

# Forecasting S&P 500 Index Closing Price using LSTM and fin BERT

Jibin Rajan Varghese,  
Dept. of Computer Science  
Texas A&M University,  
College Station, Texas, USA

Divya Susan Thomas  
Dept. Of Computer Science  
University of Texas at Austin.  
Austin, Texas USA

**Abstract-** Stock market prediction is a complex task that is invaluable to economic activity across the globe. We are currently at the end of a decade-long bull run where the stock market experienced a relatively steady rise and lower volatility. As we enter an era with more stock price fluctuations, a good price predictive model is important, especially for participants in inverse and leveraged Exchange Traded Funds (ETFs) to reduce risks and make gains in the market. Since there was no study that used LSTM trained on both structured and unstructured data to predict the Standard & Poor's 500 index, this study focused on training single and multi-layer LSTM models to predict the Standard & Poor's 500 index closing price utilizing historical, macroeconomic and technical data as well as using sentiment analysis of financial news obtained using FinBERT. The goal of this study was to find the architecture of the LSTM model with the highest accuracy in forecasting tomorrow's closing price of the Standard & Poor's 500 index and investigate if the performance of the LSTM model could be improved by adding unstructured textual information such as sentiment extracted from financial news using FinBERT, which is a pre-trained NLP model that has been shown to outperform all current state-of-the-art models in analysing sentiment of financial text.

**Keywords –** S&P 500 index, LSTM, finBERT, Sentiment Analysis, Time series forecasting, Machine Learning

## I. INTRODUCTION

Stock market prediction is a complex task that is invaluable to economic activity across the globe. We are currently at the end of a decade-long bull run where the stock market experienced a relatively steady rise and lower volatility. As we enter an era with more stock price fluctuations, a good price predictive model is important, especially for participants in inverse and leveraged Exchange Traded Funds (ETFs) to reduce risks and make gains in the market.

Since there was no study that used LSTM trained on both structured and unstructured data to predict the Standard & Poor's 500 index, this study focused on training single and multi-layer LSTM models to predict the Standard & Poor's 500 index closing price utilizing historical, macroeconomic and technical data as well as using sentiment analysis of financial news obtained using FinBERT. The goal of this study was to find the architecture of the LSTM model with the highest accuracy in forecasting tomorrow's closing price of the Standard & Poor's 500 index and investigate if the performance of the LSTM model could be improved by adding unstructured textual information such as sentiment extracted from financial news using FinBERT, which is a pre-trained NLP model that has been shown to outperform all current

state-of-the-art models in analysing sentiment of financial text [1].

## II. RESEARCH BACKGROUND

The noisy and non-parametric behaviour of the non-linear and deterministic chaotic systems that embody equity markets [2] makes it challenging to predict their behaviour. Moreover, the optimal set of variables selected for price prediction varies greatly. There are different types of studies in the literature that use either technical indicators, historical data, or a combination of both[3][4][5][6]. Feature selection is of great importance as the performance, interpretability, and robustness of the model depend on this. The predictive power of the model may be limited if too few features are selected whereas the performance may be worse if too many are selected due to over-fitting to noise.

Bhandari et al. implemented an Long Short Term Memory (LSTM) model with high predictive accuracy to predict the closing price of the Standard & Poor's 500 index using only historical, macroeconomic and technical data [7]. Other studies which used only unstructured textual data to predict stock prices also found promising results as discussed in more detail in the following paragraphs. Prior studies have found LSTM models to be quite effective for stock market prediction. Chen et al. used LSTM to predict

Chinese stock returns using historical data and found the model to be promising [8]. Similarly, in 2017, Roondiwala et al. used LSTMs to predict closing prices of NIFTY 50 using only the open, close, low, and high prices of the index from 2011 to 2016 with RMSE value of 0.0086 [9]. LSTM networks were also found to outperform other memory-free classification methods like random forest approaches, deep neural networks (DNN), and logistic regression-based classifiers in predicting movements of Standard & Poor's 500 stocks during the time period from 1992 to 2015[10]. The use of the LSTM model trained on price history along with technical indicators to forecast Standard & Poor's 500 prices one, five, and ten minutes prior to high-frequency trading also has shown promising results [11]. Furthermore, on comparing the performance of LSTM, SVM, Kalman filters, Backpropagation, etc. in predicting stock price, LSTM was shown to have the best performance for prediction accuracy with the least variance [12].

