Đề tài 1: Dự báo thời tiết ngày mai

* Dataset: https://www.meteoblue.com
* Sử dụng thuật toán:
  + Hồi quy tuyến tính
  + Mạng noron nhân tạo
  + Thuật toán Bayers
* Input: Thông số thời tiết của các ngày hôm trước
  + nhiệt độ (°C)
  + độ ẩm (%)
  + lượng mưa (mm)
  + tốc độ gió (km/h)
  + hướng gió (độ)
  + lượng mây (%)
  + ………
* Output: Thông số thời tiết cảu ngày hôm sau
* Sử dụng các thư viện:
  + Tensorflow : hỗ trợ xây dựng deep learning
  + Pandas, Numpy: để phân tích và xử lý cấu trúc dữ liệu
  + Matplotlib: vẽ đồ thị
* Hạn chế: Nguồn dataset chỉ giới hạn trong 2 tuần trước đó muốn tải thêm phải mất tiền. (Cần tìm giải pháp)

Đề tài 2: Exploratory data analysis (EDA)

* Dataset: <https://www.kaggle.com/ashishgup/netflix-rotten-tomatoes-metacritic-imdb>
* Mục đích của đề tài:
  + Đặt ra và trả lời các câu hỏi xung quanh dữ liệu
  + Tìm ra cấu trúc của dữ liệu
  + Tìm ra các quy luật và những bất thường có trong dữ liệu
  + Kiểm tra và xác nhận các giả định về dữ liệu
* Các thư viện sử dụng: Pandas, Numpy, Matplotlib
* Phần mở rộng: Xây dựng thuật toán dự báo phim nào sẽ trở thành hot trend.
  + Input: Thông tin của một bộ phim
    - Thể loại
    - Tags
    - Đạo diễn
    - Writer
    - Diễn viên
    - …
  + Output:
    - Rotten Tomatoes Score
    - Metacritic Score
    - IMDb Votes
* Đây là phần mở rộng, có thể thay đổi một số thứ.

Đề tài 3: Dự báo giá xe đã sử dụng

* Dataset: http://www.laydulieu.com/
* Sử dụng thuật toán:
  + Hồi quy tuyến tính
* Input: Thông tin xe đã qua sử dụng:

Loại xe

Nhu Cầu

Vị trí

Giá

Màu xe

Xuất xứ

............

* Output: Đưa ra dự đoán về giá xe
* Sử dụng các thư viện:
  + Pandas, Numpy: để phân tích và xử lý cấu trúc dữ liệu
  + Matplotlib: vẽ đồ thị

https://ichi.pro/vi/du-doan-gia-o-to-da-qua-su-dung-voi-machine-learning-279985563742635?fbclid=IwAR3U7rdpmEHxVu7qYttVeJUP1YCWH0mkPIHRsm0DekiBSbZrA2eNKaIE\_JU

Đề tài 4: Dự báo số biến động của số chứng khoán

* Các thư viện:
  + Sclipy: mô tả các thuật toán về đại số tuyến tính, xử lý tín hiệu và xử lý ảnh
  + Pandas: Xử lí các bảng dữ liệu (cấu trúc dữ liệu, dataframe)
  + Scikit-learn: tiền xử lý dữ liệu, học giám sát và không giám sát, lựa chọn mô hình, validate và error metrics.