

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ CƯƠNG ĐỒ ÁN NHẬP MÔN KHOA HỌC DỮ LIỆU
NHÓM 02

1. Chủ đề thực hiện:

Trong bối cảnh thị trường thiết bị điện tử cạnh tranh sôi động, cửa hàng cần hiểu rõ xu hướng và nhu cầu của khách hàng để tối ưu hóa chiến lược kinh doanh và duy trì lợi thế cạnh tranh. Dự án "Phân tích và đánh giá thị trường thiết bị điện tử tại Việt Nam" sẽ giúp cửa hàng xây dựng chiến lược bán hàng và tiếp thị phù hợp, nâng cao hiệu quả kinh doanh.

2. Nguồn dữ liệu:

Bộ dữ liệu của dự án được thu thập từ các trang thương mại điện tử như MobileCity, thegioididong.com và các cửa hàng tương tự, sử dụng các công cụ tự động như Selenium và BeautifulSoup để cào dữ liệu. Bộ dữ liệu bao gồm các thuộc tính như tên sản phẩm, giá bán, màu sắc, thời gian bảo hành, thông số kỹ thuật, đánh giá của người dùng và đường dẫn sản phẩm. Tuy nhiên, dữ liệu thiếu thông tin số lượt mua và không đồng nhất, gây khó khăn trong thu thập do số lượng trường hạn chế, dữ liệu thay đổi và chính sách bảo mật của các cửa hàng, ảnh hưởng đến tính toàn diện và độ chính xác phân tích.

3. Kết quả kỳ vọng

- Giải quyết khó khăn trong thu thập và xử lý dữ liệu, bao gồm thiếu dữ liệu và dữ liệu không đồng nhất. Cải thiện độ chính xác và đảm bảo tính toàn vẹn của dữ liệu để phân tích có giá trị.
- Phân tích: xu hướng thị trường thiết bị điện tử tại Việt Nam, xác định dòng sản phẩm phổ biến, mức giá trung bình, và các nhà cung cấp hàng đầu; các yếu tố ảnh hưởng đến quyết định mua hàng, cung cấp thông tin chuyên sâu về thị hiếu người tiêu dùng.
- Xây dựng mô hình: dự đoán giá sản phẩm và hệ thống gợi ý sản phẩm dựa trên nhu cầu, sở thích của khách hàng.
- Ứng dụng vào chiến lược kinh doanh và tiếp thị: đề xuất các chiến lược tiếp thị hiệu quả nhằm thu hút khách hàng mục tiêu và tối ưu hóa khả năng tiếp cận thị trường của các sản phẩm mới.

4. Kế hoạch thực hiện

STT	Tên công việc	Mô tả	Ngày hoàn thành
1	Xác định chủ đề	Xác định rõ mục tiêu và phạm vi của dự án, định hướng nghiên cứu cụ thể để giải quyết vấn đề đã đề ra.	20/10/2024
2	Xây dựng đề cương đồ án	Lập đề cương chi tiết cho dự án, trình bày cách thức tiến hành và các bước trong quy trình phân tích dữ liệu.	27/10/2024
3	Thu thập dữ liệu	Sử dụng các công cụ như Selenium, Requests hoặc Beautiful Soup để thu thập dữ liệu từ các trang thương mại điện tử, đảm bảo dữ liệu có đủ thông tin cần thiết cho phân tích.	03/11/2024
4	Tiền xử lý dữ liệu và khám phá dữ liệu	Làm sạch và chuẩn hóa dữ liệu, thực hiện phân tích sơ bộ để tìm hiểu các trường dữ liệu và phát hiện các vấn đề như giá trị thiếu hoặc loại dữ liệu không phù hợp. Áp dụng kỹ thuật tạo đặc trưng như trích xuất và chọn lọc đặc trưng để chuẩn bị dữ liệu cho các bước phân tích tiếp theo.	10/11/2024
5	Đặt câu hỏi phân tích	Đặt ra ít nhất 5 câu hỏi ý nghĩa có thể trả lời bằng dữ liệu đã thu thập. Mỗi câu hỏi cần nêu rõ mục đích, phương pháp phân tích và cách thức trực quan hóa để đưa ra kết luận.	17/11/2024
6	Phân tích và trực quan hóa dữ liệu	Sử dụng các công cụ trực quan hóa như Matplotlib hoặc Seaborn để minh họa dữ liệu, giúp dễ quan sát và đưa ra các kết luận.	24/11/2024
7	Xây dựng và đánh giá mô hình dữ liệu	Xây dựng các mô hình dự đoán hoặc phân loại bằng các thuật toán học máy như hồi quy, phân loại. Đánh giá mô hình bằng cross-validation và các chỉ số như độ chính xác, độ nhạy, MSE, RMSE.	08/12/2024
8	Đánh giá kết quả và báo cáo	Trình bày kết quả phân tích và mô hình hóa qua báo cáo và biểu đồ, nhấn mạnh các phát hiện quan trọng, so sánh các mô hình và nêu rõ ưu, nhược điểm của từng mô hình.	15/12/2024
9	Phản hồi và rút kinh nghiệm	Sau khi hoàn thành dự án, mỗi thành viên trong nhóm cần đánh giá lại những khó khăn gặp phải và bài học kinh nghiệm, từ đó đề xuất các cải thiện nếu có thêm thời gian.	15/12/2024
10	Thuyết trình về dự án	Chuẩn bị thuyết trình ngắn gọn, tập trung vào kết quả và những nội dung chính của dự án, minh họa bằng biểu đồ và giải thích.	22/12/2024

Bảng 1: Kế hoạch thực hiện đồ án