## **Chọn chủ đề: Phân tích xu hướng mua hàng thời trang trực tuyến tại các trang thương mại điện tử trong tương lai gần (3-6 tháng)**

**1. Giới thiệu**

**Lý do chọn đề tài:**

* Ngành thời trang luôn thay đổi nhanh chóng, việc dự đoán xu hướng giúp doanh nghiệp đưa ra quyết định kinh doanh hiệu quả.

**2. Hiểu về Business**

**Mô tả vấn đề:**

* Dự đoán xu hướng thời trang nào sẽ phổ biến trong tương lai. ( màu sắc, kiểu dáng, chất liệu vvv)
* Xác định các yếu tố ảnh hưởng đến xu hướng thời trang.
* Cần biết mặt hàng nào bán vào lúc nào.

**Mục tiêu kinh doanh:**

* Tăng doanh số bán hàng.
* Giảm thiểu rủi ro trong việc lựa chọn sản phẩm trong mùa tới
* Cải thiện hiệu quả chuỗi cung ứng.

**3. Phương pháp tiếp cận phân tích**

* Phân tích chuỗi thời gian để dự đoán nhu cầu, xem xét chu kỳ, trend cho từng dòng sản phẩm. Dựa vào đó để theo dõi tồn kho và cải thiện quy trình sản xuất và phân phối sản phẩm phù hợp
* Tìm ra mối quan hệ giữa yếu tố tác động đến thời trang ví dụ màu sắc, chất liệu, độ tuổi, giới tính vvv
* Phân tích mạng xã hội bằng mô hình xử lý ngôn ngữ tự nhiên NLP để xác định nhu cầu hiện tại, xu hướng của thị trường, đánh giá sản phẩm vvv

**Chuỗi thời gian - Time series:**

Phân tích chuỗi thời gian là quá trình nghiên cứu và đánh giá các biến thay đổi theo thời gian để hiểu rõ xu hướng, mô hình, và các yếu tố ảnh hưởng đến sự biến động của dữ liệu.

* Chuỗi Thời Gian (Time Series): Dữ liệu được thu thập liên tục theo thời gian, thường được sắp xếp theo thứ tự thời gian.
* Tính Chu Kỳ (Seasonality): Biến động theo chu kỳ có thể xuất hiện trong chuỗi thời gian, thường do yếu tố mùa vụ hoặc các yếu tố chu kỳ khác.
* Trend (Xu Hướng): Hướng diễn biến dài hạn của chuỗi thời gian, thể hiện sự tăng hoặc giảm theo thời gian.

*Ứng dụng*

* Dự Báo (Forecasting): Dự đoán giá trị tương lai dựa trên xu hướng và mô hình trong chuỗi thời gian.
* Kiểm Định Xu Hướng và Chu Kỳ (Trend and Seasonality Analysis):Phân tích xu hướng và chu kỳ để hiểu rõ sự biến động trong dữ liệu..
* Kiểm Soát Chất Lượng (Quality Control):Đánh giá và theo dõi sự biến động của dữ liệu để giữ cho quá trình sản xuất hoặc dịch vụ duy trì chất lượng.
* Thị Trường Tài Chính (Financial Markets): Dự báo giá chứng khoán, biểu đồ giá, và các yếu tố tài chính khác.
* Quản Lý Chuỗi Cung Ứng (Supply Chain Management): Dự báo nhu cầu, theo dõi tồn kho, và tối ưu hóa quy trình sản xuất và phân phối.

**Mô hình xử lý ngôn ngữ tự nhiên (NLP)**

Mô hình xử lý ngôn ngữ tự nhiên (NLP) là các hệ thống máy học được thiết kế để hiểu, tìm hiểu và tương tác với ngôn ngữ tự nhiên của con người.

*Ứng dụng*

* Tokenization: Chia văn bản thành các đơn vị nhỏ như từ hoặc câu.
* Lemmatization và Stemming: Rút gọn từ về dạng cơ bản.
* Phân loại văn bản: Xác định loại văn bản, ví dụ: tin tức, đánh giá sản phẩm, thư rác.
* Phân nhóm: Gán các mục văn bản vào các nhóm hoặc chủ đề khác nhau.
* Dự báo cảm xúc: Phân loại văn bản dựa trên cảm xúc, ví dụ: tích cực, tiêu cực, trung lập.

**4. Yêu cầu dữ liệu**

**Loại dữ liệu:**

* Dữ liệu mua sắm trực tuyến (lịch sử mua hàng, sản phẩm xem qua, độ tuổi và giới tính người dùng...)
* Giá trị hàng được mua
* Dữ liệu sản phẩm (mô tả, hình ảnh,...) của 1000 sản phẩm bán chạy nhất trong vòng 3 năm qua
* Dữ liệu bình luận của khách hàng sau khi mua để phân loại các sản phẩm theo xu hướng

**Yêu cầu dữ liệu khác:**

* Dữ liệu phải đầy đủ và chính xác.
* Dữ liệu phải được cập nhật thường xuyên.

**5. Thu thập dữ liệu**

**Phương pháp thu thập:**

* Thu thập dữ liệu từ trang web bán hàng trực tuyến như lazada, shopee, tiki mục thời trang quần áo
* Mua dữ liệu từ các nhà cung cấp dữ liệu.

**Phân tích dữ liệu:**

* Dữ liệu mua sắm trực tuyến được sử dụng để phân tích xu hướng theo thời gian, phân loại người dùng, giới tính , độ tuổi, dự đoán nhu cầu trong tương lai gần
* Dữ liệu sản phẩm được sử dụng để phân loại các sản phẩm theo xu hướng, xác định các yếu tố ảnh hưởng
* Dữ liệu mạng xã hội được sử dụng để xác định các xu hướng thời trang mới nổi, những trend đang được quan tâm

**Thống kê, mô tả dữ liệu:**

* Dữ liệu mua sắm trực tuyến được thống kê theo doanh số bán hàng, sản phẩm, khách hàng, mùa,
* Dữ liệu sản phẩm được thống kê theo loại sản phẩm, thương hiệu, giá cả, chất liệu, màu sắc..
* Dữ liệu mạng xã hội được thống kê theo số lượng bài đăng, bình luận, lượt thích,...

**Vấn đề dữ liệu và giải pháp:**

* Dữ liệu có thể thiếu sót hoặc không chính xác.
* Dữ liệu có thể bị nhiễu.

**Giải pháp:**

* Sử dụng các kỹ thuật xử lý dữ liệu để loại bỏ dữ liệu thiếu sót, không chính xác và nhiễu.
* Sử dụng các kỹ thuật thống kê để phân tích dữ liệu.