BI và DS đều có nhiều mục tiêu giống nhau ( get value từ dữ liệu, xử lý dữ liệu bẩn, chuyển đổi và xử lý dữ liệu, giúp hỗ trợ việc ra quyết định được cải thiện)

Chapter thảo luận về sự khác biệt giữa BI và DS

- question khác nhau

- đặc điểm phân tích khác nhau

- quá trình phân tích khác nhau

- mô hình dữ liệu

- view kinh doanh

DS là gì?

Khoa học dữ liệu là một ngành học mới phức tạp đòi hỏi các kỹ năng và năng lực nâng cao trong các lĩnh vực như thống kê, khoa học máy tính, khai thác dữ liệu, toán học và lập trình máy tính.

DS là tìm biến số (variables) và metrics mới để kết quả dự đoán tốt hơn

khác biệt giữa BI và DS:

1. Đưa ra câu hỏi khác nhau

Để giải thích sự khác nhau thì ta có thể bắt đầu bằng cách giải thích 2 ngành có mục tiêu khác nhau và tìm cách trả lời các câu hỏi khác nhau.

BI:

- tập trung vào descriptive analytics: "what happened?"

■ Tôi đã bán bao nhiêu vật dụng trong tháng trước?

■ Doanh số bán hàng bằng mã zip cho Giáng sinh năm ngoái là gì?

■ Có bao nhiêu đơn vị Sản phẩm X đã được trả lại vào tháng trước?

■ Doanh thu và lợi nhuận của công ty trong quý vừa qua là gì?

■ Tôi đã thuê bao nhiêu nhân viên vào năm ngoái?

BI tập trung vào việc báo cáo về tình trạng hiện tại của doanh nghiệp được gọi chung là business Performance Management (BPM).

BI cung cấp báo cáo quá khứ và trả lời các câu hỏi về hoạt động kinh doanh quá khứ. Các báo cáo và câu hỏi này rất quan trọng đối với doanh nghiệp

Bi có thể áp dụng một số phân tích đơn giản (phân tích chuỗi thời gian, so sánh giai đoạn trước,..)để doanh nghiệp thấy được những khu vực có hiệu suất thấp và cao.Những phân tích này cũng tập trung vào việc giám sát những gì xảy ra đối với doanh nghiệp.

DS:

- Mặt khác, DS tìm kiếm các biến và các số liệu là những yếu tó giúp cho dự đoán tốt hơn, nâng cao hiệu quả kinh doanh. Do đó các DS sẽ tập trung vào predictive analytics (“What is likely to happen?”) and prescriptive analytics (“What should I do?”):

Predictive Questions (What is likely to happen?) ---Câu hỏi dự đoán (điều gì có khả năng xảy ra?)

■ Tôi sẽ bán bao nhiêu vật dụng vào tháng tới?

■ Bán hàng theo mã zip sẽ là gì trong mùa Giáng sinh này?

■ Có bao nhiêu đơn vị sản phẩm X sẽ được trả lại vào tháng tới?

■ Doanh thu và profts của công ty dự kiến ​​cho quý tới là gì?

■ Tôi sẽ cần bao nhiêu nhân viên để thuê vào năm tới?

Prescriptive Questions (What should I do?) ----(tôi nên làm gì?)

■ Đặt hàng [5.000] Thành phần Z để hỗ trợ doanh số tiện ích cho tháng tới.

■ Thuê [y] đại diện bán hàng mới bằng các mã zip này để xử lý doanh số Giáng sinh dự kiến.

■ Đặt sang một bên [$ 125K] trong dự trữ fncial để trang trải sản phẩm X trở lại.

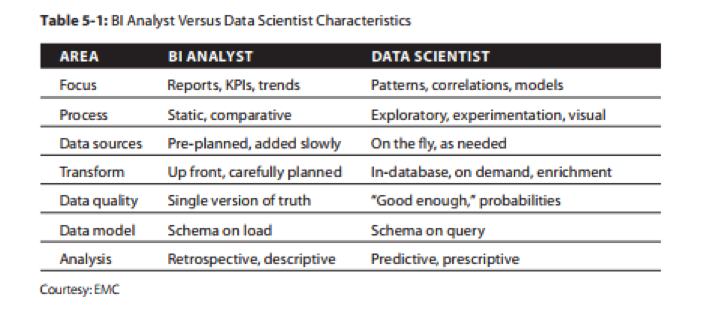
■ Bán hỗn hợp sản phẩm sau đây để đạt được doanh thu hàng quý và các mục tiêu Mar-gin.

