

Sản phẩm đề ra cần đáp ứng 5 yếu tố : **Cụ thể,Đo lường được,Có thể đạt được,Liên quan,Giới hạn Thời gian**

Nội dung buổi họp 3/6:

- + Làm Project 1
- + Thống nhất được quy trình làm việc: mọi người cần phải hiểu toàn bộ code của nhau, 2 ngày họp 1 lần vào 21:00
- + Crawl dữ liệu từ giá nhà ở Hồ Chí Minh
- + Mục tiêu đề ra: Hoàn thành việc crawl HTML trên 1 web bao gồm về nội dung mô tả, và các mục khác trên web giá nhà
- + Sản phẩm đạt được vào ngày 5 /6 : dữ liệu HTML hoàn chỉnh

Nội dung buổi 4/6:

- + Mục tiêu đề ra : Hoàn thành việc crawl data , tìm hiểu về
- + Sản phẩm đạt được vào ngày 6/6 : tìm được LLM API tối ưu nhất

Nội dung buổi 6/6:

- + Lấy giá trị 1,0, Null. Mọi người cần tự giác xem chéo lại code của nhau, có gì không hiểu thì hỏi trên nhóm. Viết report bằng tiếng việt. Mọi người cần tự thống kê những việc đã làm
- + Mục tiêu đề ra:
 - Khiêm và Huy xem sơ xử lý model cho missing values,outlier, feature selection (deadline 7/6)
 - Khôi và Khoa dow xong data LLM (deadline 7/6)
 - Khôi tìm hiểu kĩ random forest cho feature selection, Khoa tìm hiểu kĩ KNN cho missing values (deadline 7/6)
 - Khôi và Khoa viết report và comment code cho LLM, Khiêm và Huy viết report và comment code cho crawl data (deadline 9/6)

Nội dung 8/6:

- + EDA (tiền EDA-> visualize) -> Data preprocessing(missing values outlier) -> Feature engineering (is the process of creating new features or transforming existing features to improve the performance of a machine-learning model. It involves selecting relevant information from raw data and transforming it into a format a model can easily understand. The goal is to improve model accuracy by providing more meaningful and relevant information.) -> Feature selection -> Data splitting

- + Nội dung buổi họp : thống nhất chia việc tất cả mọi người cùng làm.
T3 7h30 họp
- + Khiêm (gradient boosting), Khôi (random forest, decision tree), Khoa (KNN), Huy (linear regression, lasso, ridge)

Nội dung 13/6:

- + Bàn về quy trình baseline model
- + Fix code lỗi
- + feature engineering(...)
 - Missing values định lượng
 - Missing values định tính
 - Feature selection
 - Encoder
 - Transformer+ scaler+normalizer
 - Sự so sánh baseline model linear regression
- + Trình bày sơ lược những gì đã tìm hiểu trong 2 ngày qua

Nội dung 16/6

- + Nội dung buổi họp: Báo cáo tiến độ trong 3 ngày qua
- + Thống nhất sửa các lỗi cần thiết
- + Mục tiêu đề ra : mỗi người hoàn thành xong bài model

Nội dung 19/6:

- + Nội dung đề xuất các thông số để so sánh các model với nhau và với baseline .
- + Mục tiêu đề ra : mỗi người thực hiện so sánh và tìm hiểu thông tin để cho vào phần overview

Nội dung 20/6:

- + Nội dung buổi họp: Hợp lại nội dung phần kết luận