàn nước Mỹ. Tới năm 1988, công ty này trở thành nhà bán lẻ kiếm được lợi nhuận cao nhất ở Mỹ, và tới tháng 10 năm 1989, trở thành công ty lớn nhất xét về mặt doanh thu. Từ thời điểm thành lập đến giữa thập niên 1980, về mặt phạm vi địa lý, hoạt động của Walmart chỉ giới hạn ở miền Nam và vùng hạ Trung Tây của Mỹ; nhưng tới đầu thập niên 1990, công ty này đã mở rộng sự hiện diện của mình trên toàn bộ lãnh thổ nước Mỹ – Câu lạc bộ của Sam ra mắt ở New Jersey vào tháng 11 năm 1989, và cửa hàng Walmart đầu tiên của California khai trương tại Lancaster vào tháng 7 năm 1990.

Theo danh sách Fortune Global 500 vào năm 2016, Walmart là công ty lớn nhất thế giới xét về mặt doanh thu, đồng thời đây cũng là đơn vị tuyển dụng tư nhân lớn nhất thế giới với hơn 2,2 triệu nhân viên.

Cơ hội

Trong giai đoạn 2011-2012, Walmart muốn cải thiện trải nghiệm mua sắm cho hàng triệu khách hàng trực tuyến của họ, đồng thời tăng nguồn doanh thu trực tuyến cho công ty.

Phương pháp

Quan sát các cửa hàng bán lẻ trực tuyến như Amazon và eBay, đội ngũ của Walmart tự tin rằng họ có thể cải thiện trải nghiệm khách hàng và chất lượng dịch vụ trực tuyến bằng cách tìm hiểu kỹ lưỡng hơn về khách hàng và giới thiệu các sản phẩm/dịch vụ mà có thể họ sẽ cần dựa trên những sở thích của họ – đây là thông tin rút ra được dựa trên hành vi của khách hàng thông qua quá trình bán hàng.

Một nhóm nhỏ tại Phòng thí nghiệm Walmart – trung tâm nghiên cứu và công nghệ phục vụ mục đích đổi mới sáng tạo tại Walmart – đã làm việc trong mười tháng để phát triển một công cụ tìm kiếm sử dụng công nghệ tìm kiếm theo ngữ nghĩa để dự đoán ý định của khách hàng khi tìm kiếm, từ đó đưa ra được những kết quả phù hợp nhất với họ. Đội ngũ kỹ thuật ở Phòng thí nghiệm Walmart bao gồm các chuyên gia thuộc nhiều lĩnh vực khác nhau như trích xuất thông tin, máy học, và khai thác văn bản; tất cả đều có kinh nghiệm làm việc ở các công ty hàng đầu về lĩnh vực tìm kiếm và thương mại điện tử cũng như trong các viện nghiên cứu nổi tiếng. Họ đặt tên cho công cụ tìm kiếm mới này là Polaris.

Đối với bất kỳ website thương mại điện tử nào, việc tìm kiếm đối với một kết quả mua sắm là một trải nghiệm rất khác so với hành vi tìm kiếm chung chung.

Polaris hoạt động dựa trên một nền tảng kết hợp con người với địa điểm, sự kiện, và sản phẩm, từ đó mang lại cho Walmart sự hiểu biết phong phú hơn về khách hàng và sản phẩm. Công cụ tìm kiếm mới này sử dụng các thuật toán tiên tiến, bao gồm các thuật toán tìm hiểu về nội dung câu hỏi và khai thác từ đồng nghĩa để xác định ý định của người dùng, từ đó đưa ra các kết quả tìm kiếm phù hợp. Khi người dùng gõ chữ "vải denim" vào thanh tìm kiếm, công cụ này sẽ trả các kết quả về quần bò, hoặc nếu gõ "thuốc chlorine", thì kết quả trả về sẽ là các thiết bị dùng ở bể bơi.

Polaris tập trung vào việc tìm hiểu cách thu hút sự tham gia của người dùng, theo đó nó tính đến cách hành xử của người dùng với website để đưa ra những kết quả tìm kiếm tốt nhất cho họ. Nó đưa ra một trang kết quả mới dễ sử dụng trong lúc vẫn duyệt tìm kiếm các đề tài chứ không đưa ra một danh sách các kết quả tìm kiếm vốn đã được áp dụng rộng rãi, từ đó cho phép khách mua hàng phát hiện ra những danh mục sản phẩm và dịch vụ mới mà có thể trước đây họ chưa từng nghĩ đến.

Kết quả

Năm 2012, trang Walmart.com gia tăng 10-15% tỉ lệ khách hàng hoàn thiện đơn hàng sau khi tìm kiếm sản phẩm bằng công cụ tìm kiếm mới; và trong quý đầu tiên của năm 2013, doanh thu từ hoạt động

thương mại điện tử của hãng này đã tăng hơn 30%, và à
Walmart vẫn tiếp tục rót những khoản đầu tư chiến l
lược trên các thị trường có cơ hội tăng trưởng tốt nhất. .

HUFFINGTON POST

Sử dụng dữ liệu lớn để tăng lưu lượng khách truy cập

Thông tin chung

Huffington Post là trang tổng hợp tin tức và blog
trực tuyến. Một trong những nhà sáng lập ra website
này là Ariana Huffington, một tác giả người Mỹ gốc
Hy Lạp, một nhà báo tổng hợp kiêm nữ doanh nhân
thành đạt. Website này cung cấp tin tức, tranh châm
biếm, blog, và các nội dung độc quyền, với phạm vi
đề tài bao quát là chính trị, kinh doanh, giải trí, môi
trường, công nghệ, truyền thông đại chúng, lối sống,
văn hóa, hài kịch, lối sống lành mạnh, các lợi ích của
phụ nữ, và tin tức địa phương.

Tại thời điểm năm 2014, website HuffingtonPost.com
có 22 triệu khách truy cập hằng tháng với 11 phiên
bản trên toàn cầu, dự kiến tới cuối năm 2015, họ sẽ
mở rộng thành 16 phiên bản. HuffingtonPost.com
trở thành một trong những website tin tức và giải trí
có lưu lượng truy cập đông đảo nhất với tốc độ tăng
trưởng nhanh nhất thế giới.

Cơ hội

Huffington Post thu hút được rất nhiều độc giả vì những bài viết có chất lượng, và nguồn dữ liệu này mang lại một kho kiến thức khổng lồ cùng những phát hiện mới mẻ để công ty này hoàn thiện dần những gì họ cung cấp ra thị trường và gia tăng doanh thu. Các lĩnh vực chính được công ty này ưu tiên hàng đầu là nội dung có chất lượng cao, các bình luận chính danh và xác thực, các nguồn thu nhập mới, và mang lại giá trị tốt nhất cho các nhà quảng cáo.

Phương pháp

Huffington Post đã sử dụng nhiều kỹ thuật phân tích dữ liệu lớn khác nhau để tận dụng cơ hội này. Năm 2013, các nội dung do Huffington Post cung cấp đã thu hút được 300 triệu bình luận từ độc giả. Số lượng bình luận lớn nhưng đôi khi lại kém chất lượng khiến họ nghĩ đến việc tìm kiếm một giải pháp thông minh cho vấn đề này.

Họ thực hiện một phân tích kết hợp – đây là một kỹ thuật thống kê dùng để xác định xem mọi người đánh giá các tính năng khác nhau của một sản phẩm hoặc dịch vụ như thế nào. Trong trường hợp các bình luận trên website của Huffington Post, phân tích kết hợp được sử dụng để xác định chất lượng của bình luận đến từ một người dùng ẩn danh hoặc những

người đã xác nhận danh tính thông qua việc sử dụng tên thật hoặc hình ảnh đại diện và từ các khu vực địa lý khác nhau. Phân tích này chỉ ra rằng chất lượng của các bình luận cao hơn hẳn khi chúng đến từ người ở gần khu vực của những người được khảo sát, và từ những người không ẩn danh.

Họ cũng sử dụng các kỹ thuật dữ liệu lớn để tối ưu hóa nội dung trong quá trình sản xuất nội dung và điều chỉnh lại vào thời điểm cung cấp nội dung.

Giống như nhiều nhà xuất bản nội dung khác, Huffington Post cũng thực hiện kiểm định A/B trên nội dung và các thiết kế của họ. Họ cũng sử dụng dữ liệu về thời điểm và tỉ lệ phản hồi, lấy đó làm các điểm dữ liệu để phân tích những ý kiến phản hồi đối với các bài viết họ xuất bản, và dựa trên cơ sở đó để đưa ra những quyết định xác đáng.

Huffington Post sử dụng dữ liệu lớn để đo lường mức độ thành công hay thất bại của các mô hình doanh thu quảng cáo mới. Công ty này phát triển Partner Studio, một nhánh của bộ phận quảng cáo phụ trách quảng cáo tự nhiên[1], hay marketing nội dung.