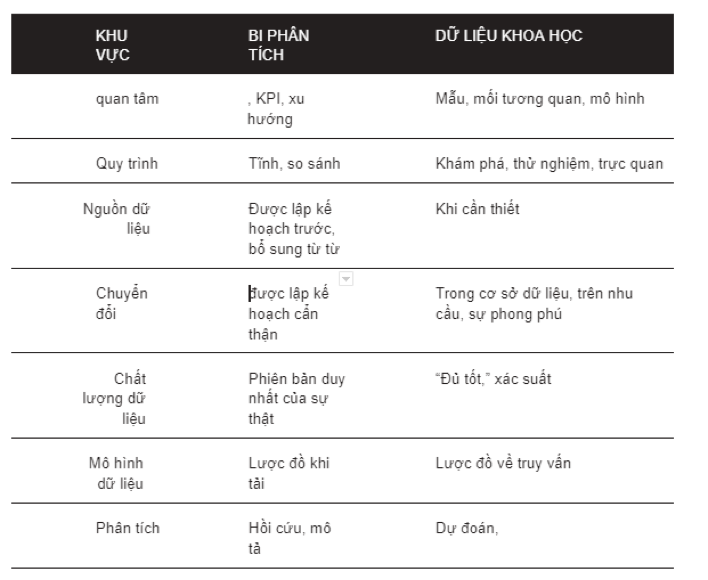
■ Tăng đường ống tuyển dụng lên 35 phần trăm để đạt được các mục tiêu tuyển dụng.

==> sẽ build model (chương 7 sẽ đề cập đến thuật toán và công cụ phân tích)

2. đăc điểm phân tích

Một lĩnh vực khác biệt giữa BI và khoa học dữ liệu là ở đặc điểm cơ bản và cách tiếp cận công việc của những người đảm nhận các vai trò đó

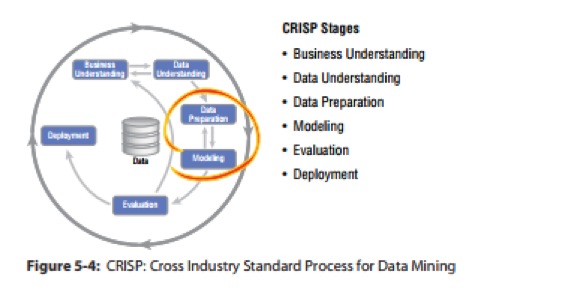




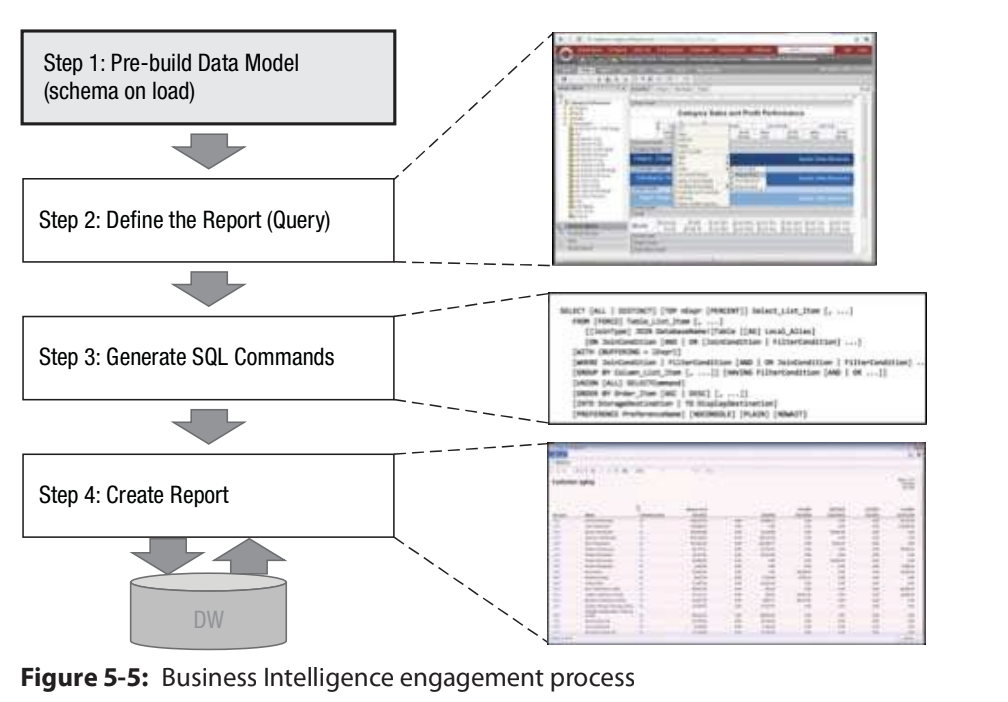
các quan điểm khác nhau về “chất lượng dữ liệu”. Đối với nhà phân tích BI đang xử lý dữ liệu lịch sử, dữ liệu cần phải chính xác 100 phần trăm. BI và các tổ chức kho dữ liệu đã đầu tư rất nhiều vào quản trị dữ liệu và quản lý dữ liệu tổng thể để đảm bảo rằng dữ liệu trong kho dữ liệu chính xác 100%.

