



# Sentiment Analysis for Meme Stock Price Prediction

Zihan Qiang, Jingyu Shen, Yu Dong Zhang

## Abstract

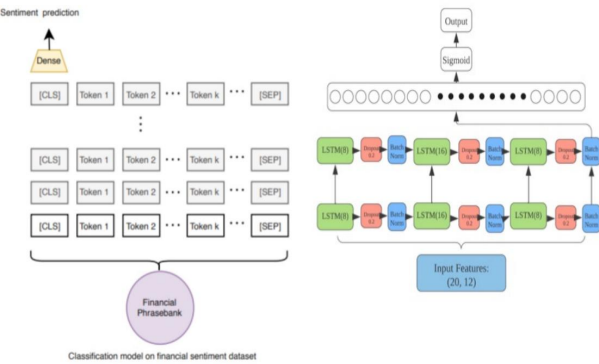
Market sentiment is one of the most significant driving forces in the financial market, and such sentiment is mostly displayed via social media platforms. However, discussions and comments from social media are often expressed in a very informal setting.

Here we present an NLP model called MemeBERT, built by adjusting configurations of FinBERT and further fine-tuning using Reddit and Twitter text data to capture sentiment from highly informal text.

Then, a LSTM neural network is later developed to generate trading decisions based on the outputted sentiment along with technical indicators. The generated strategies have reached higher return and less draw-down on the hold-out test set.

## Approach

We developed the model MemeBERT, that is a further pre-trained and fine-tuned FinBERT model (Left) on labelled Twitter and Reddit data, that are characterized by its abundance of internet slangs. Next, we developed an LSTM model to predict stock price (Right). We used grid search to find the optimal set of hyperparameters.



## Data

1. Labelled Twitter data with 1.6 million examples, each text is labelled with positive, negative and neutral. This dataset is used to train MemeBERT
2. This dataset includes sentiment score and 11 technical indicators, with the label of 0 (decrease) or 1 (increase), for 4 meme assets (i.e. GME, AMC, BTC, ETH).

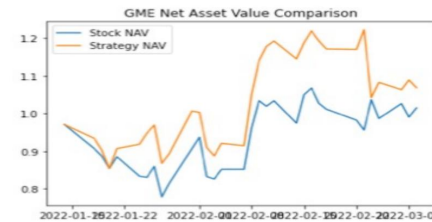
## Results and Discussions

1. MemeBERT on twitter data:  
Loss: 0.4; Accuracy: 73%; Precision: 0.79; Recall Score: 0.74; F1 score: 0.78.

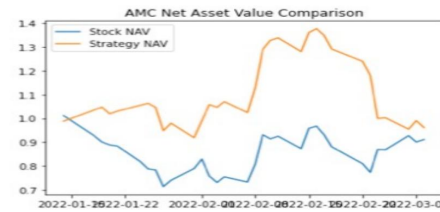
FinBERT on twitter data:  
Loss: 0.81; Accuracy: 63%; Precision: 0.63; Recall Score: 0.6; F1 score: 0.64.

MemeBERT is significantly better than FinBERT on twitter data, which shows that our MemeBERT model could understand internet slangs and achieved satisfied result.

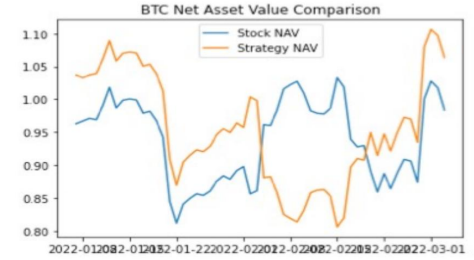
2. Price Movement Prediction:



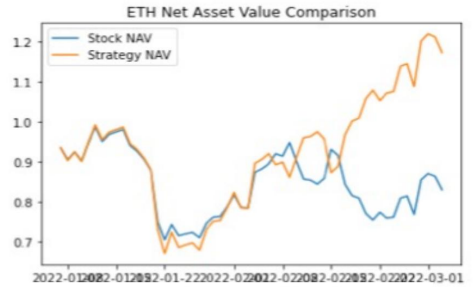
	Accuracy	NAV	SR
Train	57.88%	Strategy 1.07	0.04
Test	52.56%	Stock 1.02	0.14



	Accuracy	NAV	SR
Train	62.96%	Strategy 0.96	0.12
Test	54.22%	Stock 0.91	0.07



	Accuracy	NAV	SR
Train	62.14%	Strategy 1.06	0.031
Test	53.57%	Stock 0.98	0.01



	Accuracy	NAV	SR
Train	63.53%	Strategy 1.17	0.068
Test	54.98%	Stock 0.83	-0.051

## Conclusion

Our changes to the baseline model show significant improvements. MemeBERT has accuracy of 73%, whereas FinBERT has accuracy of 63%. The LSTM neural network, with the use of sentiment score from MemeBERT and technical indicators, has achieved higher net asset value across the 4 assets and also higher Sharpe ratio. The trading strategy is effective.