---

[1] Quảng cáo tự nhiên (native advertising): Là việc sử dụng các quảng cáo trả tiền phù hợp với vẻ ngoài, tinh thần, và chức năng của định dạng truyền thông nơi chúng xuất hiện. Quảng cáo tự nhiên thường xuất hiện trên bảng tin của các mạng xã hội, hoặc ở mục nội dung được gợi ý trên website. (ND)

Huffington Post hợp tác với các thương hiệu lớn để tạo ra các tiểu website tập trung vào nội dung, tận dụng đội ngũ tác giả, biên tập viên, và chuyên gia thiết kế của họ.

Kết quả

Rõ ràng, trong dự án này Huffington Post đã đạt được nhiều mục tiêu và tạo ra được những nội dung có chất lượng tốt hơn, thu hút được nhiều độc giả giá trị hơn, và gia tăng doanh thu cho công ty. Phân tích trên cũng chỉ ra rằng chất lượng của các bình luận cao hơn hẳn khi đó là bình luận của những người ở gần với người được khảo sát, và từ những người không ẩn danh. Từ những phát hiện này, Huffington Post quyết định triển khai tính năng yêu cầu người muốn gửi bình luận phải đăng ký, đồng thời không cho phép những người dùng ẩn danh được đăng tải bình luận lên website của họ. Huffington Post hợp tác với các thương hiệu lớn để tạo ra các tiểu website tập trung vào nội dung, tận dụng đội ngũ tác giả, biên tập viên, và chuyên gia thiết kế của họ. Một trong những bài đăng blog nổi tiếng của Huffington Post nhận được 24.000 lượt xem trong vòng 7 tiếng đồng hồ, và 41% lượt xem đó đến từ những lời giới thiệu trên mạng xã hội.

TÓM TẮT

Nội dung của chương này hướng đến việc học hỏi từ kinh nghiệm của người khác, qua đó chúng ta đã cùng nhau tìm hiểu xem các tổ chức và doanh nghiệp khác đang sử dụng dữ liệu lớn như thế nào để tạo ra giá trị. Chúng tôi cũng đề cập đến các doanh nghiệp hoạt động ở nhiều lĩnh vực ngành nghề khác nhau nhằm thể hiện hai ý sau:

1. Bản thân dữ liệu lớn chính là một giá trị và nó cũng tạo ra nhiều giá trị lớn lao ở bất kỳ lĩnh vực ngành nghề nào hoặc bất kỳ đơn vị nghiệp vụ nào trong một doanh nghiệp, với điều kiện nhà quản lý được trang bị các nguồn lực phù hợp.
2. Làm việc với dữ liệu lớn không phải là công việc quá khó khăn – miễn là bạn có những kiến thức cơ bản, một mô hình nghiệp vụ như C-ADAPT có thể giúp bạn đạt được những mục tiêu mà bạn muốn hướng đến trong dự án dữ liệu lớn sắp tới của mình.

Như vậy, trong chương này, chúng tôi đã đề cập đến nhiều trường hợp ví dụ rất đa dạng:

Ooredoo đã sử dụng phân tích dữ liệu lớn cho chiến dịch tái xây dựng thương hiệu và cho các mục đích PR của họ, Domino's Pizza đã hoàn thành mục

tiêu tăng trưởng doanh thu gộp, Tesco sử dụng dữ liệu lớn để có thể lập hồ sơ khách hàng tốt hơn, và Delta Airlines đã giải quyết được vấn đề phiền toái nhất khi đi lại bằng đường hàng không: truy tìm hành lý bị thất lạc.

Nhiều doanh nghiệp lớn nhỏ, đã tồn tại lâu đời hoặc mới thành lập, đều đang tận dụng dữ liệu lớn để kiến tạo giá trị – Intel sử dụng dữ liệu lớn để đẩy nhanh quy trình sản xuất và tiết kiệm được nhiều khoản chi phí lớn trong khía cạnh này, TXU giới thiệu thiết bị đo thông minh nhằm giúp khách hàng tiết kiệm tiền. Walmart cải thiện doanh thu gộp bằng các công cụ tìm kiếm thông minh hoạt động trên cơ sở công nghệ phân tích dữ liệu lớn.

Và chúng ta còn thấy một số công ty mới hơn không chỉ dùng kỹ thuật phân tích dữ liệu lớn cho doanh nghiệp của mình mà còn tạo ra những luồng dữ liệu phi cấu trúc khổng lồ, chẳng hạn Facebook và Huffington Post.

Sau khi đọc xong những trường hợp ví dụ trên, độc giả sẽ quen thuộc hơn với khía cạnh này, và họ cũng sẽ được trang bị khả năng triển khai những kiến thức vừa học được trong sách vào những dự án dữ liệu lớn của họ trong tương lai.

7

NHỮNG NGƯỜI THỰC HÀNH NÓI GÌ

Nghe thì dễ quên. Nhìn thì dễ nhớ. Làm thì dễ hiểu.

– Khổng Tử

Trong bất kỳ công việc nào cũng vậy, kinh nghiệm thực tế và những bài học rút ra từ đó luôn được coi là quan trọng hơn lý thuyết hay nghiên cứu mở. Thực ra, tất cả các bên liên quan đều đóng góp rất nhiều giá trị vào việc định hình nên "nền kinh tế dữ liệu lớn", tuy vậy, trong chương này chúng tôi sẽ chỉ nói đến quan điểm của các chuyên gia dữ liệu lớn, những người thực làm, đã xắn tay lăn lộn trong lĩnh vực này. Tính đến nay, ở các chương trước, chúng ta đã hiểu về dữ liệu, dữ liệu lớn là gì, chúng ta đã thấy các mô hình dữ liệu lớn và đã nghiên cứu nhiều trường hợp ví

dụ khác nhau để xem việc triển khai các mô hình đó trên thực tế như thế nào, chúng ta cũng đã gặp nhiều trường hợp ví dụ đến từ mọi lĩnh vực, mọi khía cạnh kinh doanh, và ở khắp nơi trên toàn cầu, và thấy dữ liệu lớn có thể mang lại giá trị to lớn đến nhường nào. Đối với nhiều doanh nghiệp, dữ liệu lớn vẫn được coi là một khái niệm tương đối mới mẻ, nhưng có nhiều công ty lãnh đạo thị trường đã đi những bước đầu tiên và hiện đang là lá cờ đầu dẫn đường cho các tổ chức còn lại thông qua việc tích cực triển khai các dự án dữ liệu lớn và vừa làm vừa học hỏi dần. Chúng ta cũng đã cùng nhau nghiên cứu một vài trường hợp ví dụ. Trong chương này, chúng ta sẽ cùng nhau tìm hiểu về suy nghĩ của những người thực hành, nghe họ kể về lần đầu tiếp xúc với dữ liệu lớn, những giá trị và thách thức mà họ nhìn thấy trên hành trình này, và tầm nhìn của họ về kỷ nguyên dữ liệu lớn trong tương lai.

Chúng tôi đã tiếp xúc với nhiều nhà quản lý và nhà lãnh đạo thông qua khảo sát dữ liệu lớn mà chúng tôi thực hiện với sự hỗ trợ của trường Kinh doanh Ashridge, ngoài ra chúng tôi cũng thực hiện các cuộc phỏng vấn cá nhân và thu thập ý kiến của họ từ nhiều nguồn khác nữa. Chúng tôi nhận được hơn 200 ý kiến phản hồi từ các nhà quản lý làm việc ở nhiều khu vực khác nhau trên thế giới, ở nhiều tổ chức rất đa dạng, từ các công ty khởi nghiệp đến các

tập đoàn lớn – và chúng tôi đã phải cẩn thận lựa chọn trong số các ý kiến phản hồi trên để đưa ra được một cái nhìn tổng quan và toàn diện. Thật thú vị khi thấy rằng mỗi người lại đang ở một giai đoạn khác nhau trong cuộc cách mạng này – một số người trong đó đã bắt đầu nhìn ra những giá trị lớn lao từ dữ liệu lớn – nhưng nhìn chung hầu hết họ đều nhất trí rằng dữ liệu lớn sẽ là một xu hướng quan trọng tiếp theo, và họ tỏ ra lạc quan về tương lai của dữ liệu lớn cũng như về những đóng góp của nó đối với tổ chức của họ.

Chúng tôi đã trao đổi về những nhận định của họ về hiện trạng của lĩnh vực phân tích dữ liệu lớn cùng các quan điểm, quy trình, và dự án liên quan đến kiến tạo giá trị trong các tổ chức mà họ đang làm việc. Kết hợp các ý kiến đó lại với nhau, chúng tôi có được một bức tranh vô cùng chính xác và thực tế trải rộng ở nhiều lĩnh vực ngành nghề khác nhau, và đây cũng là một góc nhìn rất thiết thực cho mọi nhà quản lý đang có kế hoạch đi sâu vào thế giới mênh mông của dữ liệu lớn. Chúng tôi tin rằng quan điểm của họ sẽ truyền cảm hứng để các độc giả của cuốn sách này cảm thấy tự tin khi đặt chân lên hành trình mang tên dữ liệu lớn.

DỮ LIỆU LỚN LÀ QUAN TRỌNG – RẤT QUAN TRỌNG!

