DSE 6300 FINAL PROJECT PROPOSAL

PROJECT NAME: Stock Prediction Application

ABSTRACT:

Stock market is a public market where we can buy and sell shares for publicly listed companies. Stock market forecasting is a popular and important study in financial and academic fields. Many methods like technical analysis, fundamental analysis, time series analysis and statistical analysis are used to predict the price in stock markets. The aim of the project is the predict the future value of the financial stocks of a company Yahoo Finance Data (Webscraped), although stock market can never be predicted with hundred per-cent accuracy due to its vague domain, this paper aims at proving the efficiency of prediction models(LSTM- Long Short Term Model) for forecasting the stock market price. Python API and Docker will implement to create the application for stock market prediction. This application allows user to choose stock and its predictions values.
Key words: Prediction, LSTM, API.

OBJECTIVES

Aim of the project is to create an application to predict the future values for the user selected stock.

CURRENT METHODS AND RELATED WORK:

Data Collection and cleaning
Training a model and its accuracy
Creating an Python API
Deploy the application (API) using Docker/Container
Subject the application (API) to real time streaming applications like Kafka(Producer/Consumer).

APPROACH:

1. Stock data is pulled from Yahoo Finance and stored as SQL query.
2. Data is cleaned in SQL and exported to Python.
3. Prediction model is created in Python.
    a. Import the libraries
    b. Load the training dataset
    c. Training the model
    d. Normalizing the dataset
    e. Building the model
    f. Fitting the model
    g. Predicting the values

    h. Plotting the stock price
4. Create a Python API that contains the predicted values and historical data.
5. Deploy the application using Docker engine.
6. Finally the data will be published and subscribed using Kafka.

IMPACT:

The successful prediction of a stock market price could yield significant profit to the user/subscriber.

SCHEDULE:

Data collection and Query Creation by April 11 th 2022

Data Cleaning and Training model by April 15 th 2022

Creating an API and Docker container by April 19 th 2022

Running in Kafka by April 23 rd 2022

TEAM MEMBERS:

Karolina Konopka
Ahalya Pitchaikannu
Priyadharshini Pichai

DATA SOURCE:

Yahoo Finance data- Direct streaming data

LIST OF TOOLS AND TECHNOLOGIES THAT WILL BE USED:

TOOLS:

- MySQL
- Python
- Docker Engine
- Kafka
- Github
- Trello

TECHNOLOGIES:

- Tensorflow
- Keras
- Scikit Learn

Pandas
Numpy
Scipy
Zipline
Pynance
Statsmodel
Quandl
Matplotlib
Flask
JSON