**PROJECT OUTLINE**

**1. Intro:**

- Mục tiêu: Dự đoán giá cổ phiếu hoặc phân tích xu hướng thị trường

- Tổng quan dataset: Dataset chứa giá cổ phiếu hàng ngày của các công ty công nghệ lớn từ 2014 đến 2023

- Link Dataset: <https://www.kaggle.com/datasets/umerhaddii/big-tech-giants-stock-price-data/data?select=big_tech_companies.csv>

**2. Data Preprocessing:**

- **Data Collection**: Tải dataset từ Kaggle

- **Data Inspection**: Kiểm tra dữ liệu

* Kiểm tra các giá trị thiếu
* Kiểm tra dữ liệu trùng lặp

- **Data Cleaning**: Làm sạch dữ liệu

* Loại bỏ hoặc điền các giá trị thiếu
* Xử lý các giá trị ngoại lệ

- **Data Transformation**: Chuyển đổi dữ liệu

* Chuyển đổi cột ‘Date’ thành định dạng datetime
* Tạo các biến mới nếu cần (ví dụ: chỉ số kỹ thuật như MA, RSI)

- **Data Normalization**: Chuẩn hóa dữ liệu

* Sử dụng các phương pháp như Min-Max Scaling hoặc Standard Scaling

**3. Exploratory Data Analysis - EDA:** Khám phá dữ liệu

- **Statistical Analysis**: Tính toán các thông số cơ bản (mean, median, standard deviation)

- **Data Visualization**: Sử dụng các biểu đồ (line charts, bar charts) để trực quan hóa dữ liệu

**=> Tập trung phân tích xu hướng thị trường => So sánh với thực tế**

**4. Data Splitting:**

- Splitting Data:

* Split into training and test sets (ví dụ: 80% training, 20% test)

**5. Model Building:**

- **Model Selection:**

* Linear Regression
* Mô hình chuỗi thời gian (ARIMA, LSTM)
* Random Forest (Sử dụng thuật toán)
* Gradient Boosting (Sử dụng thuật toán)

- **Model Training**: train models on the training set

**6. Model Evaluation:** Đánh giá mô hình

- **Performance Evaluation** (Đánh giá hiệu suất): Sử dụng các chỉ số như MAE, RMSE, MAPE

- **Model Comparison**: So sánh hiệu suất giữa các models

- **Result Visualization** (Trực quan hóa kết quả): Vẽ các biểu đồ so sánh giữa giá dự đoán và giá thực tế.

- **So sánh kết quả thực tế và kết quả dự đoán**

**Note:**

- All needed information: <https://vnexpress.net/quy-dinh-thoi-gian-giao-dich-chung-khoan-4312750.html>

- Tìm hiểu chỉ số ROA, ROE:

* Link: <https://vnexpress.net/roa-roe-la-gi-4304366.html>
* Link: <https://timo.vn/blogs/roa-roe-la-gi/>

→ Đối chiếu với thực tế để xem doanh nghiệp đó như thế nào?

- Lên trang Cafe F để xem giá cổ phiếu của một vài công ty

* Link: <https://cafef.vn/>

\*\*\*Khoảng thời gian trong dataset là từ 1/4/2010 đến 1/24/2023

=> Khoảng thời gian tương lai của data là 25/1/2023 đến hiện tại

- Phân tích giá đóng, mở => So sánh 2 trend để đánh giá model

\*\*\*Đánh giá model:

- 25/1 cho đến hiện tại là khoảng thời gian tương lai của data nhưng là quá khứ so với thực tế.

=> Nếu sai số ít thì có thể dự đoán đc khoảng thời gian trong thực tế.

## **BIG TECH GIANTS STOCK PRICE DATA**

### **1. Bối cảnh:**

- Dataset "Big Tech Giants Stock Price Data" chứa thông tin về giá cổ phiếu hàng ngày của các công ty công nghệ lớn như Apple, Microsoft, Amazon, Google, và Facebook từ năm 2014 đến 2023. Dữ liệu bao gồm các cột:

* **Date**: Ngày giao dịch
* **Open**: Giá mở cửa
* **High**: Giá cao nhất trong ngày
* **Low**: Giá thấp nhất trong ngày
* **Close**: Giá đóng cửa
* **Adj Close**: Giá đóng cửa điều chỉnh
* **Volume**: Khối lượng giao dịch

### **2. Output:**

- **Stock Price Prediction** (Dự báo giá cổ phiếu): Sử dụng phân tích chuỗi thời gian như ARIMA, LSTM để dự đoán giá cổ phiếu tương lai

- **Price Trend Classification** (Phân loại xu hướng giá):

* Xác định xu hướng và mô hình dài hạn của giá cổ phiếu
* Sử dụng các mô hình phân loại để dự đoán xu hướng giá sẽ tăng hay giảm

- **Anomaly Detection**: Phân tích và phát hiện các biến động bất thường trong giá cổ phiếu

- **Risk Assessment**: Đánh giá rủi ro

- **Phân tích biến động**: Đánh giá biến động giá cổ phiếu theo thời gian.

- **Phân tích tương quan**: Nghiên cứu mối tương quan giữa giá cổ phiếu của các công ty khác nhau.

### **3. Đặc điểm của Dataset:**

- **Đa dạng**: Bao gồm nhiều công ty công nghệ lớn với các đặc trưng khác nhau.

- **Chuỗi thời gian**: Dữ liệu có tính chất chuỗi thời gian, phù hợp cho các mô hình dự đoán theo thời gian.

- **Biến động**: Giá cổ phiếu có thể biến động mạnh do các yếu tố kinh tế, chính trị, và tin tức thị trường.

- **Tính tuần hoàn**: Dữ liệu có thể có tính tuần hoàn hoặc xu hướng theo mùa (như tăng giá vào các dịp lễ lớn).