> Không có phân tích dữ liệu lớn, các công ty sẽ đui mù và câm điếc, họ lò dò bước ra ngoài thế giới mạng như những chú hươu ngơ ngác trên xa lộ.
>
> – Geoffrey Moore

Tới lúc này, chúng ta đã thấy rõ rằng dữ liệu lớn đóng vai trò vô cùng quan trọng đối với hầu hết mọi doanh nghiệp trong thời đại ngày nay – một số công ty thực ra đã bắt đầu chớp lấy những cơ hội này, trong khi một số khác cũng đang chuẩn bị khởi sự những dự án dữ liệu lớn đầu tiên trong tương lai gần. Ngoài khả năng cải thiện hiệu suất hoạt động và gia tăng thu nhập cho doanh nghiệp, các ứng dụng dữ liệu lớn còn là cơ sở để tạo ra những giải pháp mang đậm tinh thần đổi mới sáng tạo cho những lĩnh vực ngành nghề vốn đã cũ kỹ và thiếu sức sống từ lâu.

Được thành lập năm 1999, Cornerstone kinh doanh các phần mềm quản lý nguồn nhân lực, khách hàng sau khi đăng ký thuê bao có thể tiếp cận chúng thông qua các trung tâm dữ liệu của Cornerstone đặt trên đám mây, và công ty này thu về 280 triệu đô-la doanh thu hằng năm. Trong số 2.100 khách hàng hiện tại của Cornerstone, có thể kể đến một số cái tên nổi tiếng như Hitachi, hãng tàu biển Princess Cruise, Đại học Bang Ohio, và công ty Fruit of the Loom.

Cornerstone OnDemand dùng dữ liệu lớn để cho doanh nghiệp biết họ nên tuyển dụng người nào. Cornerstone OnDemand đã mở ra một dịch vụ phân tích dữ liệu lớn mới, có thể giúp các tổ chức giải quyết vấn đề tuyển dụng nhân sự. Dịch vụ phân tích dữ liệu lớn mới mẻ này sẽ giúp các doanh nghiệp phân tích nguồn dữ liệu lớn khổng lồ về nhân viên để rút ra những phát hiện có ý nghĩa.

"Chúng ta đã chứng kiến thị trường quản lý nhân tài có những bước tiến ra sao trong một thập niên qua, khi các tổ chức toàn cầu dịch chuyển từ tập trung vào tự động hóa quy trình thông qua các giải pháp trên máy chủ sang tập trung vào mục tiêu gắn kết nhân viên thông qua các gói phần mềm quản lý nhân tài nhất quán. Bây giờ, chúng ta tiếp tục chuyển sang thế hệ tiếp theo của công nghệ quản lý nhân tài bằng dữ liệu trong thế giới của dữ liệu lớn," Adam Miller, chủ tịch kiêm CEO của Cornerstone OnDemand, chia sẻ.

TomTom, hãng cung cấp dịch vụ định vị bằng vệ tinh, đã biết sử dụng dữ liệu lớn một cách khôn ngoan và hiệu quả. Công ty lập bản đồ này đã tích hợp dữ liệu lớn vào hệ thống nhằm cải thiện chất lượng dịch vụ của mình trong suốt hơn một thập niên qua, và họ không có ý định chỉ dừng lại ở đó.

"Chúng tôi rất giỏi trong lĩnh vực dữ liệu lớn. Chúng tôi bắt đầu với nó từ rất sớm, sớm hơn tất cả

mọi người, và đó chính là lý do vì sao ngày nay chúng tôi dẫn đầu; chúng tôi giải quyết những nhu cầu thực tế bằng các ý tưởng đổi mới. Dĩ nhiên, chúng tôi không biết ai đang dùng điện thoại nào, nhưng điều đó không quan trọng: Nếu có hiện tượng tắc đường xảy ra, ở bất cứ đâu, chúng tôi sẽ thấy hàng trăm chiếc điện thoại di động và các thiết bị định vị xếp hàng nối đuôi nhau và ì ạch bò từng chút một. Chỉ vài giây sau khi biết điều đó, chúng tôi sẽ bắt đầu hỗ trợ các khách hàng của TomTom lúc này vẫn còn ở cách chỗ tắc đường khoảng 30 ki-lô-mét bằng cách thông báo cho họ đi vòng đường khác", nhà đồng sáng lập TomTom, Pieter Geelen, chia sẻ.

Trong tương lai, TomTom hy vọng rằng những tiến bộ trong công nghệ xung quanh dữ liệu lớn sẽ giúp cho việc chia sẻ các dữ liệu như thế này trở nên dễ dàng hơn, và không chỉ là chia sẻ dữ liệu giữa xe với trung tâm dữ liệu, mà còn chia sẻ dữ liệu giữa xe với xe, từ đó đẩy mạnh sự phát triển của các loại xe tự lái.

Các công ty dữ liệu lớn hiện đang bận rộn khám phá những giải pháp dữ liệu lớn mới mẻ để có thể tận dụng ở mức tối đa công nghệ phân tích dữ liệu lớn nhằm phát triển công ty của mình. Tất cả các ứng dụng dữ liệu lớn hiện đã trở thành một phần không tách rời trong cuộc sống hằng ngày của chúng ta;

tuy nhiên, dữ liệu lớn đang được sử dụng theo nhiều cách thức phi truyền thống khác nhau ở nhiều lĩnh vực ngành nghề khác nhau. Từ các công ty khởi nghiệp đến các tập đoàn lớn với lịch sử hoạt động lâu đời, các nhà quản lý đang sử dụng dữ liệu lớn cho nhiều mục đích khác nhau. Đổi mới sáng tạo là yếu tố then chốt, nhưng doanh thu và lợi nhuận là một trong những ưu tiên hàng đầu cho các tổ chức lớn. Có lý do để hiểu rõ hơn về xu hướng này – các tổ chức lớn đã tích cực tạo nguồn dữ liệu trong suốt nhiều năm, và họ cũng đang phải đối mặt với hàng loạt vấn đề cần giải quyết – trong trường hợp đó, không có gì khó hiểu khi họ muốn sử dụng nguồn lực đang có để giải quyết các vấn đề hiện tại. Ngược lại, các công ty khởi nghiệp thường làm việc trên nguyên tắc phải làm sao để cung cấp một sản phẩm/dịch vụ *nhanh hơn, tốt hơn, và rẻ hơn* so với những gì đang được cung cấp trên thị trường, vì vậy họ muốn sử dụng dữ liệu lớn để thực hiện đổi mới sáng tạo và nghĩ ra nhiều thứ mới mẻ.

Theo số liệu của Trung tâm Nghiên cứu ABI, tổng các khoản chi phí của các tổ chức trên toàn cầu vào dữ liệu lớn vượt mức 31 tỉ đô-la vào năm 2013, và dự đoán sẽ đạt con số ngất ngưởng là 114 tỉ đô-la vào năm 2018 – trong đó bao gồm cả lương cho nhân viên và các khoản đầu tư vào phần cứng và phần mềm.

Điều quan trọng cần lưu ý là những con số này chưa tính đến các công ty đổi mới sáng tạo sẽ ra đời trong tương lai và các sản phẩm sẽ hình thành trong những năm sắp tới; do đó, dường như luồng đầu tư vào lĩnh vực này sẽ còn tiếp tục gia tăng.

"Chúng ta đang gặt hái được nhiều giá trị kinh doanh thực sự từ các kỹ thuật phân tích dữ liệu lớn. Dù bạn yêu thích dữ liệu lớn hay coi nó là thứ phiền toái, thì vẫn có một thực tế không thể phủ nhận là nó quá lớn đến mức không ai có thể phớt lờ. Nhưng tất cả số dữ liệu khổng lồ đó có giá trị như thế nào và có ý nghĩa gì đối với sự thành công hay thất bại của một tổ chức? Trên cương vị một công ty đã triển khai các dự án phân tích dữ liệu lớn trong nội bộ, đồng thời cũng là công ty tư vấn cho khách hàng về cách sử dụng dữ liệu lớn, chúng tôi thực lòng tin rằng đó là một mỏ vàng rất tiềm năng," David Williams, CEO, Công ty Dịch vụ Tư vấn Tài chính Deloitte, chia sẻ.

Trong số các nhà quản lý tham gia vào cuộc khảo sát của chúng tôi, một phần lớn tin rằng dữ liệu lớn đóng vai trò quan trọng cho tổ chức của họ. 68% nhà quản lý cho biết dữ liệu lớn đóng vai trò quan trọng hoặc rất quan trọng trong công ty của họ, và 22% nhận thấy rằng nó không quan trọng tại thời điểm hiện nay. Hiển nhiên, nhận thức và kiến thức của giới quản lý về các công nghệ dữ liệu lớn cũng như về cách

vận dụng nó để tạo ra giá trị đã gia tăng nhanh chóng trong vài năm trở lại đây – cho thấy dữ liệu lớn không phải là một trào lưu nhất thời như nhận định của một số người trước đây.

