

Đại học Quốc Gia Thành phố Hồ Chí Minh

Trường Đại học Khoa học Tự nhiên

Khoa Công nghệ thông tin

Bộ môn Khoa học máy tính

BÁO CÁO LÀM VIỆC NHÓM

Đồ án Nhập môn Khoa học Dữ liệu

GVHD: Trần Trung Kiên

Danh sách thành viên:

MSSV: 18120579

Đặng Minh Thọ

MSSV: 18120584

Phạm Đình Thực

Thành phố Hồ Chí Minh, năm 2020/2021

KẾ HOẠCH HỌP NHÓM CHUNG

Giai đoạn 1: Đưa ra câu hỏi cần trả lời

Giai đoạn 2: Thu thập dữ liệu

Giai đoạn 3: Khám phá dữ liệu

Giai đoạn 4: Tiền xử lý

Giai đoạn 5: Mô hình hoá dữ liệu

➔ Trả lời cho câu hỏi ban đầu

Giai đoạn 6: Truyền thông kết quả, đưa ra quyết định

KẾ HOẠCH HỌP NHÓM CHI TIẾT

I. GIAI ĐOẠN 1: Đưa ra câu hỏi cần trả lời

Ngày 28/12/2020

1. Mô tả công việc chính

Đưa ra câu hỏi cần trả lời

2. Chi tiết

Mỗi thành viên lựa chọn ra 1 câu hỏi để trả lời, sau đó thảo luận và thống nhất câu trả lời chung

- Đặng Minh Thọ: Dự đoán giá thuê phòng của một khách sạn
- Phạm Đình Thực: Dự đoán giá mua nhà

3. Kết quả

➔ Dự đoán giá mua nhà

➔ Trang web thu thập dữ liệu được lựa chọn: 'https://rever.vn'

Thành viên	Đã hoàn thành	Chưa hoàn thành	Tỉ lệ hoàn thành
Phạm Đình Thực	Thống nhất chủ đề		100%
Đặng Minh Thọ	Thống nhất chủ đề		100%

4. Kế hoạch buổi sau

Thu thập dữ liệu bước đầu

II. GIAI ĐOẠN 2: Thu thập dữ liệu

Ngày 29/12/2020

1. Mô tả công việc chính

Thu thập dữ liệu bước đầu

2. Chi tiết

Các thành viên tiến hành các bước để thu thập dữ liệu

- Đặng Minh Thọ: Thu thập dữ liệu của 1 đối tượng
- Phạm Đình Thục: Thu thập URL của tất cả các đối tượng

3. Kết quả

Thành viên	Đã hoàn thành	Chưa hoàn thành	Tỉ lệ hoàn thành
Phạm Đình Thục	Thu thập được URL của tất cả ngôi nhà	Chưa tối ưu requests.get, hàm chạy còn chậm	100%
Đặng Minh Thọ	Thu thập được dữ liệu của căn nhà	Chưa tối ưu, chạy còn chậm	100%

4. Kế hoạch buổi sau

Tiếp tục việc thu thập dữ liệu

III. GIAI ĐOẠN 3 : Khám phá dữ liệu

Ngày 4/1/2021

1. Mô tả công việc chính

Hoàn thành việc thu thập dữ liệu

Tiến hành Khám phá dữ liệu

2. Chi tiết

Hoàn tất thu thập dữ liệu và khám phá dữ liệu

- Đặng Minh Thọ: Hoàn tất thu thập dữ liệu (70%)
- Phạm Đình Thục: Tiến hành khám phá dữ liệu. Tiền xử lý Output (xử lý missing values, chuyển về dạng số, chuẩn hoá)

3. Kết quả

Thành viên	Đã hoàn thành	Chưa hoàn thành	Tỉ lệ hoàn thành
Phạm Đình Thục	Hoàn tất việc thu thập dữ liệu Khám phá và tiền xử lý output	Sau khi thu thập, chưa tối ưu việc đặt tên cho các thuộc tính (tiếng Việt có dấu hay không?)	100%
Đặng Minh Thọ	Hoàn tất việc thu thập dữ liệu	Chưa tối ưu, hàm thu thập chạy còn chậm	80%

4 . Kế hoạch buổi sau

Hoàn tất khám phá, chuyển qua tách tập dữ liệu và tiền xử lý.

IV. GIAI ĐOẠN 3 + 4: Khám phá dữ liệu + Tiền xử lý

Ngày 6/1/2021

1. Mô tả công việc chính

Tách tập dữ liệu, Tiền xử lý

2. Chi tiết

Tách tập dữ liệu từ tập dữ liệu thu thập thành 3 tập: train, validation và test. Tiền xử lý các cột số và cột chuỗi (Chuyển về dạng hợp lý)

- Đặng Minh Thọ: Viết decription, viết markdown, viết readme
- Phạm Đình Thực: Tiền xử lý train, validation, test

3. Kết quả

Thành viên	Đã hoàn thành	Chưa hoàn thành	Tỉ lệ hoàn thành
Phạm Đình Thực	Hoàn thành tiền xử lý	Một số cột còn chưa tối ưu	100%
Đặng Minh Thọ	Hoàn thành decription, markdown, readme	Một số mô tả còn chưa chính xác vì chưa đủ thông tin	100%

4 . Kế hoạch buổi sau

Hoàn thiện tiền xử lý, tiến hành mô hình hoá dữ liệu

V. GIAI ĐOẠN 5: Mô hình hoá dữ liệu

Ngày 11/1/2021

1. Mô tả công việc chính

Hoàn thành tiền xử lý dữ liệu, Mô hình hoá dữ liệu

2. Chi tiết

Tiến hành mô hình hoá dữ liệu theo neural network. Chọn ra giá trị tốt nhất cho siêu tham số. Xác định độ lỗi của kết quả sau quá trình huấn luyện.

Trả lời cho câu hỏi nhóm đã đặt ra

- Đặng Minh Thọ: Mô hình hoá, chỉnh sửa hàm thu thập
- Phạm Đình Thực: Mô hình hoá, đo độ lỗi của quá trình huấn luyện

3. Kết quả

Thành viên	Đã hoàn thành	Chưa hoàn thành	Tỉ lệ hoàn thành
Phạm Đình Thực	Hoàn thành mô hình hoá, xác định đúng độ lỗi	Độ lỗi còn cao, chưa tối ưu	100%
Đặng Minh Thọ	Hoàn thành mô hình hoá, xác định đúng độ lỗi	Độ lỗi còn cao, chưa tối ưu	100%

4 . Kế hoạch buổi sau

Tổng hợp báo cáo, slide các buổi họp. Hoàn thiện đồ án

Link GITHUB nhóm

<https://github.com/phamdinhtthuc020100/DataScienceProject.git>