**Thành viên trong nhóm 12**

|  |  |
| --- | --- |
| Nguyễn Duy Phương | B1812294 |
| Nguyễn Quốc Hưng | B1812270 |

**Các thư viện cần dùng**

|  |  |  |
| --- | --- | --- |
| Tên thư viện | Mô tả | Lệnh cài đặt bằng CMD |
| Pickle | Thư viện lưu trữ dữ liệu, model,.. | pip install pickle-mixin |
| Pandas | Thư viện thao tác và phân tích dữ liệu | pip install pandas |
| Requests | Thư viện phép bạn gửi yêu cầu HTTP cực kỳ dễ dàng | pip install requests |
| Tqdm | Thư viện thông báo tiến độ quá trình | pip install tqdm |
| Beautifulsoup | Beautiful Soup là một thư viện giúp dễ dàng lấy thông tin từ các trang web | pip install beautifulsoup4 |
| Gensim | Thư viện xử lý ngôn ngữ tự nhiên | pip install gensim |
| Pyvi | Thư viện xử lý ngôn ngữ tự nhiên tiếng Việt | pip install pyvi |
| Sklearn | Thư viện máy học | pip install scikit-learn |
| Flask | Framework web app | pip install Flask |
| matplotlib | Thư viện biểu thị dữ liệu | pip install matplotlib |

**Cấu trúc các file trong thư mục**

|  |  |
| --- | --- |
| **TÊN FILE VÀ THƯ MỤC** | **MÔ TẢ** |
| **Baocao.docx** | File báo cáo viết bằng Word |
| **Baocao.pdf** | File báo cáo định dạng pdf |
| **Powerpoint.pptx** | File trình chiếu |
| ./FlaskApp/**main.py** | Code server web app flask chạy trên local |
| ./FlaskApp/**textProcess.py** | Code tiền xử lý văn bản |
| ./FlaskApp/**model.py** | Code xây dựng mô hình và đánh giá |
| ./FlaskApp/**Crawl.py** | Code cào dữ liệu |
| ./FlaskApp/**dulieu.pkl** | Dữ liệu đã được lưu lại sau khi cào dạng pickle |
| ./FlaskApp/**Bayes.pkl** | Mô hình đã lưu sau khi xây dựng |
| ./FlaskApp/**stopwords.txt** | Từ điển từ dừng |
| Các file khác trong thư mục FlaskApp là file cấu hình Flask không liên quan đến đề tài | |

**Chạy code**

**crawl.py:** Cào dữ liệu lưu lại thành file **dulieu.pkl**

**model.py:** training, xem kết quả đánh giá, lưu mô hình lại thành file **Bayes.pkl**

**main.py:** chạy server web local trên máy

Em đã up lên Heroku: [Phân lớp bình luận (cmtclassification.herokuapp.com)](https://cmtclassification.herokuapp.com/)