GIÁ TRỊ CHÍNH TỪ DỮ LIỆU LỚN

Các nhà quản lý cho rằng dữ liệu lớn có thể bổ sung những giá trị gì? Không ngạc nhiên khi thấy nhiều người có chung một quan điểm rằng có thể sử dụng dữ liệu lớn để cải thiện hiệu suất vận hành, chẳng hạn trong nội bộ các tổ chức. 70% nhà quản lý tin rằng có thể sử dụng kỹ thuật phân tích dữ liệu lớn để nâng cao lợi nhuận, do đó nó có thể trở thành một công cụ đắc lực để các công ty giảm bớt được tỉ lệ sai sót trong sản phẩm, tối ưu hóa quy trình, giảm những lần trì hoãn, và rút ngắn thời gian phản hồi khách hàng. Các công ty như Google và Yahoo! đã bắt tay vào nghiên cứu các công nghệ dữ liệu lớn với mục tiêu ban đầu là cải thiện hiệu suất vận hành, khả năng lưu trữ dữ liệu, và các quy trình truy xuất thông tin.

Tuy nhiên, như đã đề cập ở trước, các cuộc cách mạng dữ liệu lớn hiện đang chịu sự dẫn dắt của hai nhóm tổ chức – một là các công ty lớn lâu đời, sở hữu trong tay cả kho dữ liệu quá khứ và hiện tại với những vấn đề về vận hành cần giải quyết, và một bên là các công ty khởi nghiệp và các công ty khác muốn

sử dụng dữ liệu lớn cho mục đích đổi mới sáng tạo... Phần lớn các công ty mà chúng tôi phỏng vấn trong cuộc khảo sát đều là những doanh nghiệp lớn với lịch sử hoạt động lâu đời, và chỉ có một số ít là các công ty nhỏ mới thành lập, do đó dữ liệu của chúng tôi phản ánh rằng 70% các nhà quản lý coi dữ liệu lớn là một cơ hội giúp họ cải thiện hiệu suất vận hành.

"Phân tích dữ liệu lớn mang đến rất nhiều lợi thế," tiến sĩ Carsten Bange, CEO của BARC, cho biết. "Xếp hàng đầu trong danh sách các lợi ích mà phân tích dữ liệu lớn mang lại là các quyết định chiến lược đúng đắn hơn (69%), cải thiện việc kiểm soát các quy trình vận hành (54%), hiểu rõ khách hàng hơn (52%), và giảm bớt các khoản chi phí (47%)."

Một số doanh nghiệp đã, đang, và vẫn sẽ luôn luôn được xây dựng dựa trên dữ liệu. Đối với các doanh nghiệp khác, con đường kiến tạo lợi thế cạnh tranh thông qua dữ liệu và phân tích vẫn còn khá mới mẻ. Tuy nhiên, trong bất kỳ tổ chức nào, sự hỗ trợ và tham gia của người lãnh đạo cao nhất cũng đóng vai trò then chốt trong việc vạch ra đường lối. "Mối quan hệ giữa vị CEO với đội ngũ phân tích dữ liệu trong công ty càng gắn bó khăng khít bao nhiêu, công ty đó càng đầu tư nhiều bấy nhiêu vào các công nghệ lấy dữ liệu làm trung tâm," Matt Ariker, giám đốc điều hành của Trung tâm Phân tích Marketing Tiêu dùng tại McKinsey and Company, chia sẻ.

Khoảng 57% nhà quản lý cho rằng dữ liệu lớn và công nghệ phân tích có thể giúp doanh nghiệp kiến tạo vị thế dẫn đầu thị trường về khả năng tiết giảm chi phí – rõ ràng, vận hành tối ưu và hoạt động hiệu quả có thể giúp doanh nghiệp vươn lên dẫn đầu thị trường về khả năng tiết giảm chi phí và xây dựng lợi thế cạnh tranh dựa vào đó. Nếu một công ty có chi phí hoạt động thấp hơn nhưng không thỏa hiệp về mặt giá thành sản phẩm, thì rốt cuộc, họ cũng sẽ đạt được biên lợi nhuận gộp cao hơn. Dữ liệu lớn trang bị khả năng xác định, thu thập dữ liệu, và đo lường mọi khía cạnh của một doanh nghiệp, nhờ đó các nhà quản lý có thể tối ưu hóa quy trình và cắt giảm chi phí. Không hẳn là các tổ chức lớn với lịch sử hoạt động lâu dài chỉ quan tâm tới việc cải thiện hiệu suất vận hành. Một trong những ngành có lịch sử xa xưa nhất như ngành chăm sóc sức khỏe hiện cũng đang tiến hành đổi mới quy trình để tối ưu hóa việc hoạch định nguồn lực và cải thiện lợi nhuận cho công ty.

Thông qua việc tạo điều kiện để các chuyên gia trong lĩnh vực chăm sóc sức khỏe có thể thu thập và phân tích những luồng dữ liệu số khổng lồ liên quan đến chăm sóc sức khỏe để tìm ra những phát hiện mới, các nền tảng đám mây lưu trữ dữ liệu lớn đang mở ra một kỷ nguyên mới trong lĩnh vực chăm sóc bệnh nhân chất lượng cao với chi phí thấp hơn.

Theo một báo cáo của McKinsey and Company, trong tương lai dữ liệu lớn có thể giúp người dân Mỹ tiết kiện được 450 tỉ đô-la tiền dịch vụ chăm sóc sức khỏe hằng năm. Thực ra, một phần của tương lai ấy đã đến rồi – một trong những cách tốt nhất để kiểm soát chi phí chăm sóc sức khỏe là làm sao để bệnh nhân không phải đến bệnh viện ngay từ đầu. Với các công cụ dữ liệu mới có thể gửi cảnh báo tự động khi bệnh nhân đến lịch chủng ngừa, các bác sĩ có thể giảm tải lượng bệnh nhân nhập viện bằng cách thực hiện các biện pháp chăm sóc phòng ngừa. Bệnh viện Parkland ở Dallas, Texas đã sử dụng công nghệ phân tích và lập mô hình dự đoán để xác định các bệnh nhân có nguy cơ cao trong bộ phận chăm sóc bệnh nhân tim mạch và dự đoán các kết quả có thể xảy ra sau khi bệnh nhân được chuyển về nhà. Nhờ vậy, Parkland đã giảm được 31% tỉ lệ bệnh nhân tim mạch trở lại bệnh viện sau 30 ngày xuất viện. Đối với Parkland, con số đó tương đương với khoản tiết kiệm chi phí 500.000 đô-la mỗi năm, chưa kể đến các khoản chi phí khác mà bệnh nhân tiết kiệm được khi không phải tái nhập viện. Các bệnh viện không chỉ tiết kiệm chi phí bằng cách chủ động hỗ trợ bệnh nhân sống khỏe mạnh để tránh phải tới bệnh viện, họ còn sử dụng các nguồn lực chưa dùng đến để phục vụ cho các khách hàng mới, từ đó cải thiện được doanh thu.

Theo một nghiên cứu mới được đăng tải trên tạp chí *Health Affairs* số ra tháng 7 năm 2014, phân tích dữ liệu lớn hiện đang đặt nền móng cho hành trình cắt giảm các loại chi phí. "Những ví dụ mà chúng tôi nêu ra trong nghiên cứu này mang lại nhiều kiến giải chính cho các dự án đơn giản về dữ liệu lớn trong lĩnh vực chăm sóc sức khỏe, đưa ra nhiều bài học cho hoạt động quản lý và giám sát, đưa ra những gợi ý để giải quyết những mối lo ngại về quyền riêng tư, và nhấn mạnh nhu cầu đối với sự hỗ trợ của các hoạt động nghiên cứu về công nghệ phân tích," theo lời David Bates, giám đốc điều hành kiêm giám đốc chất lượng tại Bệnh viện Phụ nữ Brigham, đồng thời cũng là tác giả chính của nghiên cứu trên.

Dữ liệu mà nhiều ngân hàng tiếp cận chủ yếu là các dữ liệu có cấu trúc và bán cấu trúc, bao gồm cả dữ liệu về số lần nhấp chuột trên website, các hồ sơ giao dịch, các tờ khai của ngân hàng, và các file ghi âm từ tổng đài. Theo nghiên cứu của Thomas Davenport và Jill Dyché thuộc Viện Phân tích Quốc tế, các ngân hàng đang ngày càng làm tốt hơn những công việc tìm hiểu hành trình khách hàng, giám sát chất lượng dịch vụ, và xác định nguyên nhân khiến khách hàng rời bỏ. Họ thực hiện giám sát hành trình khách hàng thông qua một mạng lưới các website, tổng đài, nhân viên giao dịch, và đội ngũ nhân sự khác ở chi nhánh

để tìm hiểu về con đường khách hàng đã đi qua để đến với ngân hàng, đồng thời xác định xem liệu những con đường đó có ảnh hưởng đến tỉ lệ khách hàng rời bỏ hay đến quyết định mua một sản phẩm tài chính cụ thể hay không.

