

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN MÔN HỌC
LẬP TRÌNH CHO KHOA HỌC DỮ LIỆU

Sinh viên thực hiện:

23120038 – Lê Hoàng Mỹ Hạ

23120084 – Nguyễn Mạnh Thắng

Lớp: CQ2023/21

MỤC LỤC

| | |
|---|----------|
| 1. Thông tin thành viên..... | 3 |
| 2. Kế hoạch thực hiện và phân công chi tiết..... | 3 |
| 3. Quy trình hợp tác..... | 4 |
| 3.1. Công cụ hỗ trợ..... | 4 |
| 3.2. Phương pháp làm việc nhóm | 4 |
| 4. Tài liệu tham khảo | 5 |

1. Thông tin thành viên

| MSSV | Họ và tên | Mức độ hoàn thành |
|----------|-------------------|-------------------|
| 23120038 | Lê Hoàng Mỹ Hạ | 100% |
| 23120084 | Nguyễn Mạnh Thắng | 100% |

2. Kế hoạch thực hiện và phân công chi tiết

Dưới đây là bảng tổng hợp tiến độ thực hiện dự án theo từng tuần và phân công nhiệm vụ cụ thể cho từng thành viên:

| Thời gian và nhiệm vụ chính | Nhiệm vụ cụ thể | Người thực hiện | Mức độ hoàn thành |
|---|--|-----------------|-------------------|
| Thời gian: (22/11 – 29/11) Nhiệm vụ chính: Data Collection Data exploration | 2.1. Tổng quan về bộ dữ liệu | Thắng | 100% |
| | 2.2. Phân tích biến số số (Numerical Columns) | Hạ | 100% |
| | 2.3. Phân tích Cột Phân loại (Categorical Columns) | Thắng | 100% |
| | 2.4. Missing Data | Hạ | 100% |
| | 2.5. Mối quan hệ & Tương quan (Relationships & Correlations) | Thắng | 100% |
| | 2.6. Initial Observations & Insights | Hạ | 100% |
| Thời gian: (30/11 – 06/12) Nhiệm vụ chính: Question fomulation | 3.1. Câu hỏi 1 | Hạ | 100% |
| | 3.2. Câu hỏi 2 | Hạ | 100% |
| | 3.3. Câu hỏi 3 | Thắng | 100% |
| | 3.4. Câu hỏi 4 (Câu hỏi dùng mô hình machine learning) | Hạ, Thắng | 100% |

| | | | |
|---|--------------------|-----------|------|
| Thời gian: (07/12 – 14/12) | 4.1. Preprocessing | Thắng | 100% |
| Nhiệm vụ chính: Data Analysis | 4.2. Train Model | Hạ | 100% |
| Thời gian: (15/12 – 22/12) | 5. Project Summary | Hạ, Thắng | 100% |
| Nhiệm vụ chính: Project Summary | Viết báo cáo | Hạ, Thắng | 100% |
| Hoàn thiện đồ án | Viết Readme.md | Thắng | 100% |

3. Quy trình hợp tác

Để đảm bảo dự án vận hành tốt, nhóm đã thống nhất quy trình làm việc như sau:

3.1. Công cụ hỗ trợ

- **Quản lý mã nguồn (Source Control):** Sử dụng **GitHub** để lưu trữ code, quản lý phiên bản và merge các phần làm việc của thành viên.
- **Môi trường lập trình:** Jupyter Notebook và VS Code.
- **Giao tiếp:** Messenger để trao đổi nhanh hàng ngày.

3.2. Phương pháp làm việc nhóm

- **Chia để trị (Divide and Conquer):** Dự án được chia nhỏ thành các module độc lập (EDA, Preprocessing, Modeling) để hai thành viên có thể làm việc song song mà không bị xung đột code.
- **Code Review:** Trước khi merge vào nhánh chính (main), code của thành viên này phải được thành viên kia xem qua (pull request) để phát hiện lỗi.
- **Giải quyết xung đột:** Trong những vấn đề chung về hướng giải quyết đồ án (ví dụ: cách xử lý Outliers, Missing values), nhóm sẽ cùng thảo luận để đưa ra phương án tốt nhất.

4. Tài liệu tham khảo

- [1] <https://www.kaggle.com/code/jay25patel/spotify-music-analysis-2009-2025-eda-ml-pipeli>