PREDICTING STOCK MARKET IMPACT OF SOCIAL MEDIA TRENDS

MUHAMMAD HAMZA

IIT KHARAGPUR

Project Report:

Predicting Stock Market Impact Of Social Media Trends

1. INTRODUCTION

The project aims to analyse social media trends, specifically Twitter data, and predict their potential impact on stock market movements. The goal is to develop a machine learning model that can help investors make more informed decisions by incorporating sentiment analysis of social media discussions related to specific stocks.

2. METHODOLOGY

2.1 Data Collection

Twitter Data:

Twitter API is used to fetch tweets related to specific stock symbols. Tweets are collected in real-time to capture the latest sentiments.

Stock Price Data: Yahoo Finance API (yfinance) is utilized to retrieve historical stock price data for the chosen stock symbols.

2.2 Data Preprocessing

Text Cleaning:

Preprocessing of Twitter data involves removing URLs, special characters, and other noise to ensure accurate sentiment analysis.

Date Alignment:

Twitter data and stock price data are aligned based on date timestamps for correlation analysis.

2.3 Sentiment Analysis

VADER Sentiment Analysis:

The VADER (Valence Aware Dictionary and sentiment Reasoner) sentiment analysis tool is used to analyse the sentiment of tweets. VADER provides a compound sentiment score that indicates the overall sentiment of each tweet.

2.4 Correlation Analysis Sentiment vs. Stock Price: The sentiment scores derived from Twitter data are correlated with the daily stock price changes of the respective stocks. Statistical Analysis: Statistical techniques are applied to assess the strength and significance of the correlation between sentiment scores and stock price movements.

3. RESULTS

3.1 Data Collection and Preprocessing

Twitter Data**:**

Tweets related to stock symbols (e.g., AAPL for Apple Inc.) are collected using the Twitter API, with error handling in place for API rate limits. Stock Price Data: Historical stock price data for selected symbols are retrieved from Yahoo Finance, ensuring data consistency and accuracy.

Data Preprocessing: Text cleaning techniques are applied to remove noise and ensure the quality of sentiment analysis.

3.2 Sentiment Analysis

VADER Sentiment Scores:

Each tweet is assigned a sentiment score using VADER, providing insights into the overall sentiment (positive, negative, neutral) of Twitter discussions related to the stocks.Sentiment Trends: Visualizations such as sentiment distribution plots and time-series sentiment trends help in understanding sentiment patterns.

3.3 Correlation Analysis

Sentiment-Stock Price Correlation:

Statistical analysis (e.g., Pearson correlation coefficient) is performed to measure the correlation between sentiment scores and daily stock price changes.

Impact Assessment:

The strength and direction of the sentiment-stock price relationship are analyzed to assess the potential impact of social media trends on stock market movements.

4. DISCUSSION

4.1 Insights from Analysis

Sentiment Trends**:**

Identification of overall sentiment trends (positive/negative/neutral) in social media discussions related to stocks.

Impact Analysis:

Assessment of how sentiment shifts in social media correlate with stock price movements, providing insights into investor sentiment and market reactions.

4.2 Limitations and Future Work Data Limitations: Limited scope of Twitter data and potential biases in sentiment analysis.

Model Enhancements:

Exploration of advanced sentiment analysis models and incorporation of additional data sources (e.g., news sentiment, financial reports) for improved predictions.

Real-Time Monitoring:

Implementation of a real-time monitoring system to capture dynamic changes in social media sentiment and stock prices.

5. CONCLUSION

The project demonstrates the potential of using social media sentiment analysis as a supplementary tool for predicting stock market trends. By integrating Twitter data analysis with stock price movements, investors can gain valuable insights into market sentiment and make more informed investment decisions. Further enhancements and research avenues exist to refine the predictive accuracy and usability of the model. This project report provides a structured overview of the project's objectives, methodology, results, discussions, and conclusions, highlighting key findings and areas for future development.