Stock Price Prediction

COURSE: BIG DATA SYSTEM ENGINEERING USING SCALA (CSYE 7200)

PROFESSOR: Robin C. Hillyard

TEAM: Meenakshi Muthaih, Aswathnarayan Kirubakaran

Goals

- To determine the future value of a company stock or other financial instrument traded on an exchange

- Perform Time series analysis on Stock data using Scala and Spark

- Perform Twitter sentiment analysis for the provided stock

Data sources

- Huge stock market Dataset: https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs/version/2/home

- S&P 500 Stock Data: https://www.kaggle.com/camnugent/sandp500

- The dataset is taken is from Kaggle and has data for around 500 companies.

- Each data file has 8 columns and 2000 rows per file.

Use cases

- Provide the user with future stock price for each company

- Provide the sentiment analysis for individual stock which the user wants to invest

Methodology

- Perform Exploratory data analysis, Time series analysis on the Stock dataset.

- Perform Twitter sentiment analysis for the specified stock.

- Build a model using Spark MLlib to predict the stock price.

Acceptance Criteria

- Stock price prediction accuracy should provide accuracy above 50%

MileStones/Sprints

- Sprint 1: Perform data wrangling, integration and exploratory data analysis

- Sprint 2: Build machine learning models and perform training

- Sprint 3: Perform twitter streaming. Combine the work

- Sprint 4: Testing and Documentation

Code

- Everything in Scala
  - Spark
  - Spark Mllib
  - Spark Streaming

THANK YOU