Similarly, a study investigating the relative performance of the LSTM, Support Vector Regressor, and Autoregressive Integrated Moving Average (ARIMA) in predicting the stock price of the Standard & Poor's 500, Nikkei 225, Hang Seng Index, China Securities Index 300, DJIA and ChiNext index found that LSTM performed better than other models [13]. Finally, previous studies that have implemented single-layer and multi-layer LSTM models to predict stock market index closing prices have found the single-layer LSTM to be superior in performance to other approaches [7][14]. Kalyani et al. found encouraging results from using only qualitative inputs such as news articles about Apple Inc. and predicting its future stock prices using sentiment-based classification [15]. Gite et al. (2021) implemented a similar LSTM network to predict the closing price of the NSE (Indian stock market index) and found it to be a better predictive model than LSTM based on just historical price data[16]. Similarly, predictive models built on both stock historical stock prices and financial news showed high accuracy in predicting the index prices in both Vietnam and China [17][18]. The sentiment analysis model based on BERT was found to perform better than TextCNN, TextRNN, Att-BLSTM and TextCRNN [19]. Macroeconomic factors such as the interest rate, exchange rate and commodities index prices (e.g., gold index) were also found to be highly significant in projecting stock returns [20].

### III. METHODS

#### 1. Variable Definitions

Fundamental data collected contained the close price which is the dependent variable in this study. The close price is the price at which the security last trades when the exchange closes on a trading day. The other variables used in this study include the US dollar index (USDIX), Cboe Volatility Index (VIX), Civilian Unemployment Rate (UNRATE), Effective Federal Funds Rate (EFFR),

and the University of Michigan: Consumer Sentiment (UMC-SENT). The US dollar index (USDIX) is a metric that measures the value of the US dollar relative to six major foreign currencies, namely, the Euro, Swiss franc, Japanese yen, Canadian dollar, British pound, and Swedish krona. The index indicates the strength of the US dollar in the global markets and generally, there exists a positive correlation between the index and stock prices [21]. Cboe Volatility Index (VIX) measures the 30-day forward projection of volatility of the Standard & Poor's 500 index (SPX). It provides a measure of the market sentiment, especially fears among investors. Studies have generally shown a strong inverse relationship between VIX and the price of the Standard & Poor's 500 index [2]. Civilian Unemployment Rate (UNRATE) is a metric that represents the number of unemployed individuals who are 16 years or older, residing in one of the 50 US states, as a percentage of the US labour force. Studies have shown it to be a strong predictor of stock prices [23][24]. The monthly data was then converted to daily data using the forward-filling method.

Effective Federal Funds Rate (EFFR) is a volume-weighted median of overnight federal funds transactions reported in the FR 2420 Report of Selected Money Market Rates. It has been shown to affect inflation, growth, and unemployment and generally, has an inverse relationship with stock prices [25]. Daily EFFR and the monthly University of Michigan: Consumer Sentiment (UMCSENT) data for the 10 years was collected from the FRED website [26]. UMCSENT is a monthly survey conducted by the University of Michigan that measures consumer sentiment and expectations of the economy and future spending. It has been shown to influence stock prices [27].

Technical indicators used in this study are the three most popular indicators used by traders, namely, Moving Average Convergence Divergence (MACD), Relative Strength (RSI), and Average True Range (ATR). Daily values for these over the 10 years were computed by the historical stock market data gathered from Yahoo Finance. MACD, calculated by subtracting the 26-day exponential moving average from the 12-day exponential moving average is used by traders to determine the direction and momentum of a stock price. ATR measures the volatility of a stock over a period, typically either 14, 20, or 22 days. For the purposes of our research, 14 days was chosen as the period since it is the most used. RSI measures the momentum of a stock and is used to identify bearish and bullish price signals. All three of these indicators have been shown to have the ability to predict stock prices over shorter periods [28][29].

#### 2. S&P 500 Close Price Denoising

As can be seen in Fig.1., the closing price of the Standard & Poor's 500 shows complex, noisy behaviour. Studies have shown discrete wavelength transformations and Haar

wavelengths to be highly effective in denoising stock price data [30]. Therefore, the closing price was denoised by using the Haar wavelength with soft thresholding and the VisuShrink approach.



