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 Thục

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**: Phân tích 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**

1. **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

1. **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 Thục: Dự đoán giá mua nhà

1. **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 Thục | Thống nhất chủ đề | Không có | 100% |
| Đặng Minh Thọ | Thống nhất chủ đề | Không có | 100% |

1. **Kế hoạch buổi sau**

Thu thập dữ liệu bước đầu

1. **GIAI ĐOẠN 2**: Thu thập dữ liệu

**Ngày 29/12/2020**

1. **Mô tả công việc chính**

Thu thập dữ liệu bước đầu

1. **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

1. **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 |  |  |  |

1. **Kế hoạch buổi sau**

Tiếp tục việc thu thập dữ liệu

**Link GITHUB nhóm**

<https://github.com/phamdinhthuc020100/DataScienceProject.git>