Learning From Network Project Proposal

Mohammad Pezeshki

Title: Analyzing Sentiment Influence in Social Networks and Its Impact on Financial Markets

Motivation:

Social media platforms, particularly Twitter, have become significant sources of sentiment and public opinion, often influencing financial markets. Public sentiment expressed online, especially by influential figures or large groups of users, can impact market trends, stock prices, and trading volumes. In this project, we aim to understand:

- How effectively can Twitter sentiment data improve stock price prediction when combined with traditional financial indicators?
- To what extent does sentiment in social networks correlate with stock price movements and market volatility?
- Which machine learning models best capture the influence of social sentiment on stock prices, and do certain models handle this integration more accurately?
- Under what market conditions is social sentiment data most predictive?
- Does including specific types of sentiment, such as fear, optimism, or confidence, improve the model's accuracy in predicting price movements?

Data:

- Social Network Data: Twitter data is being gathered using the **Tweepy** library, a Python-based library that interfaces with the Twitter API. This allows for an efficient and scalable collection of tweet content, including hashtags, user posts, and engagement metrics, while adhering to Twitter's data access policies.
- Financial Market Data: We will obtain historical stock price data, trading volumes, and volatility indices (such as the VIX) from sources like Yahoo Finance or Alpha Vantage focusing on major indices like the S&P 500 or key stocks that are sensitive to public sentiment.

Total Data Available:

- **Twitter Data:**
  - #Tesla Hashtag Dataset: Initially 465,721 tweets; after preprocessing, it reduced to 9,402 tweets.
  - Elon Musk's Tweets Dataset: Initially, there were 6,251 tweets; after preprocessing, it was reduced to 4,663 tweets.
- **Stock Price Data:** Approximately 2.75 years of daily Tesla stock prices, including adjusted closing prices and moving averages (7, 20, and 50 days).

Method:

- **Problem**: This research aims to determine if Twitter sentiment scores can effectively predict stock prices and assess which machine learning model provides the best predictive accuracy.
- **Algorithms:** Six machine learning algorithms are employed for the predictions: Logistic Regression (LR), Decision Trees (DT), Random Forest Regressor (RF), Support Vector Regressor (SVR), Extreme Gradient Boosting (XG Boost), and Facebook Prophet (FBS).
- **Approach:** Sentiment scores derived from tweets and stock price data, including moving averages, are combined to train and test each algorithm for predicting stock price movements.

Intended Experiments:

- **Implementation**: We will utilize the implementations of these machine learning algorithms available in the Scikit-Learn and Facebook Prophet libraries and use the implementation available in this article ("https://doi.org/10.47852/bonviewJCBAR42022006")
- **Experiments**: Each algorithm will be run on two datasets—Tesla-related tweets and Elon Musk's tweets—both with and without the inclusion of moving averages. Models will be evaluated based on mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) to compare prediction accuracy, runtime, and overall model performance. The Random Forest and XG Boost models are hypothesized to perform best based on preliminary results.
- **machine for experiments:** Acer Aspire 5 with 7-core CPU and 16 GB RAM DD4 and Google Colab

Reference:

1. Impact of Social Media Sentiments on Stock Market Behavior: A Machine Learning Approach to Analyzing Market, Theeshanthani Kandasamy and Kamal Bechkoum Dynamics
2. Power of 280: Measuring the Impact of Elon Musk's Tweets on the Stock Market, Sanjeev Metta*, NidheeshMadhavan†& KrishnamoorthyKrishnamoorthy Narayanan
3. Elon Musk's Twitter and Its Correlation with Tesla's Stock Market, Daniel Pyeong Kang Kim, Jongwhee Lee, Jungwoo Lee, Jeanne Suh