**KẾ HOẠCH LÀM VIỆC NHÓM**

**Nhập Môn Khoa Học Dữ Liệu**

**2020 - 2021**

# 

# *Giáo viên hướng dẫn: Trần Trung Kiên-Phan Thị Phương Uyên*

# Group Information(Nhóm 16)

|  |  |  |
| --- | --- | --- |
| STT | MSSV | Họ và Tên |
| 1. | 18127199 | Lê Thanh Sơn |
| 2.. | 18127190 | Lê Hồng Quang |

# Bảng báo cáo:

|  |  |  |
| --- | --- | --- |
| Các mốc thời gian | 18127199-Thanh Sơn | 18127190-Hồng Quang |
| 16/08/2021->19/08/2021 | Tham khảo các trang web cung cấp API tại https://github.com/public-apis/public-apis | |
| Lên ý tưởng lấy API tại trang <https://www.coindesk.com/api/> | Không thể đặt câu hỏi có ý nghĩa thực tế cho dữ liệu này |
|  | Đề xuất lấy data từ <https://openweathermap.org/api/road-risk> |
|  | Kiểm tra thấy api chỉ được tiếp cận khi gửi email xin cấp quyền và xác nhận mục đích sử dung |
| Đề xuất láy dữ liệu thời tiết từ trang <https://weatherstack.com/> |  |
| Kiểm tra thì phát hiện nếu dùng gói miễn phí chỉ giới hạn 250 calls/ tháng muốn dữ liệu nhiều hơn phải trả phí |  |
|  | Đề xuất lấy data từ trang <https://openweathermap.org/api/air-pollution> |
| Kiểm tra và thấy có thể thu thập dữ liệu khá nhiều , hoàn toàn miễn phí chỉ bằng cách đăng ký tài khoản mà nhận api-key và có thể đặt được câu hỏi có ý nghãi trong thực tế. Quyết định sẽ xử lý trên tập dữ liệu từ API của trang web này. | |
| 20/8/2021 | Tiến hành quá trình thu thập dữ liệu từ web API và xuất ra file csv (lấy dữ liệu ở TPHCM) | Xác định các tham số để gọi các câu lệnh lấy dữ liệu từ web API |
| 21/8/2021->1/9/2021 | Làm Homework 3 | |
| 2/9/2021 |  | Khám phá dữ liệu và phát hiện tỉ lệ các lớp tại cột output chưa được phân bố đều |
| 3/9/2021 | Tiến hành thu thập thêm dữ liệu tại một số tỉnh/thành phố |  |
| 4/9/2021 |  | Đặt câu hỏi và đưa ra ý nghĩa thực tế |
| 5/9/2021 | Tìm hiểu một số thuật toán phân lớp từ slide bài giảng và trên doc của sklearn | |
| 6/9/2021 | Kiểm tra và chỉnh sửa | Mô hình hóa dữ liệu bằng mô hình neural network sử dụng thuật toán MLP Classifier |
| 7/9/2021 | Mô hình hóa dữ liệu bằng mô hình neural network sử dụng thuật toán Softmax Regression | Sửa lại danh sách siêu tham số alpha trong mô hình MLP Classifier do chọn alpha quá nhỏ dẫn đến độ lỗi lý tưởng (bằng 0%) |
|  | Kiểm tra và sửa lỗi ở hai mô hình Softmax và Kneighbors | Mô hình hóa dữ liệu bằng mô hình neural network sử dụng thuật toán Kneighbors Classifier |
| 8/9/2021 | Trực quan hóa dữ liệu | Nêu nhận xét và chọn mô hình tốt nhất |
|  | Làm mục Reflection trong slide FinalProject và tài liệu tham khảo | |
|  | Làm file thuyết trình đồ án | Làm slide (tóm tắt file notebook) và báo cáo công việc nhóm |
|  |  |  |

Ghi chú: Trong quá trình làm đồ án nhóm có sử dụng hai phương tiện là Facebook-Messenger và Github để trao đổi, liên lạc , bàn bạc và chỉnh sửa code

# Tham khảo

1. <https://openweathermap.org/api/air-pollution>
2. <https://scikit-learn.org/>
3. <https://stackoverflow.com/>
4. Video bài giảng + demo của môn học+ slide bài giảng
5. Homework:1,2,3