**Name:** Talia Qaiser

**AndrewID:** Tqaiser

# **Programming Assignment: Stock Price Prediction and GameStop Short Squeeze**

**Models and Approach:**

My approach to the assignment was first doing a sentiment analysis, I did it using vader sentiment and then used the Naïve Bayes Classifier as the model to predict the sentiments. After that I compiled an LSTM model for prediction using the yfinance stock price data for 2021 and predicted the price for June, July and August. Then I combined and merged the datasets, the vader sentiment with the close and open prices and ran an LSTM model on that.

**Sentiment Analysis:**

Data Used: rsuperstonk\_dataset\_features.csv

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TUMIPC>

Ran a sentiment analysis and built a Naïve Bayes Classifier model to predict the sentiment. With a 73% accuracy.

A screenshot of a computer screen

Description automatically generated

**LSTM Model:**

The LSTM model using yfinance but no sentiment score.

A graph showing the price of a stock

Description automatically generated

Analysis: The model captures patterns but does not predict the values properly.

**Combining the Model:**

Added the vader sentiment to the yfinance dataset and ran an LSTM model.

A graph of a line graph

Description automatically generated with medium confidence

Analysis: This model does not capture the patterns that well.

A black background with white text

Description automatically generated

The MSE is very high but the RMSE and MAE are not too high.

**Limitations:**

The model only uses the Close prices and the Vader Sentiment from reddit, that is why it is performing worse than the previous one. The model can be improved if more features are added to it like Volume of stock, Open prices, sentiments from other social media sources. The timeframe of the data used is also just limited to 2021, feeding the model more data will also improve the predictions.

**Sensitivity:**

To check the sensitivity of the model to spikes, I conducted a sensitivity analysis with a factor of 5.

A graph showing the price of a stock

Description automatically generated

A graph of blue and orange lines

Description automatically generated

The model is not very sensitive to the changes in data points for the vader sentiment. So, if there is an extreme amount of change in the social media posts about the company, only then we will see huge changes in the stock prices.

**Event Analysis:**

The spikes can occur due to the following reasons:

1. High Social Media Activity: this happened in January 2021 for GameStop
2. New and Media Coverage
3. Market Dynamics
4. Influencer opinions

We can keep a track of these spikes and identify them by:

1. Continuous monitoring of trends and trend words for social media.
2. Do a sentiment analysis periodically

**Algorithmic Changes Proposed (Online and ChatGPT suggestions):**

1. **Temporal Analysis:** Implement temporal analysis in your model. Consider time-dependent features to capture the temporal dynamics of sentiment spikes and their impact on stock prices.
2. **Adaptive Sentiment Weights:** Introduce adaptive sentiment weights. Assign different weights to sentiment features based on their historical significance and relevance to recent spikes.
3. **Incorporate Lagged Sentiment:** Include lagged sentiment features to account for potential delays in the market's response to sentiment changes. This can help capture the persistence of sentiment effects.
4. **Ensemble Models:** Explore ensemble models that combine the strengths of multiple algorithms. Combining sentiment analysis with traditional financial indicators and market data may enhance the model's robustness.

**Summary:**

In this assignment, a model was developed to predict stock prices using social media sentiment, particularly during the GameStop short squeeze in June – August 2021. The model demonstrated accuracy in capturing general sentiment trends, and then captured the prediction along well but when combined exhibited limitations in handling the sentiment and prediction together. It also is limited in sensitivity of the model as it only shows a limited impact in extreme cases. Sensitivity analysis revealed the model's challenges in responding to extreme events, suggesting the need for improvements.

**Discussion:**

The whole assignment underscored the limitations of traditional forecasting models in adapting to rapidly evolving market dynamics. Traditional models struggled to anticipate and respond to the influence of social media sentiment, especially during volatile events like short squeezes. Conversely, incorporating social media sentiment data can prove valuable in enhancing predictive power and providing real-time insights, offering a more nuanced understanding of market behavior it did not do that for this model that I built but if given more time and more data the model could have performed better.

**Proposal (ChatGPT):**

Future research directions should focus on improving stock price prediction models that integrate social media sentiment and addressing challenges. Potential areas for exploration include:

1. **Behavioral Economics Integration:** Incorporate principles from behavioral economics to better understand and model how social sentiment influences investor behavior.
2. **Sentiment Dynamics Modeling:** Develop models that explicitly capture sentiment dynamics, accounting for factors such as sentiment momentum and sudden shifts.
3. **Cross-Domain Sentiment Analysis:** Explore the application of sentiment analysis across multiple domains, including social media, financial news, and expert opinions, for a more comprehensive view of market sentiment.
4. **Robustness Testing:** Conduct robustness testing on models to assess their performance in extreme scenarios, improving resilience in the face of unexpected events.
5. **Larger Dataset:** Use a larger dataset to train the model so that it is more accurate and can learn better.

By addressing these research directions, future models can become more accurate, and robust, providing a deeper understanding of the interplay between social media sentiment and stock prices in dynamic financial markets.

**Appendix:**

For code I used ChatGPT, the code given in the recitation.

Other links used:

<https://www.analyticsvidhya.com/blog/2022/01/sentiment-analysis-with-lstm/>

<https://github.com/JordiCorbilla/stock-prediction-deep-neural-learning/blob/master/stock_prediction_lstm.ipynb>

Other instances where I got help from ChatGPT have been quoted in the report.