Fig.1. Movement of S&P 500 closing price between 2010-2020

### 3.Sentiment Analysis on financial news headlines

As mentioned in section II, using the Kaggle financial news headlines dataset, daily positive, neutral, and negative sentiment scores were obtained for the ten years from 2010-2020. First, only headlines pertaining to the largest 35 companies in the Standard & Poor’s 500 index were extracted into a data frame. Then Auto Tokenizer was called with the `from_pretrained()` method to return the correct tokenizer class instance for the finBERT model and then the pre-trained finBERT model was fetched from Hugging Face. Each headline in the data frame was then tokenized and fed to the finBERT model. The output logits were passed through a softmax layer to get a tensor with 3 values representing the positivity, neutrality, and negativity of each headline. The scores for all the

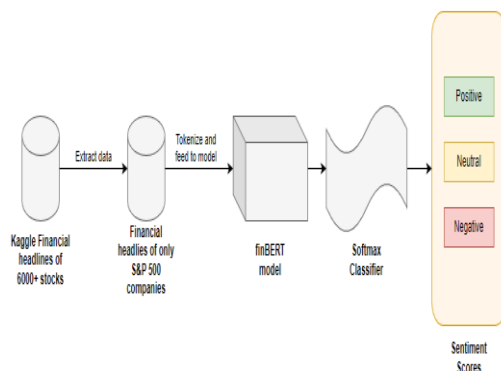


Fig.2. Workflow for extracting sentiment scores

headlines are then aggregated by date and the average is found to give a positive, negative, and neutral average score for each day. These average sentiments are taken to be the daily Standard & Poor’s 500 positive, neutral, and negative sentiment scores. These were then joined to the

structured data using the date as a key. This process is depicted in Fig.2.

### 3.Data Normalization

The range and standard deviation of the input variables also varied greatly from each other. For example, as shown in Fig.3., the Close price ranges from 250 to 3000 while sentiment score variables range from 0 to 1. Therefore, there was a need to perform feature scaling to ensure that the machine learning algorithms performed well. Min-max normalization technique using `MinMaxScaler` from `sci-kit learn` was used to scale features to values between 0 and 1.

## III. EXPERIMENTS

### 1.Data Sources

This study utilized both structured and unstructured data to predict Standard & Poor’s 500 index closing price. Structured data collected included fundamental, macroeconomic, and technical data that were identified to have an impact on the stock price after reviewing the literature. Unstructured data consisted of financial news headlines from Benzinga. The data was collected from January 2010 to June 2020 i.e., a 10-year duration, which included the Covid-19 pandemic. In total, 12 variables were collected. These are summarized in Table 1. Daily closing prices of the Standard & Poor’s 500 were collected from Yahoo Finance for 10 years except at weekends and holidays when the market was closed. Index data used in this study was the daily Adjusted Close price (ticker symbol `DX-Y.NYB`) collected from Yahoo Finance for 10 years [31]. In this study, the historical daily adjusted Close price for the 10 years from 2010-2020 obtained from Yahoo Finance was used [32]. Additionally, the monthly Civilian Unemployment Rate for the 10 years collected from FRED [26] was also used in this study.

Unstructured data used in this study was the ‘Daily Financial news for 6000+ Stocks’ dataset found on Kaggle which contains around 4 million financial article headlines from the 2009-2020 period scrapped from Benzinga. This massive data set contained headlines pertaining to around 6000 stocks, many of which did not belong to the Standard & Poor’s 500 index. Nearly 50% of the Standard & Poor’s 500 index’s price is determined by the largest 35 companies. Therefore, in this study, only headlines related to these 35 companies were extracted to make the dataset more manageable. Sentiment analysis was conducted on all the daily headlines between 2010 and 2020 pertaining to these companies from the dataset to obtain positive, negative and neutral Standard & Poor’s 500 news sentiment scores for every day for ten years.

Two datasets were prepared for the study. Dataset 2 consisted only of the fundamental, macroeconomic and technical indicators. Dataset 1 contained in addition to these, the sentiment scores of headlines. This has been

done so we can investigate if incorporating unstructured textual information such as sentiment extracted from financial news and adding the textual data can improve the performance of the LSTM network. This information is summarized in Table.1.

