

Bài 1- Cơ bản về phân tích dữ liệu

A.Lý Thuyết

1. Phân loại các kiểu Data Analytics

Data Analytics có thể được phân loại thành bốn loại chính với các **mục tiêu** khác nhau và **phương pháp** thực hiện khác nhau

- **Descriptive Analytics (Phân tích mô tả):**
 - **Mục tiêu:** tập trung vào việc mô tả và hiểu các sự kiện đã xảy ra trong quá khứ từ đó tạo thành một bức tranh rõ ràng về những gì đã xảy ra trong một khoảng thời gian cụ thể
 - **Công cụ:** phổ biến bao gồm báo cáo, tổng hợp dữ liệu, và biểu đồ.
 - **Diagnostic Analytics (Phân tích chẩn đoán):**
 - **Mục tiêu:** đi sâu vào nguyên nhân của các sự kiện hoặc kết quả cụ thể. Nó không chỉ tập trung vào việc "cái gì" và "khi nào" mà còn cố gắng giải thích "tại sao".
 - **Công cụ:** Các phương pháp như phân tích nguyên nhân, phân tích tương quan, và đánh giá rủi ro thường được sử dụng.
 - **Predictive Analytics (Phân tích dự đoán):**
 - **Mục tiêu:** dự đoán sự kiện hoặc xu hướng tương lai dựa trên dữ liệu hiện tại và lịch sử.
 - Các mô hình thống kê, máy học, và khoa học dữ liệu được áp dụng để xác định khả năng xảy ra của các sự kiện tương lai.
 - **Prescriptive Analytics (Phân tích chỉ đạo):**
 - **Mục tiêu:** Mức độ phức tạp và tiên tiến nhất, nó không chỉ dự đoán tương lai mà còn đề xuất các hành động cụ thể để đạt được kết quả mong muốn hoặc giảm thiểu rủi ro.
 - **Phương pháp:** Nó kết hợp cả dữ liệu, kỹ thuật phân tích dự đoán và các quy tắc kinh doanh để đưa ra quyết định.
-

Ví dụ:

- **Descriptive Analytics (Phân tích mô tả):**
 - **Ví dụ:** Một công ty bán lẻ muốn hiểu về xu hướng doanh số bán hàng trong 12 tháng qua. Phân tích mô tả giúp họ tổng hợp dữ liệu bán hàng, biểu diễn thông qua các biểu đồ và báo cáo tổng hợp, từ đó hiển thị mức tăng trưởng doanh số theo từng quý, sản phẩm bán chạy nhất, và khu vực có doanh số cao nhất.
- **Diagnostic Analytics (Phân tích chẩn đoán):**
 - **Ví dụ:** Sau khi nhận ra rằng doanh số bán hàng của một sản phẩm cụ thể giảm trong quý vừa qua, công ty quyết định sử dụng phân tích chẩn đoán để xác định nguyên nhân. Bằng cách phân tích các yếu tố như chiến dịch tiếp thị, đánh giá của khách hàng, và giá cả so với đối thủ, họ có thể xác định rằng một chiến dịch tiếp thị không hiệu quả là nguyên nhân chính dẫn đến sự sụt giảm.

- **Predictive Analytics (Phân tích dự đoán):**
 - **Ví dụ:** Một ngân hàng muốn dự đoán khả năng vỡ nợ của khách hàng để quản lý rủi ro tín dụng tốt hơn. Sử dụng phân tích dự đoán, họ áp dụng các mô hình máy học vào lịch sử tín dụng, giao dịch, và dữ liệu cá nhân của khách hàng để dự đoán những khách hàng có nguy cơ cao vỡ nợ trong tương lai.
 - **Prescriptive Analytics (Phân tích chỉ đạo):**
 - **Ví dụ:** Một công ty sản xuất muốn tối ưu hóa chuỗi cung ứng của mình để giảm chi phí và cải thiện hiệu quả. Sử dụng phân tích chỉ đạo, công ty này có thể kết hợp dữ liệu về nhu cầu thị trường, thời gian vận chuyển, chi phí sản xuất và vận chuyển, cùng với các ràng buộc về tài nguyên, để đề xuất một kế hoạch sản xuất và phân phối tối ưu. Mô hình này có thể đề xuất số lượng sản phẩm cần sản xuất, thời điểm sản xuất, và cách phân phối sản phẩm một cách hiệu quả nhất để đáp ứng nhu cầu thị trường với chi phí thấp nhất.
-