Cũng theo báo cáo trên, ngân hàng Bank of America đã điều chỉnh cấu trúc của mình để tăng hiệu quả cho dữ liệu lớn. "Ngân hàng này trước đây từng tuyển dụng một số chuyên gia phân tích định lượng, nhưng để chuẩn bị cho kỷ nguyên dữ liệu lớn, họ đã hợp nhất và tái cấu trúc bộ phận này với sơ đồ tổ chức theo kiểu ma trận, theo đó đội ngũ này vừa trực thuộc nhóm phân tích trung tâm, vừa báo cáo cho các đơn vị nghiệp vụ khác. Chẳng hạn, bộ phận phân tích ngân hàng bán lẻ bao gồm các chuyên gia phân tích định lượng và khoa học gia dữ liệu; bộ phận này nằm dưới sự quản lý của Aditya Bhasin, đây cũng là người đứng đầu mảng Marketing Tiêu dùng và Ngân hàng số. Bộ phận này hiện đang làm việc chặt chẽ hơn bao giờ hết với người đứng đầu các đơn vị nghiệp vụ khác nhau."

Rất nhiều công ty cũng đang âm thầm làm việc để cải thiện trải nghiệm khách hàng. Tập đoàn sản xuất ô tô BMW đã bắt tay vào sử dụng dữ liệu lớn thu thập được từ các đại lý bán xe và các đơn vị sản xuất ở khắp nơi trên thế giới. Chẳng hạn, các mẫu xe mô hình của BMW có thể tạo ra trung bình 15.000

điểm dữ liệu từ động cơ và bộ truyền động đến phanh và giảm xóc. Với phân tích dữ liệu lớn, BMW có thể phát hiện, thậm chí là khắc phục những điểm sơ hở xuất hiện trong quá trình sản xuất. Những dữ liệu vốn trước đây thường mất hàng tháng trời để phân tích thì nay có thể diễn ra chỉ trong vài ngày. Với dữ liệu thu thập được trong các cuộc chạy thử xe và các quy trình khác, người ta có thể phát hiện ra lỗi sai để xử lý trước khi dòng xe mới được đưa vào sản xuất. Bằng nỗ lực này, BMW đã cải thiện được sự hài lòng của khách hàng bằng cách cho ra đời những chiếc xe có chất lượng cao hơn. Thời gian ở trong cửa hàng của những chiếc xe này sẽ được giảm bớt, và việc sửa chữa những hư hỏng phát sinh cũng sẽ đòi hỏi ít thời gian hơn. Điều chắc chắn là khách hàng càng phải bỏ ít thời gian và tiền bạc cho xe, thì họ càng hài lòng.

Gần đây, hãng hàng không Southwest Airlines cho biết họ đã quyết định triển khai công nghệ đám mây để phục vụ mục đích lưu trữ và xử lý dữ liệu với bộ giải pháp mới về tối ưu hóa lực lượng lao động và liên hệ khách hàng do Aspect cung cấp. Động thái này cho phép Southwest Airlines cung cấp dịch vụ khách hàng với chất lượng còn tốt hơn nữa nhờ tích cực sử dụng các kỹ thuật phân tích dữ liệu. Đội ngũ nhân sự hoạt động ở tuyến đầu trực tiếp tiếp xúc với khách hàng sẽ nhận được các bảng báo cáo tóm tắt theo thời

gian thực về các chỉ số đo lường hiệu suất công việc (KPI) liên quan đến các mục tiêu về chiến lược và vận hành. Họ sẽ sử dụng kỹ thuật phân tích giọng nói để rút ra những thông tin có ý nghĩa và sâu sắc trong các cuộc trao đổi ghi âm trực tiếp giữa nhân viên phục vụ và khách hàng. Điều này sẽ mang lại nhiều thông tin hơn giúp Southwest Airlines xác định được khách hàng đang cần tìm kiếm gì, và trải nghiệm của họ ở Southwest Airlines như thế nào. Các chỉ số đo lường khác nhau sẽ hướng dẫn đội ngũ nhân viên cách cung cấp dịch vụ chất lượng cao cho khách hàng.

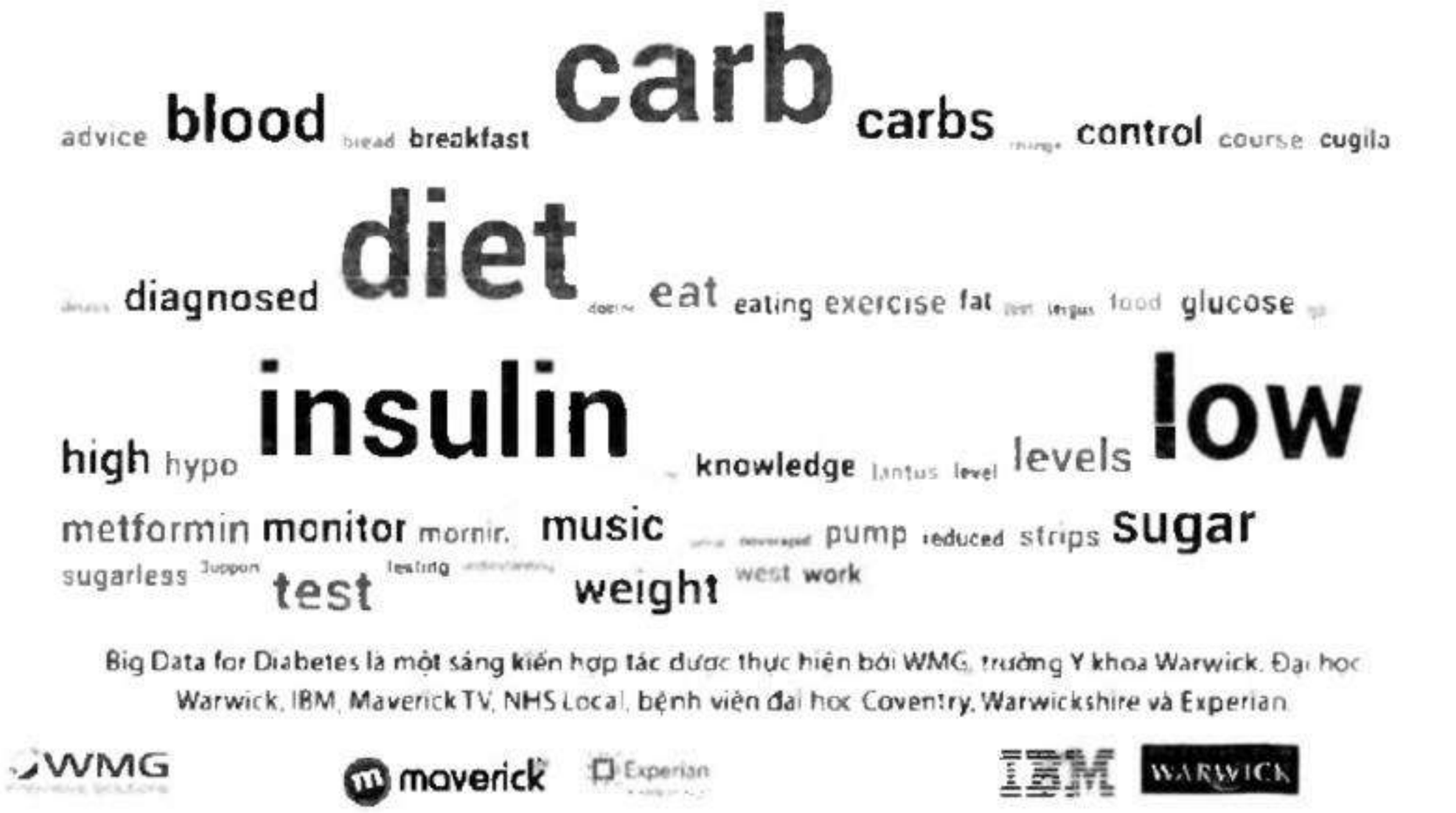
Ngoài ra, Southwest Airlines còn có thể cải thiện hơn nữa chất lượng dịch vụ khách hàng bằng cách phân tích các dữ liệu thời gian thực như phân tích mạng xã hội để qua đó cung cấp cho khách hàng những giải pháp phù hợp riêng với họ. Nhờ có dữ liệu lớn, Southwest Airlines biết họ nên quảng bá sản phẩm nào cho khách hàng nào vào thời điểm nào và thông qua kênh liên lạc nào. Trải nghiệm mang tính cá nhân hóa cao độ này giúp họ gia tăng tỉ lệ chuyển đổi thành khách hàng và củng cố sự hài lòng ở các thượng đế.

Phân tích dữ liệu lớn cũng hỗ trợ bạn sáng tạo và phát triển sản phẩm/dịch vụ mới – đây là ý kiến của 76% các nhà quản lý tham gia vào cuộc khảo sát của chúng tôi. Các công ty như Odimax hiện đang xây

dựng nhiều cơ sở dữ liệu khách hàng tiềm năng dựa trên phân tích ngữ nghĩa đối với các cuộc trao đổi trên mạng xã hội, nhờ đó họ có thể cung cấp các thông tin liên hệ có chất lượng cho khách hàng của mình. Một ví dụ đầy tính sáng tạo nữa xuất phát từ dự án phân tích ngữ nghĩa của IBM đối với các cuộc trao đổi trên mạng xã hội liên quan đến đề tài "Bệnh tiểu đường" (xem Hình 7.1). Từ các kết quả thu về, có thể thấy rõ rằng trong một số vùng trên thế giới, người ta đang sử dụng âm nhạc làm liệu pháp phòng bệnh tiểu đường – và hiện nay IBM đang tiến hành những nghiên cứu sâu thêm về cách trị bệnh sáng tạo này.