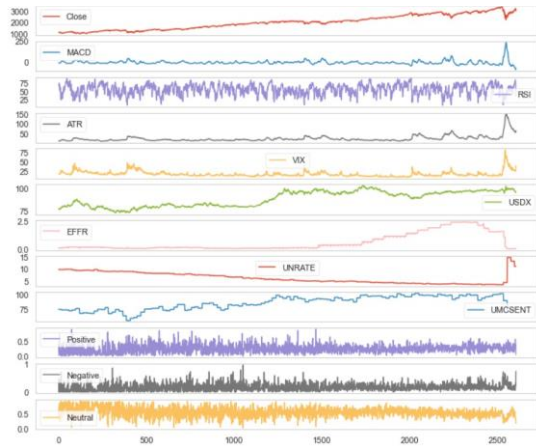


Fig.3. Snapshot of input variables showing their varying ranges.

Table.1.Datasets and Variables used in this study

Dataset	Content	Variables
1	Fundamental data, Macroeconomic data, Technical indicators, Sentiment score from headlines	Close price, VIX, EFRR, UNRATE, UMCSENT, USDX, MACD, ATR, RSI, Positive sentiment score, Neutral sentiment score, Negative sentiment score
2	Fundamental data, Macroeconomic data, Technical indicators	Close price, VIX, EFRR, UNRATE, UMCSENT, USDX, MACD, ATR, RSI.

## 2. Data Split

After this, the dataset was a two-dimensional array with several observations and features, but LSTM requires input data in three dimensions (number of input features, number of observations, and size of time step). Therefore, the data was reshaped to meet this. The size of the time step used in this study is 5. The data was split into training and testing sets using an 80/20 split ratio. The order of the time series data was preserved during the split. The first

80% of the data ranging from 2010 till 2018 was used for training and the next 20% of the data from 2018 to 2020 was kept as the testing set. The training set was used to train the LSTM models while the testing set was used to evaluate the performance of the models. The last 20% of the data points in the training set after splitting were used for validation during the hyperparameter tuning process and once tuning was completed, this data was included back into the training set.

## 3.Hyperparameter Tuning

To find the best model architecture, several variations were tested. Both single-layer LSTM and multi-layer LSTM models were fitted to the data. For single-layer LSTM networks, 7 models with 2,10,30,50,100,150 and 200 neurons respectively were tested to find the best model. For the multi-layer LSTM architecture, 6 different model variations were tested. These multi-layer configurations were (8,4), (10,5), (10,5,2), (50,20), (150,100), and (100,50,20) where L1, L2, and L3 in (L1, L2, [L3]) represent the number of neurons in the first, second and third hidden layer respectively. For each model, hyperparameter tuning was performed to find the optimal values for each hyperparameter. The hyperparameters tuned in this study were batch size, learning rate, and type of optimizer. The optimizers tested were Adam, Adagrad and Nadam. Learning rates tested were 0.1,0.01 and 0.001 and the batch sizes tested were 8,16 and 32.

Hyperparameter tuning was performed for each combination of hyperparameters and each model, with 50 epochs, a time step of size 5, and repeated 10 times to account for the stochastic nature of LSTM models. Early Stopping call-back with validation loss as the performance metric and patience (number of epochs with no improvement after which training will be stopped) of 5 was also used. Average Root Mean Square Error (RMSE) values were calculated on the validation data and the combination that resulted in the lowest RMSE value was the optimal hyperparameter setting for that model. Table 2 lists the hyperparameter settings used for tuning and Tables 3-5 list the optimized hyperparameter setting found after tuning for every LSTM model in this study. These were the hyperparameter settings on which the optimized models were trained and evaluated on the test data.

Table.2.Overview of hyperparameter settings for tuning.

Hyperparameter tuned	Settings Tested
Batch Size	8, 16, 32
Learning Rate	0.001, 0.01, 0.1
Optimizer	Adam, Adagrad, Nadam

Table.3.Values of optimized hyperparameters for single layer LSTM models on Dataset 1.