Mặt khác, DS đang cố gắng dự đoán những gì có khả năng xảy ra trong tương lai và kết quả là xử lý các xác suất, mức mật độ, F distribution, t test và p value. Tương lai không bao giờ chính xác 100 phần trăm, vì vậy các DS phát triển cảm giác về điều gì là “đủ tốt” trong việc cố gắng dự đoán những gì có thể xảy ra và khuyến nghị những hành động cần thực hiện.

Cần có một thái độ khác để trở thành một DS, một thái độ chấp nhận thất bại như một công cụ để học hỏi. Các DS học cách đón nhận thất bại như một phần của phương pháp tiếp cận nhanh, nhanh thất bại của họ trong việc tìm kiếm nhằm phát hiện ra các chỉ số và biến số mới là những yếu tố dự đoán hiệu suất tốt hơn. Một cách tiếp cận phổ biến mà các DS áp dụng được mô phỏng theo Cross Industry Standard Process for Data Mining (CRISP) model



3. cách tiếp cận phân tích



Bước 1: Chuẩn bị Mô hình Dữ liệu.

Quá trình bắt đầu bằng việc xây dựng mô hình dữ liệu cơ bản. Cho dù bạn sử dụng data warehouse or data mart or hub-and-spoke approach, whether you use a star, snowflake, normalized or dimensional schema, BI phải trải qua một quy trình thu thập yêu cầu chính thức với người dùng doanh nghiệp để xác định tất cả (hoặc ít nhất là phần lớn) các câu hỏi mà người dùng doanh nghiệp muốn trả lời. Trong quá trình thu thập yêu cầu này, BI phải xác định các câu hỏi cấp một và cấp hai mà người dùng doanh nghiệp muốn giải quyết để xây dựng một mô hình dữ liệu mạnh mẽ và có thể mở rộng. Ví dụ:

**Câu hỏi cấp một**: Tháng trước chúng ta đã điều trị cho bao nhiêu bệnh nhân?

■ Câu hỏi cấp hai: So với tháng trước như thế nào?

■ Câu hỏi cấp hai: Các loại DRG chính được xử lý là gì?

**Câu hỏi cấp một**: Có bao nhiêu bệnh nhân đến ER đêm qua?

■ Câu hỏi cấp hai: So với đêm trước đó như thế nào?

■ Câu hỏi cấp hai: Những lý do nhập học hàng đầu là gì?

**Câu hỏi cấp 1**: Tỷ lệ giường bệnh được sử dụng tại Bệnh viện X tuần trước là bao nhiêu?

■ Câu hỏi cấp hai: Xu hướng sử dụng giường trong năm qua là gì?

■ Câu hỏi cấp hai: Những bộ phận nào có mức tăng sử dụng giường lớn nhất?

Sau đó, nhà phân tích BI làm việc data warehouse team để xác định và xây dựng các mô hình dữ liệu cơ bản hỗ trợ các loại câu hỏi này.

Bước 2: Xác nhận Báo cáo (Truy vấn).