Trong trận chung kết Cúp bóng đá Thế giới năm 2014, người Đức đã sử dụng các kỹ thuật phân tích dữ liệu lớn để dự đoán trận đấu của Argentina, cách họ sẽ chơi, và kế hoạch chơi của họ. Sau đó, họ lập công thức cho chiến lược chơi của mình với sự hỗ trợ của công nghệ phân tích dữ liệu lớn. Trong môn cricket, hãng truyền hình ESPN sử dụng dữ liệu lớn để dự đoán cách đánh bóng của một cầu thủ. Họ cũng nắm rõ về điểm mạnh và điểm yếu của cầu thủ bằng cách phân tích hiệu quả chơi trước đây của cầu thủ.

HÌNH 7.1 Dự án phân tích ngữ nghĩa của IBM về các cuộc trao đổi trên mạng xã hội liên quan đến chủ đề bệnh tiểu đường



Động vật trong tự nhiên vốn rất khó theo dõi và cách duy nhất để thực hiện việc này là tận dụng một hệ thống theo dõi khá thô sơ. Hãng HP đã hợp tác với Tổ chức Bảo tồn Quốc tế (Conservation International – CI) để tạo ra Earth Insights thông qua phân tích dữ liệu lớn.

Hệ thống phân tích dữ liệu lớn này được thiết kế để tạo ra một hệ thống cảnh báo sớm cho các động vật thuộc nhóm gặp nguy hiểm. Bằng cách sử dụng những thiết bị cảm biến, các camera trong hệ thống phân tích này thu thập dữ liệu từ khoảng 1.000 thiết bị, và sau đó tổng hợp dữ liệu lại để tìm ra số lượng

cá thể của từng loài. Sau đó, dữ liệu lớn đã thu thập sẽ được đưa vào nền tảng HP Vertica để phân tích các thông số một cách chính xác nhằm xác định các khu vực hoặc địa điểm những loài động vật gặp nguy hiểm đang sinh sống. Điều này giúp bộ phận bảo tồn có thể đầu tư thời gian và tiền bạc vào các khu vực cụ thể để cứu những loài gặp nguy hiểm.

Các nhà quản lý còn gợi ý một số cách áp dụng dữ liệu lớn trong doanh nghiệp của họ như sau:

- Cải thiện quy trình bán hàng và xác định những thay đổi trong các thông điệp truyền thông
- Thu thập và sử dụng các điểm dữ liệu liên kết với nhau
- Tối ưu hóa hiệu quả về độ phủ của các chiến dịch marketing
- Triển khai chiến lược gia nhập thị trường
- Tạo thêm nhiều dòng thu nhập thông qua các dữ liệu thu thập được và kết quả phân tích dựa trên đó
- Tác động tới các chương trình huấn luyện nhân viên
- Nâng cao kết quả và quy trình nghiên cứu và phát triển

- Nghiên cứu hiệu quả về thị trường, tình hình cạnh tranh, và lĩnh vực hoạt động

CÁC THÁCH THỨC KHI TRIỂN KHAI DỰ ÁN DỮ LIỆU LỚN

Từ nhiều năm nay, chúng ta đều đã thấy rằng một trong những thách thức lớn nhất khi triển khai bất kỳ công nghệ hay hệ thống mới nào trong một doanh nghiệp là sự thiếu hụt nguồn lực, nhất là nguồn lực con người để có thể thu thập, tìm hiểu, và học hỏi những kỹ năng và kỹ thuật mới nhất một cách nhanh chóng để bắt đầu triển khai dự án. Trong số các nhà quản lý tham gia vào khảo sát của chúng tôi, 73% đồng ý rằng việc thiếu kiến thức và trình độ chuyên môn là một trong những thách thức chính trong việc triển khai các dự án dữ liệu lớn. Không rõ đó là kiến thức kinh doanh để xây dựng chiến lược hay kiến thức kỹ thuật để triển khai các chiến lược đó, nhưng chúng tôi cho rằng dữ liệu lớn vẫn là một lĩnh vực nghiên cứu tương đối mới mẻ, và để có được những hiểu biết phù hợp và cần thiết, nhà quản lý cũng như các thành viên khác trong đội ngũ của nhà quản lý cần nỗ lực rất nhiều. Các doanh nghiệp có thể tranh thủ sự hỗ trợ từ các ông lớn trong lĩnh vực dữ liệu lớn như IBM, Google, các công ty tư vấn hoặc các công ty liên kết với những học viện giáo dục hoạt động trong lĩnh vực dữ liệu lớn.

57% các nhà quản lý tham gia vào cuộc khảo sát của chúng tôi cũng nêu ra vấn đề ngân sách là một trong những thách thức họ phải đối mặt. Theo quan điểm của chúng tôi, đội ngũ lãnh đạo doanh nghiệp phải coi đầu tư vào dữ liệu lớn là một khoản đầu tư chiến lược dài hạn. Họ có thể dành ra những khoản ngân sách phù hợp cho các dự án dữ liệu lớn, bởi chúng không chỉ giúp họ tăng trưởng doanh thu trước mắt mà còn cải thiện hiệu suất vận hành và tạo lợi thế về chi phí cho doanh nghiệp trên thị trường. Có nhiều phương pháp trực tiếp và gián tiếp để đo lường giá trị được tạo ra (chúng tôi đã bàn đến việc này ở Chương 3) – theo đó nhà quản lý có thể kiến tạo giá trị và thuyết phục lãnh đạo phân bổ ngân sách phù hợp cho hoạt động đầu tư vào dữ liệu lớn.

Một số ít các nhà quản lý tham gia vào khảo sát của chúng tôi, chiếm khoảng 35%, chủ yếu đến từ châu Âu, cho rằng an ninh là một trong những thách thức lớn trong việc triển khai các dự án dữ liệu lớn. Các vấn đề an ninh chủ yếu thiên về vấn đề tổ chức, và nhà quản lý nên xử lý chúng theo từng trường hợp cụ thể.

Các tổ chức rõ ràng đã bắt tay vào xây dựng những chiến lược để quản lý việc phân tích và các dự án liên quan đến dữ liệu lớn. 17% các nhà quản lý nói

rằng họ đã chuẩn bị sẵn sàng một chiến lược chính thức để triển khai các dự án dữ liệu lớn, và 39% cho biết tổ chức của họ hiện đang xúc tiến xây dựng chiến lược. 40% các tổ chức cho rằng việc hoạch định ra một chiến lược chính thức không có ý nghĩa gì, nhưng chúng tôi cho rằng con số này sẽ giảm đi đáng kể trong những năm tới đây. Dữ liệu lớn và kiến tạo giá trị là một cơ hội lớn đối với các doanh nghiệp và phần lớn các tổ chức sẽ sớm hướng theo trào lưu này, có thể trên tinh thần tự nguyện hoặc rơi vào tình thế bắt buộc phải đi theo. Thực ra, 65% các nhà quản lý tham gia vào cuộc khảo sát của chúng tôi nhất trí rằng dữ liệu lớn sẽ làm biến đổi bộ mặt lĩnh vực ngành nghề mà họ đang hoạt động trong vòng 5-10 năm tới, và nó sẽ làm thay đổi cách thức hoạt động của các doanh nghiệp ở hầu khắp các lĩnh vực. Trong khi đó, 11% nhà quản lý phản đối, và phần còn lại không có ý kiến gì cả – tuy không biết cụ thể những người này hoạt động trong lĩnh vực nào, nhưng chúng tôi cho rằng có một số lĩnh vực thay đổi chậm hơn.

Trong các tổ chức sử dụng dữ liệu lớn ngày nay, đa phần đều cảm thấy hài lòng với kết quả mà họ đạt được. Khi chúng tôi hỏi về kết quả của các dự án đó, từ cả những doanh nghiệp tính đến nay mới chỉ thực hiện được ít nhất một dự án dữ liệu lớn, thì 73% cho biết họ đạt được kết quả đúng như kỳ vọng hoặc tốt hơn so với kỳ vọng ban đầu.

Nhìn chung, các nhà quản lý đến từ nhiều lĩnh vực ngành nghề khác nhau đều hào hứng về dữ liệu lớn và cách vận dụng nó để tạo ra thêm nhiều giá trị cho doanh nghiệp. Bây giờ vẫn là buổi bình minh của công nghệ này, nhưng các tổ chức đang tích cực và chủ động đón nhận công nghệ mới này một cách nhanh chóng và gấp rút lấp đầy những lỗ hổng về nhân tài cũng như các nguồn lực khác. Các học viện giáo dục, các website trực tuyến, và đội ngũ huấn luyện viên tư nhân cũng đang đóng góp nhiều công sức để lấp đầy lỗ hổng về nhân lực trên thị trường, và không bao lâu nữa trên thị trường sẽ có đầy đủ nguồn lực. Chúng tôi cũng cảm thấy rằng với tinh thần đổi mới sáng tạo và sự dồi dào trong thị trường nhân lực, chi phí triển khai nói chung và những yêu cầu về ngân sách đối với các dự án dữ liệu lớn sẽ sớm giảm xuống.