Number of Neurons	Batch Size	Learning Rate	Optimizer
2	8	0.1	Adam
10	8	0.001	Adam
30	16	0.1	Adagrad
50	16	0.1	Adagrad
100	32	0.1	Adagrad
150	32	0.01	Adagrad
200	32	0.01	Adagrad

#### 4.Training and testing with optimized hyperparameters

Once all the hyperparameters were optimized, these models with the optimized hyperparameter setting shown in Tables 3-5 were fitted on the complete training data and their performance was evaluated on the test data. Performance metrics used in this study were Mean Absolute Percentage Error (MAPE), Pearson correlation coefficient (R), and Average Root Mean Square Error (RMSE), which were computed as shown in the equations below where  $y_i$  is the actual value at the time  $i$ ,  $\hat{y}_i$  is the predicted value at time  $t$ ,  $\bar{y}_i$  is the mean of the actual time series,  $\bar{\hat{y}}_i$  is the mean of the forecasted time series and  $N$  is the number of observations. Smaller values of RMSE and MAPE indicate smaller errors and hence better model performance. In contrast, the larger value of R is better since that indicates a better linear correlation between the actual and predicted data points.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

$$MAPE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N-1} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$R = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 (\hat{y}_i - \bar{\hat{y}})^2}}$$

Each of the models with the optimized hyperparameters was trained using 100 epochs, a time step of size 5, and repeated 30 times to account for the stochastic nature of LSTM models. EarlyStopping with loss performance metric and patience of 5 is also employed. Both single-layer and multi-layer models were trained using the same values for these parameters. The average of RMSE, MAPE and R values from the 30 replications were computed.

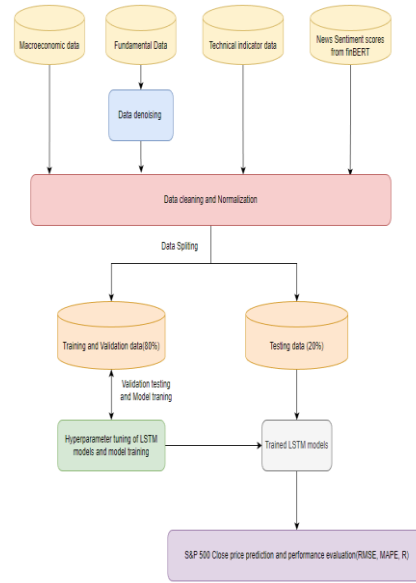


Fig.4. Flowchart of proposed methodology.

Table.4.Values of optimized hyperparameters for multi-layer LSTM models on Dataset 1.

Multi-layer model neurons	Batch Size	Learning Rate	Optimizer
[8, 4]	16	0.01	Adam
[10, 5]	32	0.01	Adam
[10, 5, 2]	4	0.01	Adam
[50, 20]	4	0.01	Adagrad
[150, 100]	16	0.01	Adagrad
[100, 50, 20]	8	0.001	Adagrad

Table.5.Values of optimized hyperparameters for single layer LSTM models on Dataset 2.

Number of Neurons	Batch Size	Learning Rate	Optimizer
2	8	0.1	Adam
10	16	0.01	Adam
30	16	0.1	Adagrad
50	8	0.1	Adagrad
100	32	0.1	Adagrad
150	32	0.01	Adagrad
200	32	0.01	Adagrad

The average RMSE value was chosen as the main selection criteria. So, the model with the smallest average RMSE, followed by the smallest MAPE value and largest R was the best model. To decide if adding sentiment data improves the LSTM model, the models were trained on 2 datasets – As explained in the prior section, Dataset 1 included the sentiment score data while Dataset 2 did not include this. Average RMSE scores of the models trained on both datasets were used to decide if sentiment scores improved model performance.

A flowchart of the methodology proposed in this study is depicted in Fig.4.

## V. EXPERIMENTAL ANALYSIS



Fig.5.Results for the 7 different single layer LSTM models on Dataset 1 over 30 replications'

Using the optimized hyperparameters, the 7 single-layer LSTM models, namely with 2,10,30,50,100,150 and 200 neurons were tested to find the best model for dataset 1 which is the complete dataset that contained both structured data (Macroeconomic, technical, and fundamental data), and unstructured data (sentiment scores from headlines). The model with 10 neurons had the lowest average RMSE value of 52.07, MAPE value of 1.26, and the largest R-value of 0.97. Hence it was found to be the best model. The line plot in Fig.5 shows the variation in Average RMSE, MAPE, and R for the 7 different models. The box plots show the variation in RMSE, MAPE, and R over 30 replications for the seven single-layer LSTM networks tested in this study. In the model with 10 neurons had seen in Fig.5, MAPE and RMSE(average) increase as the neuron count drops below 10 and increase as the number of increases above 10. This might be because a larger number of neurons result in over-fitting while a lower number of neurons result in under-fitting.

