Big Data Course

Capstone Project   
Action Plan

For students (instructor’s review required)

ⓒ2023 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung Innovation Campus, you must receive written consent from copyright holder.

|  |  |
| --- | --- |
| Course | Big Data Course |
| Team Name | Nhóm 5-Bigdata |
| Team Leader/  Members | Lê Anh Xuân /  Nguyễn Kiều Phước,  Trần Thị Thanh Thảo,  Phú Thanh Thiện,  Hồ Đức Trí,  Nguyễn Văn Hải |
| Project Title | **Phân tích dữ liệu đơn hàng thương mại điện tử và xây dựng mô hình dự đoán sản phẩm** |
| Goal |  |
| -Tìm hiểu về các kỹ thuật phân tích dữ liệu từ Bigdata  -Nghiên cứu về mô hình Machine Learning  -Phát triển 1 mô hình dự đoán - Recommendation | |
| Abstract |  |
| Project này tập trung vào việc áp dụng các kỹ thuật phân tích dự đoán trong lĩnh vực Big Data để xây dựng một hệ thống gợi ý. Bằng cách sử dụng các thuật toán Machine Learning, Hệ thống gợi ý sẽ sử dụng phương pháp lọc cộng tác để cá nhân hóa trải nghiệm người dùng, tăng cường sự hài lòng và tối ưu hóa doanh thu cho các doanh nghiệp. | |
| Method |  |
| -Thu thập dữ liệu từ các nguồn mở để lấy dữ liệu về hành vi mua sắm của khách hàng, bao gồm thông tin sản phẩm, lịch sử giao dịch và đánh giá sản phẩm.  -Các phương pháp tiền xử lý dữ liệu như loại bỏ các giá trị thiếu, dữ liệu không hợp lệ và chuẩn hóa định dạng.  -Trực quan hóa dữ liệu  -Sử dụng phương pháp lọc cộng tác để xây dựng mô hình dự đoán  -Triển khai mô hình lên web | |

|  |  |
| --- | --- |
| Data |  |
| Dữ liệu cho dự án này sẽ được thu thập từ nền tảng Kaggle, nơi cung cấp nhiều bộ dữ liệu phong phú và đa dạng liên quan đến thương mại điện tử và hành vi mua sắm của khách hàng. Các bộ dữ liệu này thường bao gồm thông tin về sản phẩm, lịch sử giao dịch, đánh giá của khách hàng, và các yếu tố khác.  Dữ liệu sẽ được sử dụng để thực hiện phân tích mô tả nhằm hiểu rõ hơn về hành vi mua sắm của khách hàng và các yếu tố ảnh hưởng đến quyết định mua hàng.  Sử dụng dữ liệu để phát triển mô hình gợi ý thông qua các thuật toán lọc cộng tác và Machine Learning. Dữ liệu sẽ giúp xác định các mẫu hành vi và sở thích của khách hàng. | |
| Expected  Outcome |  |
| < Describe in a few sentences the outcome you expect from your project.>  <Also explain about the benefits if such outcome were to be obtained.>  Từ dự án này, tôi kỳ vọng sẽ phát triển một hệ thống gợi ý thông minh có khả năng dự đoán chính xác hành vi mua sắm của khách hàng dựa trên dữ liệu lớn. Hệ thống này sẽ sử dụng các phương pháp lọc cộng tác để cung cấp các gợi ý sản phẩm cá nhân hóa, giúp nâng cao trải nghiệm người dùng.  Tăng cường Trải nghiệm Khách hàng:  Hệ thống gợi ý sẽ giúp khách hàng tìm thấy các sản phẩm phù hợp với sở thích và nhu cầu của họ, từ đó nâng cao sự hài lòng và trung thành với thương hiệu.  Tối ưu hóa Doanh thu:  Bằng cách cung cấp các gợi ý chính xác, doanh nghiệp có thể tăng tỷ lệ chuyển đổi và doanh thu, nhờ vào việc khuyến khích khách hàng mua sắm nhiều hơn.  Hiểu biết Sâu về Hành vi Khách hàng:  Dự án sẽ cung cấp cái nhìn sâu sắc về hành vi và sở thích của khách hàng, giúp doanh nghiệp đưa ra các quyết định chiến lược tốt hơn trong tiếp thị và phát triển sản phẩm. | |
| Role by  Member |  |
| <The role assigned to each team member.>   |  |  | | --- | --- | | **Member** | **Role** | | Lê Anh Xuân | Xây dựng mô hình dự đoán | | Nguyễn Kiều Phước, | Xây dựng web | | Trần Thị Thanh Thảo, | Phân tích dữ liệu | | Phú Thanh Thiện, | Tiền xử lý dữ liệu | | Hồ Đức Trí, | Triển khai mô hình lên web | | Nguyễn Văn Hải | Thu thập dữ liệu | | |

|  |  |
| --- | --- |
| Schedule  Summary |  |
| <Provide a summarized schedule.>  <Use the provided WBS workbook for more detailed action plan.>  Trong dự án này, chúng tôi sẽ bắt đầu bằng việc thu thập dữ liệu từ Kaggle và đưa dữ liệu vào kho dữ liệu (data warehouse). Tiếp theo, chúng tôi sẽ tiến hành tiền xử lý dữ liệu, bao gồm xử lý dữ liệu thiếu và trùng lặp, tổng hợp dữ liệu, chuẩn hóa dữ liệu, xử lý dữ liệu thừa và giảm chiều dữ liệu. Sau khi hoàn thành bước tiền xử lý, chúng tôi sẽ thực hiện phân tích dữ liệu, tập trung vào phân tích dữ liệu đơn hàng, dữ liệu khách hàng và dữ liệu sản phẩm. Tiếp theo, chúng tôi sẽ xây dựng mô hình dự đoán bằng cách chia tập dữ liệu và áp dụng các mô hình SVD và KNN, sau đó kiểm tra hiệu quả của các mô hình này. Cuối cùng, chúng tôi sẽ triển khai mô hình lên web bằng cách xây dựng trang web và đưa mô hình lên trang. Đồng thời, chúng tôi sẽ viết báo cáo cho từng phần của dự án và tạo slide thuyết trình cho các phần đã hoàn thành. | |
| Comment &  Assessment |  |
| <Comment and assessment **by the instructor.**> | |