TÓM TẮT

Ở chương này, chúng tôi đã trình bày về quan điểm của những người đã từng trải qua những tình huống như các độc giả của cuốn sách này – một số người đã bắt tay vào triển khai các dự án dữ liệu lớn, một số khác đang lên kế hoạch. Kết quả khảo sát thực hiện trên hơn 200 người tham gia mà chúng tôi đích thân thực hiện đã chỉ rõ rằng dữ liệu lớn có ý nghĩa vô cùng quan trọng đối với doanh nghiệp của họ. Trong hầu

hết các trường hợp, giá trị chính mà dự án dữ liệu lớn mang lại là cải thiện hiệu suất vận hành, tạo điều kiện để xây dựng quy trình ra quyết định thông minh, và khích lệ sự đổi mới sáng tạo – thực ra, chúng tôi cũng đã bàn khá chi tiết về các chủ đề này ở Chương 4.

Trong chương này, chúng tôi cũng trình bày nội dung một số cuộc phỏng vấn với những người tham gia khảo sát hiện đang sử dụng dữ liệu lớn trong tổ chức hoặc cho khách hàng của họ. Nhìn chung, quan điểm chủ yếu của hầu hết những người này là dữ liệu lớn đóng một vai trò hết sức quan trọng và nhiều triển vọng trong bất kỳ doanh nghiệp nào.

8

KẾT LUẬN VÀ BÀN LUẬN

Dữ liệu lớn là một trong những khía cạnh quan trọng nhất đối với hầu hết các doanh nghiệp trong thời đại ngày nay. Xu hướng tiếp nhận công nghệ mới không phải điều gì quá mới mẻ đối với chúng ta – bất kể đó là xu hướng nào, từ internet đến ERP (hoạch định nguồn lực doanh nghiệp), tất cả hầu như đều có tỉ lệ tiếp nhận chậm chạp ban đầu, sau đó rốt cuộc hầu như doanh nghiệp nào cũng đón nhận. Trong trường hợp dữ liệu lớn, không thiếu người từng tuyên bố rằng đây chỉ là một trào lưu sớm nở tối tàn trên thị trường, nhưng dần dần, những tuyên bố đó đã lặng lẽ bị đẩy lui. Thực ra, có lý do để một số người cho rằng dữ liệu lớn chỉ là một trào lưu nhất thời: Họ đã từng chứng kiến hành trình của dữ liệu và khoa học phân tích trong quá khứ, và họ không thể nhìn ra tầm vóc mới của yếu tố "lớn" sau chữ "dữ liệu", không biết nó sẽ bổ

sung thêm những giá trị gì. Nếu bạn nhìn nhận mọi việc thông qua lăng kính cũ, thì rốt cuộc bạn cũng chỉ thấy những điều cũ kỹ mà thôi.

Bất chấp thực tế, và bất chấp quan điểm của các tác giả của cuốn sách này, rằng dữ liệu lớn và khoa học phân tích sẽ còn tồn tại, phát triển và mang lại rất nhiều giá trị cho các doanh nghiệp, song chúng tôi sẽ không khẳng định rằng tốc độ tiếp nhận nó sẽ diễn ra nhanh chóng, hay thậm chí chúng tôi cũng không dám kết luận rằng cuộc cách mạng này sẽ tạo ra giá trị bền vững cho tất cả các doanh nghiệp. Dữ liệu lớn là một yếu tố quan trọng đối với các doanh nghiệp và chúng ta không thể bỏ qua chuyện đó – thực ra, dù có nhận thức được điều này hay không, chúng ta đang đóng góp vào cuộc cách mạng này trên cương vị người dùng các thiết bị như điện thoại di động và các ứng dụng như Facebook hay Twitter... Nhưng như chúng ta đã thấy trong cuốn sách này, giá trị tạo ra phải được đo lường theo phương pháp đúng đắn và phải tồn tại trong tất cả các hoạt động của doanh nghiệp, bao gồm cả những khía cạnh như phân tích dữ liệu lớn.

Chúng tôi tin rằng những kiến giải còn ẩn giấu và tiềm năng giá trị nằm trong mảng dữ liệu phi cấu trúc của dữ liệu lớn sẽ lớn hơn nhiều so với những gì dữ liệu có cấu trúc có thể mang lại. Chúng tôi có hơn

40 năm kinh nghiệm sử dụng tình báo doanh nghiệp chính thức ở nhiều ngành nghề (và phi chính thức trong suốt nhiều năm), và chúng tôi đã thử nghiệm cũng như đầu tư rất nhiều vào lĩnh vực phân tích dữ liệu có cấu trúc, và trình bày kết quả phân tích. Tuy nhiên, trong phần lớn các trường hợp, chúng tôi hoặc đã bỏ qua hoặc không thể đưa dữ liệu phi cấu trúc vào phần phân tích chính. Chỉ đến khi điều này trở thành một vấn đề thực sự nghiêm trọng đối với các công ty internet như Yahoo! và Google, chúng tôi mới quay sang tìm cách thu thập và phân tích dữ liệu phi cấu trúc – vấn đề này đã được đưa ra bàn luận ở trước. Như vậy, một số người trong chúng ta mới chỉ bắt đầu nghĩ đến tầm quan trọng của dữ liệu phi cấu trúc từ khoảng hơn một thập niên trở lại đây. Đó là một sự bổ sung tương đối mới mẻ, thú vị, nhưng ít được khám phá trong toàn bộ địa hạt của khoa học phân tích dữ liệu – do đó, nó là một phần quan trọng có tiềm năng mang lại rất nhiều giá trị. Thật may mắn (nhưng cũng có thể không may mắn), dữ liệu phi cấu trúc hầu hết lại được kết hợp với dữ liệu có cấu trúc, vì thế cấp độ phân tích tiếp theo bị thiên lệch về hướng dữ liệu phi cấu trúc.

Trong dữ liệu lớn, cơ hội tồn tại ở cả phía phần cứng/lưu trữ và trích xuất dữ liệu và phần mềm/phân tích, nhưng phía nào quan trọng hơn và cần đầu tư

bao nhiêu lại phụ thuộc vào điều kiện và nhu cầu của từng doanh nghiệp cụ thể. Do những yêu cầu về mặt vận hành hoặc tuân thủ quy định, nhiều tổ chức tài chính có thể sẽ phải đầu tư rất nhiều vào cơ sở hạ tầng và cân nhắc kỹ đến việc bảo đảm an toàn cho dữ liệu hay trung tâm dữ liệu riêng của họ, sau đó mới chuyển sang triển khai các hoạt động phân tích – và trong cả hai trường hợp, họ đều có thể tạo ra giá trị tài chính. Ngược lại, nhiều công ty khởi nghiệp hiện đang sử dụng cơ sở hạ tầng làm dịch vụ và tránh phải bỏ ra khoản đầu tư ban đầu lớn, họ tập trung trực tiếp vào khía cạnh phân tích. Trong cuốn sách này, chúng tôi đã bàn nhiều đến khía cạnh phân tích vốn mang lại nhiều cơ hội ngàn năm có một để các doanh nghiệp kiến tạo giá trị.

Nói về vấn đề kiến tạo giá trị ở bất kỳ doanh nghiệp nào, có sự khác biệt rõ ràng giữa giá trị tài chính và giá trị phi tài chính – nhưng phần lớn mọi người đều cho rằng vai trò trước hết của một doanh nghiệp là tạo ra của cải cho các cổ đông, vì vậy việc kiến tạo giá trị tài chính được ưu tiên hơn so với giá trị phi tài chính. Các doanh nghiệp nên cân nhắc đến *khía cạnh mềm* của việc kiến tạo giá trị và nên sử dụng các biện pháp đo lường đại diện để quan tâm đến lợi ích của tất cả các bên liên quan, nhưng nếu trọng tâm chính của doanh nghiệp là giá trị phi tài chính, khi đó sự bền vững của

doanh nghiệp sẽ là một câu hỏi để ngỏ xét về lâu dài. Vì thế, bất kỳ dự án nào trong một doanh nghiệp cũng đều phải tự giải quyết được phần chi phí tài chính và đóng góp tài chính trước tiên. Từ góc nhìn đó, các nhà quản lý và nhà lãnh đạo trong tổ chức sẽ đưa ra được các biện pháp đo lường hiệu quả và kịp thời.

