

Project Proposal:

Trend Analysis of Stock Prices and Financial News

Group:

Akshita Chandiramani, Ruchita Jadhav, Sanjeev Ramasamy Seenivasagamani

Overview:

Stock market is an important part of the economy of a country. It plays an important role in the growth of the industry and commerce of the country that eventually affects the economy of the country to a great extent.

Fluctuations in stock prices are based on supply and demand of that stock. This affiliation is tangled with the type of news that is issued at any time. News regarding any firm influences the price of their stocks. Positive news creates a good impression regarding the firm and causes investors to buy stocks, thereby increasing the demand. This, in turn, increases the price of the stock. Negative news, on the other hand, such as a poor earnings report, political uncertainty, or a scandal will cause the investors to sell stocks which transforms into a decrease in stock price.

We aim to derive a quantifiable relationship between financial news and stock prices and predict stock prices of a firm based on analysis of financial news of that firm. This will help investors and firms in understanding the impact of a particular news on stock prices.

Goals and Problems:

Goal 1: Data acquisition and munging

Tasks:

Collect stock price information from Yahoo Finance website in the form of a csv.

Collect financial news from Reuters dataset in the form of json files

Since the required attributes are spread across different files, we need to integrate these into one consolidated dataset by matching stock ticker symbols with firm names

Expected results:

A cleaned up combined dataset that enables the next steps in data analysis

Expected problems:

Considering the difference in ticker symbols and company names, an efficient method to join data needs to be identified

Goal 2: Analyze Trend between news and stock prices over a 30 day period

Tasks:

Perform sentiment analysis of financial news for various firms using scikit-learn

Understand the trend of sentiment by using plotting features like Matplotlib

Compare stock prices to news sentiment to understand the relation between news and stocks

Expected results:

A correlation between news and stock prices to confirm that stocks are affected by the news

Expected problems:

Implementation of sentiment analysis might have drawbacks in understanding finance-related news. For example: “Apple is evaluated at \$1.2Bn” Is this positive or negative news? This might need some background information relating to Apple’s past performance and evaluation.

Sentiment for a firm may be analyzed as negative if the context it is used in is negative. For example: If there is news about a murder, but the article also mentions that a photo of the murderer was posted on Facebook, the news may be marked as negative for Facebook.

Goal 3: Modeling: To understand the relationship between news sentiment and stock prices

Tasks:

Developing statistical models for regression and evaluating the models’ performance based on various evaluation metrics like accuracy, RMSE

Expected results:

A model that predicts the variation of stock prices based on the sentiment of news about the firm

Expected problems:

The model may not have high accuracy since stock prices depend on external parameters apart from financial news, and the stock prices may have outliers which cannot be explained

Tools and Models:

Dataset: Yahoo Finance Stock Data, Reuter Financial News Data (~100 MB)

Tools: Apache Spark, Spark ML, Spark SQL, Jupyter Notebooks, Python 3 (Numpy, Pandas, Matplotlib, Seaborn), Hadoop, Scikit-Learn (Count Vectorizer, NGram), Dash by Plotly

Models: Linear regression with feature modeling, Topic Modelling on the news data