Khi các yêu cầu phân tích đã được chuyển thành mô hình dữ liệu, thì bước 2 của quy trình là BI sử dụng công cụ BI — SAP, MicroStrategy, Cognos, Qlikview, Pentaho, v.v. — để tạo truy vấn dựa trên SQL để xây dựng báo cáo và / hoặc trả lời các câu hỏi của doanh nghiệp. BI sẽ sử dụng giao diện đồ họa người dùng (GUI) của công cụ BI để tạo truy vấn SQL bằng cách chọn các thước đo và kích thước; chọn bộ mô tả trang, cột và trang; xác định ràng buộc, tổng phụ và tổng; tạo ra các phép tính đặc biệt (trung bình, trung bình động, rank,...); và chọn tiêu chí sắp xếp. Công cụ BI GUI che giấu phần lớn sự phức tạp của việc tạo SQL.

Bước 3: Tạo lệnh SQL.

Khi nhà phân tích BI hoặc người dùng doanh nghiệp đã xác định yêu cầu báo cáo hoặc truy vấn mong muốn, công cụ BI sẽ tự động tạo các lệnh SQL cần thiết (câu lệnh SQL). Trong một số trường hợp, nhà phân tích BI có thể sửa đổi các lệnh SQL được tạo bởi công cụ BI để bao gồm các lệnh SQL duy nhất có thể không được công cụ BI hỗ trợ.

Bước 4: Tạo Báo cáo.

Trong bước 4, công cụ BI đưa ra các lệnh SQL đối với kho dữ liệu và tạo báo cáo hoặc tiện ích bảng điều khiển tương ứng. Đây là một quá trình lặp đi lặp lại nhiều, trong đó nhà phân tích BI sẽ chỉnh sửa SQL (sử dụng GUI hoặc viết tay câu lệnh SQL) để điều chỉnh yêu cầu SQL. Các nhà phân tích BI cũng có thể chỉ định các tùy chọn kết xuất đồ họa (biểu đồ thanh, biểu đồ đường, biểu đồ hình tròn) cho đến khi họ nhận được báo cáo và / hoặc đồ họa chính xác mà họ muốn

Tóm lại, cách tiếp cận BI dựa trên mô hình dữ liệu được xây dựng trước (giản đồ khi tải), cho phép người dùng truy vấn dữ liệu nhanh chóng và dễ dàng — miễn là dữ liệu mà họ muốn truy vấn đã được xác định và tải vào kho dữ liệu . Nếu dữ liệu không có trong kho dữ liệu, thì việc thêm dữ liệu vào kho hiện có có thể mất hàng tháng để thực hiện. Việc sửa đổi kho dữ liệu để bao gồm một nguồn dữ liệu mới không chỉ đòi hỏi một lượng thời gian đáng kể, mà quá trình này có thể rất tốn kém, vì các lược đồ dữ liệu phải được cập nhật để bao gồm nguồn dữ liệu mới, các quy trình ETL mới phải được xây dựng để chuyển đổi và chuẩn hóa dữ liệu thành các lược đồ dữ liệu được cập nhật, đồng thời các báo cáo và trang tổng quan hiện có có thể phải được cập nhật để đưa vào dữ liệu mới.

#### **Quy trình tham gia của DS**

**Bước 1: Xác định giả thuyết để kiểm tra.**

Bước 1 bắt đầu với việc các DS xác định dự đoán mà họ muốn đưa ra hoặc giả thuyết mà họ muốn kiểm tra. Đây là kết quả của việc hợp tác với chuyên gia về vấn đề kinh doanh để hiểu các nguồn chính của sự khác biệt trong kinh doanh (ví dụ: làm thế nào cty mang lại giá trị) và sau đó xây dựng các giả thuyết hoặc dự đoán liên quan.

**Bước 2: Thu thập dữ liệu… và dữ liệu khác.**

DS thu thập dữ liệu có liên quan hoặc có khả năng thú vị từ nhiều nguồn — cả nội bộ và bên ngoài cơ quan hóa — và đẩy dữ liệu đó vào hồ dữ liệu hoặc hộp cát phân tích. data lake là một khả năng nền tảng tuyệt vời cho quá trình này, vì các DS có thể thu thập và nhập bất kỳ dữ liệu nào họ muốn (nguyên trạng), kiểm tra dữ liệu về giá trị của nó khi đưa ra giả thuyết hoặc dự đoán và sau đó quyết định có đưa dữ liệu đó vào hay không trong mô hình phân tích. Đây là nơi mà một bài tập hình dung có thể tăng thêm giá trị đáng kể trong việc tạo điều kiện cho sự hợp tác giữa người dùng doanh nghiệp và các DS để xác định các nguồn dữ liệu *có thể* giúp cải thiện kết quả dự đoán.

