

## Đóng góp của các thành viên

| MSSV     | Họ và tên                     | Đóng góp   |
|----------|-------------------------------|--|
| 22280049 | Vũ Đăng Khôi<br>(Nhóm trưởng) | <p><b>Quá trình quản lý nhóm :</b></p> <ul style="list-style-type: none"> <li>- Sắp xếp các buổi họp</li> <li>- Ghi lại tiến trình công việc</li> <li>- Sắp xếp, phân chia công việc cho các thành viên</li> </ul> <p><b>Quá trình crawl data:</b></p> <ul style="list-style-type: none"> <li>- Thực hiện nghiên cứu tìm hiểu crawl data bằng selenium và beautiful soup</li> </ul> <p><b>Quá trình LLM:</b></p> <ul style="list-style-type: none"> <li>- Thực hiện và viết báo cáo về LLM</li> </ul> <p><b>Quá trình Data preprocessing và feature engineering</b></p> <ul style="list-style-type: none"> <li>- Chỉnh sửa file xử lý dữ liệu và EDA (làm phụ)</li> <li>- Thực hiện quá trình xử lý dữ liệu và EDA để so sánh với các thành viên khác</li> <li>- Tổng hợp code sau khi xử lý dữ liệu và EDA</li> </ul> <p><b>Model:</b></p> <ul style="list-style-type: none"> <li>- Áp dụng mô hình Regression tree, random forest</li> </ul> <p><b>Sản phẩm :</b></p> <ul style="list-style-type: none"> <li>- Cập nhật lại giao diện website</li> </ul> |
| 22280041 | Trảo An Huy                   | <p><b>Quá trình Crawl Data:</b></p> <ul style="list-style-type: none"> <li>- Đồng định hướng và viết báo cáo về phần crawl data.</li> <li>- Thu thập dữ liệu tọa độ</li> </ul> <p><b>Vẽ Heatmap:</b></p> <ul style="list-style-type: none"> <li>- Chịu trách nhiệm vẽ heatmap theo địa chỉ của bất động sản</li> </ul> <p><b>Quá trình ETL (Extract, Transform, Load)</b></p> <ul style="list-style-type: none"> <li>- Xây dựng, trình bày quá trình ETL</li> </ul> <p><b>Data Preprocessing và EDA</b></p> <ul style="list-style-type: none"> <li>- Thực hiện quá trình xử lý dữ liệu và EDA để so sánh với các thành viên khác</li> <li>- Lập bảng so sánh các phương pháp xử lý giá trị ngoại lai và giá trị bị thiếu</li> <li>- Đưa ra mục lục cơ bản.</li> </ul>  |

|          |               |  |
|----------|---------------|--|
|          |               | <b>Áp dụng mô hình</b> <ul style="list-style-type: none"> <li>- Xây dựng mô hình base line (hồi quy tuyến tính)</li> <li>- Áp dụng thử các phương pháp tối ưu quá trình.</li> <li>- Hồi Quy ridge.</li> <li>- Hồi quy lasso</li> <li>- Hồi quy eslastic.</li> </ul>  |
| 22280048 | Thái Anh Khoa | <b>Crawl Data :</b> <ul style="list-style-type: none"> <li>- Dùng Beautiful soup, trích xuất ra các thông tin cơ bản về giá nhà, phòng,... Sau đó dùng LLM để xử lý, truyền vào file csv</li> </ul> <b>LLM:</b> <ul style="list-style-type: none"> <li>- Code file LLM, chỉnh sửa và viết Prompt, viết Report cho phần LLM.</li> </ul> <b>Data preprocessing và EDA:</b> <ul style="list-style-type: none"> <li>- Code file xử lý cơ bản data và encoding các category, chỉnh sửa và bổ sung file Data Preprocessing and Feature Engineering 1 , file này sau đó đã được gộp chung file 2.</li> </ul> <b>EDA:</b> <ul style="list-style-type: none"> <li>- Vẽ biểu đồ phục vụ cho EDA, xử lý outlier.</li> </ul> <b>Model:</b> <ul style="list-style-type: none"> <li>- Mô hình K-Nearest Neighbors Regression</li> </ul> <b>Sản phẩm:</b> <ul style="list-style-type: none"> <li>- Dùng streamlit để thực hiện sản phẩm dự đoán giá nhà, sau đó dùng HuggingFace để tạo link truy cập.</li> </ul> |
| 22280045 | Đặng Lê Khiêm | <b>Quá trình Crawl Data</b> <ul style="list-style-type: none"> <li>- Định hướng trình thực hiện, tổ chức dữ liệu cho hợp lý trong quá trình Crawl Data và Trích xuất thông tin bằng LLM.</li> <li>- Thử nghiệm, so sánh, áp dụng Selenium với Request+Beautiful Soup 4 vào bài.</li> <li>- Chịu trách nhiệm chính trong quá trình viết code và report cho quá trình Crawl Data</li> <li>- Lưu trữ, quản lý các file data</li> </ul> <b>Trích xuất thông tin từ mô tả bài đăng bằng LLM</b> <ul style="list-style-type: none"> <li>- Hoàn thiện prompt. Đưa ra và chọn các feature cần trích xuất từ mô tả bài đăng.</li> </ul>   |

|  |  |   |
|--|--|---|
|  |  | <ul style="list-style-type: none"> <li>- Sửa lỗi nhỏ trong phần code sử dụng LLM để trích xuất dữ liệu. (Lỗi: chèn mã định danh article_id vào prompt)</li> </ul> <p><b>Quá trình ETL (Extract, Transform, Load)</b></p> <ul style="list-style-type: none"> <li>- Chịu trách nhiệm một phần code cho quá trình ETL</li> </ul> <p><b>Data Preprocessing và EDA</b></p> <ul style="list-style-type: none"> <li>- Thực hiện quá trình Data Preprocessing và EDA theo góc nhìn của cá nhân để so sánh với các thành viên còn lại.</li> <li>- Trình bày, bổ sung, chỉnh sửa toàn bộ file notebook quá trình Data Preprocessing và EDA.</li> <li>- Đưa ra phương hướng giải quyết cho các giá trị ngoại lai ở các cột.</li> </ul> <p><b>Áp dụng mô hình</b></p> <ul style="list-style-type: none"> <li>- Chịu trách nhiệm cho quá trình áp dụng mô hình Gradient Boosting vào bài.</li> <li>- Chỉnh sửa các pipeline của các mô hình.</li> <li>- Bổ sung, chỉnh sửa toàn bộ file so sánh giữa các mô hình</li> </ul> <p><b>Toàn bộ quá trình</b></p> <ul style="list-style-type: none"> <li>- Lưu trữ, quản lý toàn bộ các file.</li> </ul> |
|--|--|---|