### 1. Single Layer LSTM models' results on Dataset 1

The original closing price data and the forecasted closing price values obtained from the best single-layer model with 10 neurons which had the lowest average RMSE value are depicted for the ten years of the study in Fig.6. Between 2010 to June 2018, the green curve represents predictions in the training data whereas, between June 2018- June 2020, it shows predictions in the test data. It can be seen from that the forecasted values closely overlap with the actual values for the most part except for the 2020 area where more variation between the curves is visible. This deviation could be due to unusual market conditions during the Covid-19 pandemic. However, even during the unexpected period of 2020 in the test data, the model does capture the suddendrop and the subsequent V-shaped sharp recovery of the market. This is indicative of the

strong forecasting ability of the model and suggests that it is not over-fitted.

### 2. Multi-Layer LSTM models' results on Dataset 1

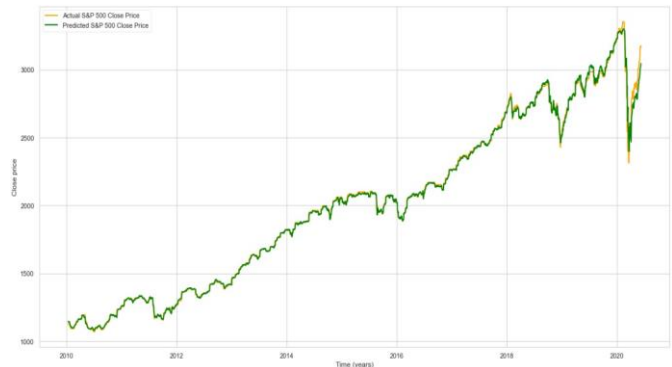


Fig.6.Actual vs forecasted Standard & Poor's 500 closing price using best multi-layer LSTM on dataset 1 (10-5-2 model)

Even though single-layer LSTM gave pretty good results, several multi-layer LSTM models were also tested to see if any of them would result in improving the accuracy of predicting the closing price of the Standard & Poor's 500 index. Using the optimized hyperparameters, the 6 multi-layer LSTM models were tested to find the best model for dataset 1. These were models with hidden layers (8,4), (10,5), (10,5,2), (50,20), (150,100), and (100,50,20) where the numbers represent the number of neurons in each hidden layer.

Multi-layer LSTM model with 3 hidden layers having 10, 5, and 2 neurons respectively had the lowest average RMSE value of 55.93, MAPE value of 1.44, and R-value of 0.96. Hence it was found to be the best model. The line plot in Fig.7 shows the variation in average RMSE, MAPE, and R for the 6 different models. The box plots show the variation in RMSE, MAPE, and R over 30 replications for each of the 6 multi-layer LSTM models tested in this study. The original closing price data and the forecasted closing price values obtained from the best multi-layer model with 3 hidden layers having 10, 5 and 2 neurons respectively and which had the lowest average RMSE value is depicted for the ten years period of the study in Fig.8. Between 2010 to June 2018, the green curve represents predictions in the training data whereas, between June 2018- June 2020, it shows predictions in the test data. The forecasted values closely overlap with the actual values for the most part except for the 2020 area where more variation between the curves is visible. However, the overlap is not as close as in the case of the best single-layer model, especially around 2020 when unusual market conditions created by the covid-19 pandemic arose. This is to be expected since the best single-layer model has a lower RMSE value compared to the multi-layer model. However, even during the

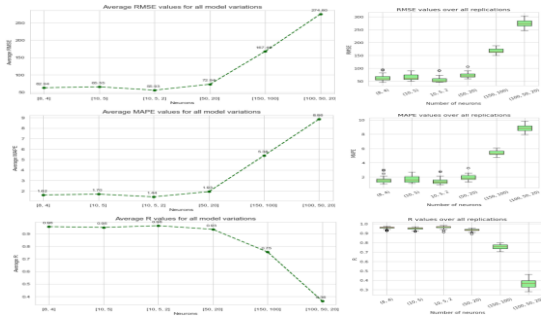


Fig.7. Results for the 6 different multi-layer LSTM models on dataset-1 over 30 replications.

unexpected time of 2020 in the test data, the model does capture the sudden drop and the subsequent V-shaped sharp recovery of the market. This is indicative of the strong forecasting ability of the model and suggests that it is not over-fitted. Therefore, even the best multi-layer model appears to have good predictive power. However, since the single-layer LSTM model is simpler and has a lower average RMSE value, it is a better choice than the multi-layer LSTM model since multiple layers do not seem to improve the performance of the network but rather reduce it.