**Bước 3: Xây dựng Mô hình Dữ liệu.**

Bước 3 là nơi các DS xác định và xây dựng lược đồ cần thiết để giải quyết giả thuyết đang được kiểm tra. Các DS không thể xác định lược đồ cho đến khi họ biết giả thuyết rằng họ đang thử nghiệm và hiểu họ sẽ sử dụng nguồn dữ liệu nào để xây dựng mô hình phân tích của mình.

GHI CHÚ Lược *đồ về* quá trình truy vấn này đáng chú ý khác với lược đồ kho dữ liệu truyền thống *về quá trình tải* . DS không dành nhiều tháng để tích hợp tất cả các nguồn dữ liệu khác nhau lại với nhau thành một mô hình dữ liệu chính thức đầu tiên. Thay vào đó, DS sẽ xác định lược đồ khi cần thiết dựa trên dữ liệu đang được sử dụng trong phân tích và các yêu cầu của công cụ phân tích và / hoặc thuật toán. DS có khả năng sẽ lặp lại qua một số phiên bản khác nhau của lược đồ cho đến khi kết thúc một lược đồ hỗ trợ mô hình phân tích với độ tốt phù hợp là chấp nhận hoặc bác bỏ giả thuyết đang được kiểm tra.

**Bước 4: Trực quan hóa dữ liệu.**

sử dụng nhiều công cụ trực quan hóa dữ liệu nổi bật hiện có để khám phá các mối quan hệ, tương quan và ngoại lệ trong dữ liệu. Các DS sẽ sử dụng các công cụ trực quan hóa dữ liệu để bắt đầu quá trình phân tích của họ bằng cách cố gắng xác định các mối tương quan trong dữ liệu đáng được điều tra và các ngoại lệ trong dữ liệu có thể cần được xử lý đặc biệt (ví dụ: biến đổi nhật ký). Các công cụ trực quan hóa dữ liệu như Tableau, Spotfre, DataRPM và ggplot2 là những công cụ trực quan hóa dữ liệu tuyệt vời để khám phá dữ liệu và xác định các biến mà các DS *có thể* muốn kiểm tra.

**Bước 5: Xây dựng mô hình phân tích.**

sử dụng các công cụ phân tích tiên tiến như SAS, SAS Miner, R, Mahout, MADlib, Alpine Miner, H2O, v.v. để tương quan các biến số khác nhau trong nỗ lực xây dựng các mô hình phân tích. Các DS sẽ khám phá các kỹ thuật và thuật toán phân tích khác nhau để cố gắng tạo ra các mô hình tiên đoán nhất. Một lần nữa, hãy nghĩ đến xác suất, mức độ hạn chế, *F distribution*, *t test* và *p value*. Chương 7 sẽ trình bày một số thuật toán phân tích khác nhau mà các DS có thể sử dụng và trong bối cảnh nào.

**Bước 6: Đánh giá độ tốt của mô hình.**

xác định độ tốt của mô hình của fit. Độ tốt của fit của một mô hình thống kê mô tả mức độ tốt của mô hình so với một tập hợp các quan sát (*F*-test, *p*và *t*-statistic). Một số kỹ thuật phân tích khác nhau sẽ được sử dụng để xác định độ tốt của fit bao gồm kiểm định Kolmogorov – Smirnov, kiểm định chi bình phương của Pearson, phân tích phương sai (ANOVA) và ma trận nhầm lẫn (hoặc sai số)

“Độ tốt của ﬁ t” đo lường mức độ tốt của một mô hình phân tích đối với một tập hợp các quan sát và đo lường mức độ mà dữ liệu quan sát phù hợp với các giá trị được mong đợi hoặc dự đoán bởi mô hình phân tích.

