TASK- B.7 REPORT (EXTENSION REPORT)

Enhancing Stock Price Prediction through Sentiment Analysis of Social Media Data

1. Introduction

The following report discusses a number of ways in which stock price prediction techniques might be improved by incorporating sentiment analysis of social media, in this case, Twitter. Traditional models of stock price predictions are limited to using historical price data as an input. This often makes them less accurate and not reliable. We are going to try and improve the prediction accuracy by adding other factors like social sentiment, giving a wider view of the trend in markets.

2. Summary of Research

2.1. Approaches for Stock Price/Trend Prediction
Following are some of the ways researched to improve stock price prediction:

Fundamental Analysis: Based on the financial health of a company, its management, industrial conditions, etc.
Technical Analysis: The study that is used to forecast future movements based on past price and volume.
Machine Learning Models: These are models based on algorithms like linear regression analysis, decision trees, and neural networks that study patterns in stock price fluctuation.
Sentiment Analysis: The public sentiment of social media platforms, news articles, and financial reports are analysed to get a view of market reaction and trends.

2.2. Approach Used: Sentiment Analysis
Sentiment analysis is a process that determines the emotional character of text data. From these tweets about specific stocks, one can extract sentiment scores on what the public speaks about a company or its stock and eventually use this information along with historical stock price data to make better prediction models.

Advantages in Using Sentiment Analysis:

Real-time Data: Social media platforms give views in real time, thus availing timely information about the public opinions.
Market Sentiment Indicator: Public sentiment generally goes along with market movements; thus, giving quite good insight into making any prediction.

3. Implementation Summary
3.1. Data Collection
We shall implement the sentiment analysis by collecting through Python's Tweepy library tweets about certain stocks. This shall be carried out in the following manner:

Logging into the Twitter API with the correct credentials
Fetching tweets of stock-related keywords

3.2. Sentiment Analysis

Sentiment Analysis: The scraped tweets would have to go through either TextBlob or VADER sentiment analysis. Each tweet will be given a sentiment score ranging from -1 (negative sentiment) to 1 (positive sentiment).

3.3. Integration

The derived sentiment scores will be averaged every day and integrated with the historical stock price data. Activities:

Determining the average daily sentiment for each stock.
The sentiment data will then be joined to the historical stock prices based on some common date.
Now that we have integrated sentiment scores as features in our dataset, we can go ahead to train a machine learning model for the prediction of stock prices. Some of the models possible are:

Linear Regression
Decision Trees
LSTM Networks
The performance of the best model will be evaluated by metrics such as Mean Squared Error, which can tell how well the model is able to predict future stock prices given both historical data and sentiment.

4. Evaluation

After the sentiment analysis approach is applied and integrated with historical stock data, model performance will be evaluated by comparing the MSE of the model trained with the sentiment features against a baseline model using only historical price data. We will analyze whether the sentiment helps in giving better accuracy to the predictions; this will give us a fine tune on the effectiveness of the extension done.

5. Conclusion

Sentiment analysis of social media data in integrated stock price prediction models holds great promise in developing far more accurate predictions. This model leverages real-time public sentiment, capturing market trends more robustly than could a model based on historical data alone.

6. Future Work

Other research data sources include news articles and financial reports for sentiment analysis.
Evaluation of different machine learning models to work out the best approach.
The results can also be further improved by doing a greater analysis of some events or trends in the world that may influence the stock price. This will be very important at creating better predictions.

7. References

Twitter API Documentation
TextBlob Documentation
VADER Sentiment Analysis