### 3. Comparing Single Layer LSTM on Dataset 1 and 2

Though the results showed that the single-layer LSTM model was better than the multi-layer model in this scenario, further exploration was needed to find out if adding the sentiment scores extracted from financial headlines improves the LSTM model. To do this, the single-layer LSTM models were trained on Dataset 2 which did not contain the sentiment data. Using the optimized hyperparameters for Dataset 2, the 7 single-layer LSTM models, namely with 2, 10, 30, 50, 100, 150 and 200 neurons were tested to find the best model. The model with 50 neurons had the lowest average RMSE value of 49.75, MAPE value of 1.21 and the largest R-value of 0.97. Hence it was found to be the best model. The line plot in Fig.5 shows the variation in Average RMSE, MAPE and R for the 7 different models and the boxplots show the variation in RMSE, MAPE and R over 30 replications for each of the 7 single-layer LSTM models tested in this study.

As can be seen from Fig.5 and Fig.9, the MAPE and average RMSE increase as the number of neurons increases above 50 neurons but when the number of neurons decreases, the error initially increases but interestingly drop for the model with 10 neurons before increasing again. The original closing price data and the forecasted closing price values were derived from the best network with 50 neurons which had the lowest average RMSE value is depicted for the ten years period of the

study in Figure 10. Between 2010 to June 2018, the green curve represents predictions in the training data whereas, between June 2018- June 2020, it shows predictions in the test data. The forecasted values closely overlap with the actual values for the most part except for the 2020 area where more variation between the curves is visible. The overlap is slightly even better than that found for the 10-neuron single-layer LSTM model trained on dataset 1 which included the sentiment scores. This is to be expected since this best single-layer model with 50 neurons had a lower RMSE value compared to the 10-neuron model trained on dataset 1. However, there are still deviations from the actual value even in this model though the model does capture the sudden drop and the subsequent V-shaped sharp recovery of the market just like other previously mentioned models and is indicative of the strong forecasting ability of the model and suggests that it is not over-fitted. Based on the results shown above, it appears that the addition of financial news headlines sentiment scores does not seem to improve the predictability of the single-layer LSTM model.

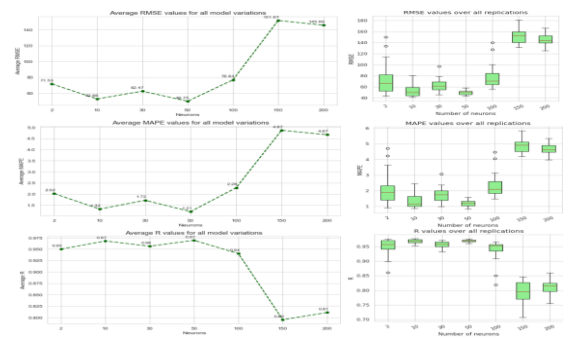


Fig.8. Results for the 7 different single layer LSTM models on dataset2 over 30 replications.

When sentiment scores are included, the best model, an LSTM model with a 10-neuron single layer, has an average RMSE of 52.07 whereas when sentiment scores are not included the best model is an LSTM with a 50-neuron single layer which gives an average RMSE of 49.7 over 30 replications. This might suggest that since the RMSE values are similar, it might be better to include sentiment scores since then the LSTM model architecture needs to have only 10 neurons. However, since the 10-neuron single-layer model trained on dataset 2 (without sentiment scores) also has an average RMSE score of 52.6, it suggests that even if the goal is to have a simpler LSTM model, adding the sentiment scores from the financial headlines does not seem to help improve the model performance. This is perhaps because including the daily financial headlines might be adding more noise to the dataset than any actual extra information that is not already contained in the macroeconomic and technical

data variables included in the study which includes customer sentiment (UMCSENT).