**Mô hình dữ liệu khác nhau**

#### **Mô hình hóa dữ liệu cho BI**

truy vấn, báo cáo, trang tổng quan) yêu cầu kỹ thuật mô hình hóa dữ liệu cho phép người dùng doanh nghiệp tạo báo cáo và truy vấn của riêng họ

=> lược đồ sao

* Lược đồ hình sao bao gồm một hoặc nhiều [bảng dữ kiện](https://wivi.wiki/wiki/Fact_table) tham chiếu đến bất kỳ số lượng [bảng thứ nguyên nào](https://wivi.wiki/wiki/Dimension_(data_warehouse))

Lược đồ hình sao phân tách dữ liệu quy trình kinh doanh thành các dữ liệu, chứa dữ liệu định lượng, có thể đo lường về doanh nghiệp và các thứ nguyên là thuộc tính mô tả liên quan đến dữ liệu thực tế. Ví dụ về dữ liệu thực tế bao gồm giá bán, số lượng bán và các phép đo thời gian, khoảng cách, tốc độ và trọng lượng. Các ví dụ về thuộc tính thứ nguyên có liên quan bao gồm mẫu sản phẩm, màu sản phẩm, kích thước sản phẩm, vị trí địa lý và tên nhân viên bán hàng.

* bảng fact (metrics hoặc thước đo)

transactional systems such as orders, shipments, sales, returns, premiums, claims, accounts receivable, and accounts payable. Facts are typically numeric values that can be aggregated (e.g., averaged, counted, or summed).

* bảng dimension

các thuộc tính về thứ nguyên đó) chứa "danh từ" của hệ thống giao dịch cụ thể đó như sản phẩm, thị trường, cửa hàng, nhân viên, khách hàng và các biến thể khác nhau của thời gian.

Thứ nguyên là nhóm phân cấp và bộ mô tả mô tả các sự kiện. Chính các thuộc tính chiều này cho phép khám phá phân tích, các thuộc tính như kích thước, trọng lượng, vị trí (đường phố, thành phố, tiểu bang, mã vùng), tuổi, giới tính, nhiệm kỳ, v.v.

Mô hình thứ nguyên lý tưởng cho người dùng doanh nghiệp vì nó hỗ trợ câu hỏi tự nhiên của họ- các quá trình thăm dò và trả lời. Mô hình hóa thứ nguyên hỗ trợ các khái niệm BI, chẳng hạn như đi sâu vào (điều hướng qua các thứ nguyên) và đi sâu vào / đi sâu vào (điều hướng lên và xuống phân cấp thứ nguyên, chẳng hạn như phân cấp thứ nguyên sản phẩm của danh mục sản phẩm ➪ thương hiệu).

#### **Mô hình hóa dữ liệu cho DS**

Trong thế giới khoa học dữ liệu, Hadoop cung cấp cơ hội để *suy nghĩ khác* về cách chúng tôi thực hiện mô hình hóa dữ liệu. Hadoop ban đầu được thiết kế bởi Yahoo để xử lý các log web rất dài. Hadoop được thiết kế với các khối dữ liệu rất lớn (kích thước khối mặc định của Hadoop là 64 MB đến 128 MB so với kích thước khối cơ sở dữ liệu quan hệ thường là 32 Kb trở xuống). Để tối ưu hóa lợi thế về kích thước khối này, nhóm khoa học dữ liệu muốn các bản ghi rất dài và các mô hình dữ liệu dài.

Ví dụ: một số DS thích "thu thập" lược đồ hình sao bằng cách cộng hoặc tích hợp các bảng chiều bao quanh bảng dữ kiện thành một bản ghi duy nhất để xây dựng và thực hiện các truy vấn dữ liệu phức tạp hơn mà không cần phải sử dụng các phép nối

Như ví dụ trong Hình 5-10, thay vì ba lược đồ sao khác nhau với các kích thước phù hợp hoặc chia sẻ để liên kết các lược đồ sao khác nhau, nhóm khoa học dữ liệu muốn ba lược đồ dài, với dữ liệu khách hàng sau:

■ Nhân khẩu học của khách hàng (tuổi, giới tính, địa chỉ nhà hiện tại và trước đây, giá trị của ngôi nhà hiện tại và trước đây, lịch sử tình trạng hôn nhân, con cái và tuổi và giới tính của họ, thu nhập hiện tại và trước đây, v.v.)

■ Lịch sử mua hàng của khách hàng (mua hàng năm bao gồm các mặt hàng đã mua, trả lại, giá đã thanh toán, chiết khấu, phiếu giảm giá, địa điểm, ngày trong tuần, thời gian trong ngày, điều kiện thời tiết, nhiệt độ)

■ Các hoạt động xã hội của khách hàng (toàn bộ lịch sử các bài đăng trên mạng xã hội, lượt thích, lượt chia sẻ, lượt tweet, lượt yêu thích, lượt retweet, v.v.)