2. Quy trình để thực hiện phân tích

Số lượng bước cụ thể tùy thuộc theo phương pháp tiếp cận tuy nhiên quy trình phân tích dữ liệu thường bao gồm 5 bước chính:

1. **Define Objectives (Xác định mục tiêu):** Xác định mục tiêu rõ ràng cho dự án phân tích, bao gồm những gì bạn muốn tìm hiểu hoặc giải quyết từ dữ liệu.
 2. **Data Collection (Thu thập dữ liệu):** Bước này bao gồm việc thu thập dữ liệu từ các nguồn khác nhau, bao gồm dữ liệu nội bộ và dữ liệu bên ngoài, để tạo ra một tập dữ liệu đầy đủ cho phân tích.
 3. **Data Processing and Cleaning (Xử lý và làm sạch dữ liệu):** Làm sạch và xử lý dữ liệu để loại bỏ lỗi, dữ liệu thiếu, hoặc dữ liệu nhiễu, đồng thời chuẩn hóa dữ liệu để phân tích có thể được thực hiện một cách hiệu quả.
 4. **Data Analysis (Phân tích dữ liệu):** Sử dụng các phương pháp thống kê, máy học, hoặc phân tích dữ liệu khác để khám phá và mô hình hóa dữ liệu, nhằm trích xuất thông tin hữu ích hoặc nhận diện xu hướng/mẫu.
 5. **Presentation and Application (Trình bày và áp dụng kết quả):** Trình bày kết quả của việc phân tích trong một dạng dễ hiểu (ví dụ: báo cáo, biểu đồ, dashboard) và đề xuất các hành động dựa trên những phát hiện, nhằm hỗ trợ quyết định hoặc cải thiện quy trình kinh doanh.
-

Ví dụ:

Để thực hiện phân tích doanh số từ file Excel chứa dữ liệu thông tin các mặt hàng đã bán trong tháng 12/2023, bạn có thể tuân theo quy trình phân tích dữ liệu như sau:

1. **Xác định Mục Tiêu:**
 - Mục tiêu ở đây có thể là hiểu rõ về doanh số bán hàng trong tháng 12/2023, bao gồm:
 - Tổng doanh số bán hàng trong tháng.
 - Các sản phẩm bán chạy nhất.
 - Xu hướng bán hàng theo từng tuần trong tháng.

- So sánh doanh số giữa các sản phẩm hoặc danh mục sản phẩm.

2. Thu Thập Dữ Liệu:

- Trong trường hợp này, dữ liệu đã được thu thập và lưu trữ trong một file Excel. Bạn cần đảm bảo rằng file chứa tất cả thông tin cần thiết cho phân tích, bao gồm tên sản phẩm, số lượng bán, giá bán, ngày bán, và bất kỳ thông tin phụ trợ nào khác.

3. Xử Lý và Làm Sạch Dữ Liệu:

- Mở file Excel và xem xét dữ liệu:
 - Loại bỏ hoặc điều chỉnh bất kỳ mục nào bị thiếu hoặc sai lệch.
 - Định dạng lại ngày tháng, số lượng, và giá bán nếu cần.
 - Kiểm tra và loại bỏ dữ liệu trùng lặp.
 - Đảm bảo rằng tất cả dữ liệu được tổ chức một cách nhất quán.

4. Phân Tích Dữ Liệu:

- Sử dụng các công cụ phân tích như Excel, Power BI, hoặc Python (pandas, numpy) để:
 - Tính tổng doanh số bán hàng trong tháng.
 - Xác định các sản phẩm bán chạy nhất dựa trên số lượng và doanh thu.
 - Phân tích xu hướng bán hàng theo tuần để xem có sự biến động nào không.
 - So sánh doanh số giữa các danh mục sản phẩm.

5. Trình Bày và Áp Dụng Kết Quả:

- Tạo báo cáo hoặc dashboard trực quan để trình bày kết quả phân tích:
 - Biểu đồ tổng doanh số bán hàng trong tháng.
 - Biểu đồ top sản phẩm bán chạy.
 - Biểu đồ xu hướng bán hàng theo tuần.
 - So sánh doanh số giữa các danh mục sản phẩm.
- Đề xuất hành động dựa trên kết quả phân tích, như điều chỉnh chiến lược bán hàng, tăng cường quảng cáo cho sản phẩm bán chậm, hoặc điều chỉnh giá sản phẩm.