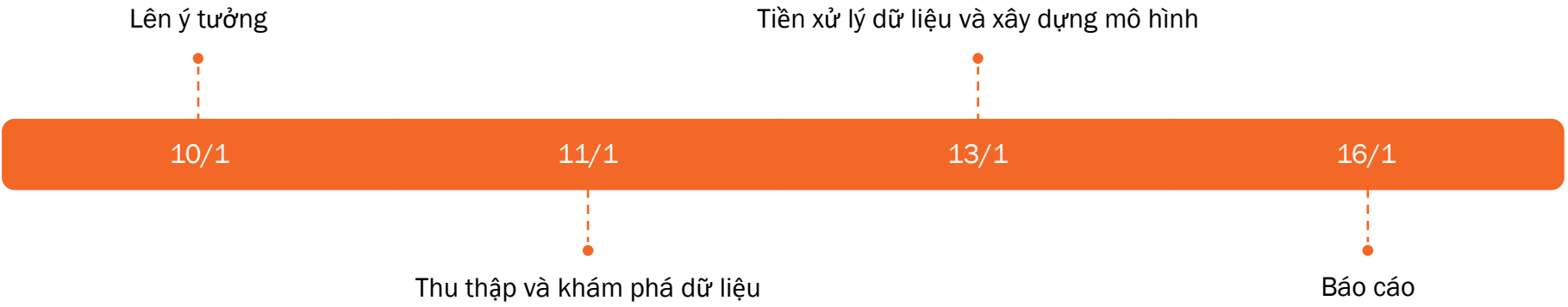


ĐỒ ÁN CUỐI KÌ

NHẬP MÔN KHOA HỌC DỮ LIỆU





Lên ý tưởng

- Dự đoán giá nhà
- Dự đoán thời tiết
- Dự đoán giá laptop

Lựa chọn cuối cùng: Xây dựng mô hình dự đoán giá của laptop thông qua cấu hình của máy

Thu thập và khám phá dữ liệu

-Dữ liệu máy tính bán trên Amazon

-Tìm được file CSV trên internet

-Tìm hiểu ý nghĩa của các cột

-Company: Tên công ty sản xuất

-Memory_Structure_Categories:

-Gpu_Type: Tên GPU

-Gpu_Company: Tên công ty sản xuất GPU

-OpSys: Hệ điều hành

-Cpu_Company: Tên công ty sản xuất chip

-Inches: Kích thước màn hình

-CPU_Frequency_Numeric: Xung nhịp

-Ram_Numeric: Dung lượng ram

-Size_Numeric: Dung lượng bộ nhớ trong

-Weight_Numeric: Trọng lượng laptop

-Price_dollars: Giá laptop

Tiền xử lý dữ liệu

- Kiểm tra kiểu dữ liệu của các cột
- Chuyển kiểu dữ liệu các cột về đúng dạng
- Kiểm tra các cột có bị thiếu giá trị hay không
- Tiền hành điền vào các ô bị thiếu (mean,mode,..)
- Tìm hiểu giá trị các cột được phân bố như thế nào
- Xử lý các dữ liệu dạng object bằng phương pháp Onehotencoder
- Tách các tập

Xây dựng mô hình

- Xử dụng mô hình Linear Regression

- Huấn luyện cho tập train

- So sánh với tập test và chấm điểm

Những điều nhóm chưa làm được

- Chưa crawl dữ liệu từ trên internet về, mà sử dụng file csv có sẵn
- Chưa kiểm tra độ lỗi của mô hình đã làm

Kết thúc

Cảm ơn thầy đã lắng nghe!