**Góc nhìn/quan điểm business khác nhau**

Thay vì cố gắng xây dựng “phiên bản duy nhất của sự thật” hoặc tạo ra “góc nhìn 360 độ về khách hàng”, nhóm DS sẽ xây dựng hồ sơ phân tích về từng đơn vị kinh doanh chính của tổ chức hoặc danh từ chiến lược ở cấp độ thực thể riêng lẻ.

Một trong những khái niệm DS mạnh mẽ nhất là cấu hình phân tích. Nhóm DS xây dựng hồ sơ phân tích chi tiết nắm bắt hành vi, xu hướng, sở thích và xu hướng của các thực thể kinh doanh riêng lẻ (ví dụ: khách hàng, thương gia, sinh viên, bệnh nhân, bác sĩ, tuabin gió, động cơ phản lực, máy ATM).

Hồ sơ phân tích là sự kết hợp của các chỉ số, chỉ số hiệu suất chính, điểm số, quy tắc liên kết và thông tin chi tiết phân tích kết hợp với xu hướng, hành vi, xu hướng, liên kết, tình cảm, sở thích và đam mê đối với một thực thể riêng lẻ (khách hàng, thiết bị, đối tác, máy ).

Quá trình phân tích hoạt động như sau:

**Bước 1**: Thiết lập một giả thuyết mà bạn muốn kiểm tra. Trong ví dụ về khách hàng của chúng tôi, giả thuyết thử nghiệm của chúng tôi là "Thành viên thẻ vàng cao cấp có nhiều ngày hơn mà không mua hàng hoặc tương tác với ứng dụng dành cho thiết bị di động có xác suất bỏ cuộc cao hơn từ 25 đến 30% so với những khách hàng tương tự."

**Bước 2**: Xác định và định lượng các chỉ số hoặc điểm số quan trọng nhất để xác định trước một kết quả kinh doanh nhất định. Trong ví dụ của chúng tôi, các chỉ số và điểm số mà chúng tôi sẽ sử dụng để kiểm tra giả thuyết về mức độ tiêu thụ của khách hàng bao gồm Điểm hài lòng của khách hàng, Số lần mua hàng trung bình hàng tháng và Điểm mức độ trung thành của khách hàng. Lưu ý rằng các chỉ số không có cùng trọng số (hoặc mức quy định). Một số chỉ số và điểm số quan trọng hơn những số liệu khác trong việc dự đoán hiệu suất với giả thuyết kiểm tra.

**Bước 3**: Sử dụng các số liệu dự đoán để xây dựng hồ sơ chi tiết cho từng khách hàng riêng lẻ liên quan đến giả thuyết được kiểm tra.

* Dự đoán các tình huống tiêu hao của khách hàng
* Xác định và định lượng các biến dự báo
* Xây dựng hồ sơ khách hàng cá nhân
* So sánh các hoạt động hiện tại của khách hàng với hồ sơ để chấm điểm các điểm bất thường
* Tinh chỉnh (kiểm tra / học hỏi) các biến cấu hình

Giảm phương sai

Thêm các biến mới

Xóa các biến

* Tích hợp điểm số giữ chân khách hàng vào các hệ thống hoạt động (tiếp thị, trung tâm cuộc gọi, POS)

**Bước 4:** So sánh các hoạt động gần đây và trạng thái hiện tại của một cá nhân với hồ sơ của họ để xác định các hành vi và hành động bất thường có thể là dấu hiệu của vấn đề giữ chân khách hàng. Trong ví dụ về tỷ lệ giữ chân khách hàng, chúng tôi có thể muốn tạo điểm “Sự chú ý của khách hàng” đánh giá khả năng khách hàng cụ thể đó sẽ rời đi và sau đó tạo các đề xuất cụ thể về những hành động hoặc “ưu đãi tốt nhất tiếp theo” có thể được gửi đến giữ chân khách hàng đó.

**Bước 5:** Tiếp tục tìm kiếm các nguồn dữ liệu mới và các chỉ số mới có thể là những yếu tố dự đoán tốt hơn về mức tiêu hao. Đây cũng là một phần của quy trình DS nhằm liên tục cố gắng cải thiện độ chính xác và mức độ hội tụ của các chỉ số và điểm số bằng cách sử dụng phân tích và mô phỏng độ nhạy như thí nghiệm Monte Carlo.

