

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

Khoa Công nghệ thông tin



**BÁO CÁO THỰC HÀNH**

**ĐỒ ÁN CUỐI KÌ**

**| CHỦ ĐỀ |**

**PHÂN TÍCH DỮ LIỆU VỀ CHẤT LƯỢNG**

**KHÔNG KHÍ**

*( Thời gian hoàn thành quy định: 04 Tháng 01 Năm 2024 )*

**Giảng viên hướng dẫn: Thầy Lê Ngọc Thành – Cô Nguyễn Thị Thu Hằng**

**Môn: NHẬP MÔN KHOA HỌC DỮ LIỆU**

**Thành viên nhóm:**

- Hồ Sỹ Kiên - 21120091
- Hoàng Thành Nam - 21120099
- Nguyễn Hoài An - 21120035
- Đinh Thị Thúy Hường - 21120176

Thành phố Hồ Chí Minh – 2024

Phần 1: Bảng phân công công việc

Tasks

All tasks Board +

	Aa Task name	Assignee	Due
☑	Data Collecting	K Kiên Hồ Sỹ	December 10, 2023 → December 13, 2023
☑	First Explore (Meaning, Shape)	K Kiên Hồ Sỹ	December 18, 2023
☑	Explore Cont.(Handle Data type, Missing)	👤 Hoài An Nguyễn	December 18, 2023
☑	Explore Cont.(Numeric Distribution)	N Nam	December 18, 2023
☑	Explore Cont.(Non-numeric Distribution)	H Hường Thúy	December 18, 2023
☑	Ask Question (2)	K Kiên Hồ Sỹ	December 26, 2023
☑	Ask Question (1)	👤 Hoài An Nguyễn	December 26, 2023
☑	Ask Question (1)	N Nam	December 26, 2023
☑	Ask Question (1)	H Hường Thúy	December 26, 2023
☑	Model Training (Decision Tree)	K Kiên Hồ Sỹ	January 2, 2024
☑	Model Training (SVM)	H Hường Thúy	January 2, 2024
☑	Model Training (KNN)	N Nam	January 2, 2024
☑	Model Training (MLP)	👤 Hoài An Nguyễn	January 2, 2024
☑	Edit report slide	K Kiên Hồ Sỹ H Hường Thúy 👤 Hoài An Nguyễn N Nam	January 3, 2024

Phần 2: Tự đánh giá

STT	Họ tên	Khó khăn mắc phải	Bài học rút ra
1	Hồ Sỹ Kiên	Lần đầu sử dụng git nên còn nhiều bỡ ngỡ, chưa thực sự thành thạo.  Tự cảm thấy phần đặt câu hỏi và trả lời chưa thực sự sâu sắc.  Code còn hơi dài dòng và chưa thực sự tối ưu.	Giúp làm quen với git và cách làm việc nhóm trên git, Cách quản lý và phân chia công việc trên Notion.  Biết cách xây dựng một mô hình đơn giản thông qua các dữ liệu đã có.
2	Hoàng Thành Nam	Chưa có nhiều thời gian cho bài tập.  Gặp rắc rối trong việc xử lí các kiểu biểu đồ minh họa.	Cần chỉnh thời gian phù hợp. Học thêm về cách trực quan hóa bằng biểu đồ

3	Nguyễn Hoài An	Chưa nắm rõ về các thuật toán để xây dựng mô hình.  Code chưa thực sự tối ưu.	Học hỏi thêm về nhiều thuật toán.  Biết sử dụng thêm các hàm thư viện có sẵn để tối ưu code.
4	Đinh Thị Thúy Hường	Chưa nắm rõ về các mô hình.  Tự cảm thấy phần đặt câu hỏi và trả lời chưa thực sự sâu sắc, nếu có thời gian sẽ tìm hiểu về dữ liệu nhiều hơn để đưa ra một câu hỏi hay hơn.	Hiểu hơn về dữ liệu, thêm được nhiều kiến thức.

### **Nếu có nhiều thời gian hơn, nhóm em sẽ:**

- Tổ chức nhiều buổi họp hơn để cùng nhau thảo luận, tìm hiểu sâu hơn về thông tin dữ liệu mà nhóm đã thu thập. Từ đó nâng cao hiệu quả làm việc.
- Học cách tiền xử lý dữ liệu kĩ hơn để có thể loại bỏ các dữ liệu gây nhiễu hoặc các dữ liệu không có giá trị về mặt ý nghĩa gây ảnh hưởng đến mô hình học máy.
- Đưa ra các câu hỏi mang tính ý nghĩa, sâu sắc hơn nhằm khai thác triệt để hơn những thông tin mà dữ liệu mang lại.
- Tìm hiểu thêm về các mô hình học máy để tối ưu hơn hiệu quả của các mô hình.
- Hoàn thiện phân chia file notebook, tuân thủ clear-coding, viết code ngắn gọn và đơn giản nhất có thể, chú thích tường minh cho từng bước xử lý.
- Hoàn thiện hơn slide thuyết trình về mặt nội dung và thẩm mỹ.

Link github đồ án: [https://github.com/kienlqd1234/Intro2DS\\_FinalProject](https://github.com/kienlqd1234/Intro2DS_FinalProject)

Link Notion đồ án:

<https://www.notion.so/36707dd2fd9b4ae39e69e7a48786e359?v=efb5c1e8fddd44bd80fdad5533db2021&pvs=4>