Nhiều doanh nghiệp được xếp vào nhóm khác và về bản chất, họ chính là các công ty dữ liệu thực thụ, trong đó có Facebook và Twitter – bất luận mức độ nổi tiếng của họ ra sao, cơ sở khách hàng và ý tưởng kinh doanh là gì, thì tài sản quan trọng duy nhất của các công ty này chính là kho dữ liệu mà họ tạo ra thông qua các hoạt động của người dùng, và dữ liệu của họ trực tiếp hoặc gián tiếp tạo ra dòng doanh thu nhất quán cho họ. Các công ty khác như GNIP (được Twitter mua lại vào năm 2014) và Datasift hiện đang sử dụng dữ liệu của người dùng (như Facebook hay Twitter) và tạo ra giá trị bằng cách khiến dữ liệu phi cấu trúc dễ sử dụng hơn cho các công ty khác, giúp họ dựa vào đó để xây dựng hệ thống riêng của mình. Các công ty như thế này đang đảm nhiệm việc đáp ứng những yêu cầu lớn về cơ sở hạ tầng đối với lĩnh vực quản lý dữ liệu. Họ cung cấp các dịch vụ để một công ty nhỏ hay một công ty khởi nghiệp cũng có thể trực tiếp tiếp cận với nguồn dữ liệu phi cấu trúc một cách dễ dàng, đơn giản, và nhất quán mà không cần

đầu tư hàng trăm nghìn đô-la vào công việc chuẩn bị cho các dự án dữ liệu lớn... Các công ty này, trong đó bao gồm Facebook và Twitter, hiện đang tạo ra nhiều giá trị lớn, không chỉ từ việc xây dựng công nghệ phân tích trên dữ liệu lớn, mà còn từ việc bán lại và sử dụng dữ liệu ở nơi khác.

Hiển nhiên, cơ hội khổng lồ mà dữ liệu lớn mang lại, xét về mặt bản chất, nằm ở khía cạnh kỹ thuật, nhưng các nhà quản lý kinh doanh không nên hốt hoảng – bao giờ bên cạnh các bạn cũng có các đội ngũ chuyên gia công nghệ sẵn sàng hỗ trợ bạn về phần thực hành, tức là giải đáp khía cạnh "như thế nào". Quan trọng hơn, nhà quản lý nên đưa ra câu trả lời cho các câu hỏi "tại sao" và "cái gì". Chúng tôi viết cuốn sách này dành cho các nhà quản lý trong doanh nghiệp xuất phát từ lý do rằng các nhà quản lý ngày nay nên được trang bị những kiến thức và kỹ thuật phù hợp để đưa ra được những lý do chiến lược và mang tính thương mại để giải thích vì sao họ cần triển khai bất kỳ dự án nào trong tổ chức của mình, và họ sẽ đo lường mức độ thành công hay thất bại của các dự án đó ra sao. Dữ liệu lớn hiện đang là cơ hội mang lại giá trị tốt nhất cho các doanh nghiệp và các nhà quản lý không nên phụ thuộc hoàn toàn vào bộ phận công nghệ để đưa ra các quyết định mang tính chiến lược. Và nếu bạn cũng nằm trong số họ, thì có

lẽ sau khi đọc xong cuốn sách này, bạn đã hiểu rõ tầm quan trọng của dữ liệu lớn và phân tích dữ liệu lớn đối với doanh nghiệp của mình. Dữ liệu và tác động của nó đối với quá trình kiến tạo giá trị có thể là yếu tố nội bộ hoặc bên ngoài, vì vậy doanh nghiệp nên có cách tiếp cận bao quát và cởi mở để có thể tạo thêm nhiều giá trị cho các bên liên quan. Chúng tôi đã trình bày về các giai đoạn trong phân tích dữ liệu lớn cùng với nhiều trường hợp ví dụ về cách các tổ chức kiến tạo giá trị. Một trong những bài học chủ đạo rút ra từ cuốn sách này, cũng như từ bất kỳ cuốn sách nào khác về quản lý, là: tư duy và hoạch định chiến lược cho một dự án phân tích dữ liệu lớn để tạo giá trị cho doanh nghiệp còn quan trọng hơn bản thân quá trình triển khai dự án. Trong lúc trình bày vấn đề này, chúng tôi cũng đã giới thiệu nhiều trường hợp ví dụ bằng ngôn ngữ đơn giản để các nhà quản lý không có chuyên môn về kỹ thuật cũng có thể cân nhắc đến phân tích dữ liệu lớn mà không cần lo lắng về khía cạnh công nghệ. Sau đó họ cũng có thể tự tin lãnh đạo một dự án tương tự với tầm nhìn chiến lược.

Một trong những đóng góp lớn nhất của cuốn sách này là mô hình C-ADAPT để phân tích dữ liệu lớn. Mô hình này do chính các tác giả của cuốn sách phát triển sau thời gian làm việc với nhiều doanh nghiệp hoạt động ở nhiều lĩnh vực ngành nghề khác nhau với các mục tiêu khác nhau. Mô hình C-ADAPT

được giới thiệu ở đây cùng với bảng tính C-ADAPT rất hữu hiệu mà bất kỳ nhà quản lý nào cùng đội ngũ nhân viên dưới quyền đều có thể sử dụng để xác định và triển khai thành công các dự án phân tích dữ liệu. Bây giờ đã đến lượt bạn tỏa sáng – chúng tôi hy vọng rằng sau khi đọc cuốn sách này, bạn sẽ có được hiểu biết rõ ràng hơn về mô hình C-ADAPT của chúng tôi, và bạn sẽ có thể triển khai nó nhằm tạo ra giá trị bền vững cho tổ chức của mình.

Bạn có thể sử dụng bảng tính C-ADAPT (có thể tải về từ địa chỉ www.creatingvalue.org.uk) và bắt đầu làm việc với nhóm của mình để cùng nhau vạch ra các mục tiêu kinh doanh, khoanh vùng các khía cạnh liên quan, các tập dữ liệu và các chuyên gia phân tích. Bạn có thể làm việc tuần tự từng bước một như đã trình bày ở Chương 7 – và bạn sẽ sớm xác định được điểm khởi đầu cho dự án phân tích dữ liệu lớn tiếp theo của mình – đội ngũ công nghệ bây giờ sẽ có thể làm việc với bạn hiệu quả hơn.

Hết

BIG DATA CHO NHÀ QUẢN LÝ

NHÀ XUẤT CÔNG THƯƠNG

Trụ sở: 655 Phạm Văn Đồng, Bắc Từ Liêm, Hà Nội

Tel: (024) 3934 1562

Fax: (024) 3824 2551

Website: http://nhaxuatbancongthuong.com

Chịu trách nhiệm xuất bản:

Giám Đốc - Tổng Biên Tập: **NGUYỄN MINH HUỆ**

Biên tập viên nhà xuất bản: Đồng Thị Thu Thủy

Sửa bản in: Jean-Marie

Thiết kế bìa: Vương Mai

Trình bày: Đàm Oanh

Biên mục trên xuất bản phẩm của Thư viện Quốc gia Việt Nam

Malviya, Atal

Big data cho nhà quản lý / Atal Malviya, Mike Malmgren ; Khánh Trang dịch. - H. : Công Thương ; Công ty Sách Thái Hà, 2020. - 388tr. ; 21cm

Tên sách tiếng Anh: Big data for managers: Creating value

ISBN: 978-604-9963-04-9

1. Dữ liệu 2. Giá trị 3. Quản lí

005.7 - dc23

COM0064p-CIP

In 1.500 cuốn, khổ 13 x 20,5 cm tại Công ty TNHH In Thanh Bình, địa chỉ: Số 432 đường K2 phường Cầu Diễn, quận Nam Từ Liêm, Hà Nội. Số đăng ký kế hoạch xuất bản: 1526-2020/CXBIPH/04-63/CT. Quyết định xuất bản số: 102/QĐ-NXBCT, cấp ngày 01/06/2020. In xong và nộp lưu chiểu năm 2020.

Trong thế giới số đang phát triển với tốc độ chóng mặt ngày nay, các trang web, điện thoại di động, mạng xã hội, và nhiều nền tảng số khác không ngừng sản xuất ra những luồng dữ liệu khổng lồ hàm chứa trong đó trí tuệ và những thông tin giá trị. Khi được sử dụng đúng đắn, nó sẽ tạo ra giá trị bền vững ở nhiều hình thức khác nhau cho các doanh nghiệp. ***Big data cho nhà quản lý*** cung cấp một khung đơn giản và mang đến một cách tiếp cận dễ hiểu nhằm giúp độc giả có thể hiểu, hấp thu, và phân tích dữ liệu, đồng thời thực hiện những phân tích có ý nghĩa để từ đó đưa ra những quyết định xác đáng có cơ sở thông tin vững vàng. Nó sẽ hỗ trợ cho quá trình tạo giá trị trong doanh nghiệp, từ chăm sóc khách hàng cho đến cải tiến sản phẩm, từ bán hàng và marketing cho đến hiệu suất hoạt động.

THAI HA BOOKS
Serve to Lead

SÁCH THÁI HÀ

THAIHABOOKS
100%
CÓ BẢN QUYỀN

ISBN: 978-604-9963-04-9
9 786049 963049
8 935280 906112

Giá: 149.000đ