**Bước 6:** Tích hợp thông tin chi tiết về phân tích, điểm số và khuyến nghị vào các hệ thống hoạt động chính (có thể là CRM, tiếp thị trực tiếp, điểm bán hàng và trung tâm cuộc gọi cho sáng kiến ​​kinh doanh giữ chân khách hàng) để đảm bảo rằng thông tin chi tiết được khám phá bởi phân tích là có thể được hành động bởi nhân viên tuyến đầu hoặc nhân viên gắn bó với khách hàng.

chương 6

Chương 6 giới thiệu một số thuật toán phân tích khác nhau mà người dùng doanh nghiệp nên biết và thảo luận về thời điểm thích hợp nhất để sử dụng loại thuật toán nào. Bạn không cần phải là một nhà DS để hiểu khi nào và tại sao áp dụng các thuật toán phân tích này. Hiểu biết chi tiết hơn về các thuật toán phân tích khác nhau này sẽ giúp người dùng doanh nghiệp cộng tác với nhóm DS để khám phá các biến và chỉ số có thể là những yếu tố dự đoán tốt hơn về hiệu suất kinh doanh.

■ Mang đến trải nghiệm tích cực và hấp dẫn hơn cho khách trong một thị trường giải trí ngày càng cạnh tranh;

■ Xác định giá trị trọn đời của khách tiềm năng tối đa (MPGTV) để sử dụng làm cơ sở xác định chi tiêu khuyến mại và giảm giá của khách, đồng thời ưu tiên các vé vào cửa Ưu tiên và phòng khách sạn trong công viên;

■ Quảng cáo các điểm tham quan 3D nhiều công nghệ mới (Terror Airline và Zombie Apocalypse) để đảm bảo việc áp dụng thành công và khả năng tồn tại lâu dài của những trò chơi mới thu hút các phân khúc khách mới;

■ Đảm bảo sự thành công của các nhân vật phim và truyền hình mới nhằm tăng doanh thu cấp phép liên quan và đảm bảo khả năng tồn tại lâu dài của nhân vật cho các phần phim và chương trình truyền hình mới.

Công viên đang triển khai một ứng dụng di động có tên Fairy-Tale Chaperon để thu hút khách khi họ di chuyển qua công viên và giúp khách tận hưởng các điểm hấp dẫn, giải trí, cửa hàng bán lẻ và nhà hàng khác nhau. Fairy-Tale Chaperon sẽ:

■ Cung cấp vé vào cửa Ưu tiên cho các điểm tham quan khác nhau và thưởng cho những vị khách quan trọng nhất của họ bằng các phiếu giảm giá kỹ thuật số, chiết khấu và “Fairy Dust” (số tiền tương đương chỉ có thể chi tiêu trong công viên).

■ Quảng cáo các bài đăng trên mạng xã hội để thúc đẩy gamifcation và phần thưởng xung quanh các cuộc thi, chẳng hạn như hầu hết các bài đăng trên mạng xã hội, các bài đăng phổ biến nhất trên mạng xã hội cũng như ảnh và video phổ biến nhất.

■ Theo dõi các mô hình, xu hướng và xu hướng bò của khách và trong công viên để xác định điểm tham quan nào cần quảng bá (để tăng sức hút) và điểm tham quan nào khách nên tránh vì thời gian chờ đợi lâu.

■ Đưa ra các đề xuất ăn uống và giải trí cho khách trong thời gian thực dựa trên lĩnh vực khách quan tâm và chỗ ngồi / bàn còn trống cho một số nhà hàng và khu giải trí.

■ Thưởng cho những khách chia sẻ thông tin trên mạng xã hội có thể được sử dụng để theo dõi mức độ hài lòng và thích thú của khách trong thời gian thực qua Facebook, Twitter và Instagram. Nó cũng tạo cơ hội để quảng bá những bức ảnh được chọn để bắt đầu các chiến dịch tiếp thị lan truyền.

**Các thuật toán EDA:**

■ Trend analysis

■ Boxplots

■ Geography (spatial) analysis

■ Pairs plot

■ Time series decomposition

More advanced analytic algorithms that are covered in this chapter are:

■ Cluster analysis

■ Normal curve equivalent (NCE) analysis

■ Association analysis

■ Graph analysis

■ Text mining

■ Sentiment analysis

■ Traverse pattern analysis

■ Decision tree classifier analysis

■ Cohorts analysis