Fig.9. Actual vs forecasted Standard & Poor's 500 closing price using best single-layer LSTM on dataset 2 (50 neuron model)

## VI. CONCLUSION

The ability to predict stock prices accurately has always been coveted by many traders around the world. Prior studies have shown LSTM models, in particular single-layer LSTM models trained on macroeconomic, technical fundamental data to be very effective at predicting the close price of the Standard & Poor's 500 index. Studies using sentiment score data from financial news and other unstructured data to predict stock prices have also shown promise.

This study investigated if incorporating daily averaged sentiment scores of news headlines pertaining to companies in the Standard & Poor's 500 index along with the macroeconomic, technical, and fundamental data would help improve the predictive power of LSTM models. First, the study used finBERT, a state-of-art BERT model pretrained on financial text model to extract sentiment scores from financial headlines which were then added to the dataset. A variety of single-layer and multi-layer LSTM models were then implemented and extensively tuned to find the optimal hyperparameter settings. These models were then fitted on both the complete dataset and the dataset with the sentiment scores removed. Their performance was then evaluated using Average RMSE, MAPE and R scores. Results showed that the single-layer LSTM model performed better than multi-layer LSTM models.

This validates previous studies that have also shown the same result. The single-layer LSTM network with 10 neurons was found to be the best fit and offered the best prediction accuracy. The model was able to forecast rapid fluctuations of the closing price of the Standard & Poor's 500 index even in highly unusual market environments during the COVID-19 pandemic. This is a very promising outcome and suggests that traders seeking to predict the markets might find it beneficial to give LSTM models a closer look. Results also showed that adding the sentiment scores did not improve the model performance but rather

decreased it slightly. This might be because the daily financial headlines sentiment scores might be adding noise but no new information to the dataset since it already contained technical and macroeconomic data including consumer sentiment.

## VII. FUTURE WORK

Many tasks can be explored in future studies. The financial headlines analysed in this study only came from Benzinga. Even though this is a popular news source among traders, headlines collected from only one news source might have biases. To avoid this, headlines can be collected from a variety of popular news sources. Similarly, sentiment from other unstructured data such as social media sentiment, analyst reports and annual reports can also be incorporated. Furthermore, hybrid models that combine LSTM with other deep learning models can also be developed and tested to compare their performance. Finally, to test the robustness and adaptability of the proposed models, they could be used to predict market indices in other countries that are like the United States in terms of economic and consumer behaviours.

## REFERENCES

- [1] Z. Liu, D. Huang, K. Huang, Z. Li, J. Zhao, Finbert: A pre-trained financial language representation model for financial text mining, in: Proceedings of the twenty-ninth international conference on artificial intelligence, 2021, pp. 4513–4519.
- [2] R. G. Ahangar, M. Yahyazadehfar, H. Pournaghshband, The comparison of methods artificial neural network with linear regression using specific variables for prediction stock price in tehran stock exchange, arXiv preprint arXiv:1003.1457 (2010).
- [3] L. Di Persio, O. Honchar, Artificial neural networks architectures for stock price prediction: Comparisons and applications, International journal of circuits, systems and signal processing 10 (2016) (2016) 403–413.
- [4] J. Wang, J. Kim, Predicting stock price trend using macd optimized by historical volatility, Mathematical Problems in Engineering 2018 (2018) 1–12.
- [5] M. Qiu, Y. Song, Predicting the direction of stock market index movement using an optimized artificial neural network model, PloS one 11 (5) (2016) e0155133.
- [6] S. Wu, Y. Liu, Z. Zou, T.-H. Weng, S<sub>i</sub>-lstm: stock price prediction based on multiple data sources and sentiment analysis, Connection Science 34 (1) (2022) 44–62.
- [7] H. N. Bhandari, B. Rimal, N. R. Pokhrel, R. Rimal, K. R. Dahal, R. K. Khatri, Predicting stock market index

- using lstm, Machine Learning with Applications 9 (2022) 100320.
- [8] K. Chen, Y. Zhou, F. Dai, A lstm-based method for stock returns prediction: A case study of china stock market, in: 2015 IEEE international conference on big data (big data), IEEE, 2015, pp. 2823–2824.21
- [9] M. Roondiwala, H. Patel, S. Varma, Predicting stock prices using lstm, International Journal of Science and Research (IJSR) 6 (4) (2017) 1754–1756.
- [10] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European journal of operational research 270 (2) (2018) 654–669.
- [11] Z. Lanbouri, S. Achhab, Stock market prediction on high frequency data using long-short term memory, Procedia Computer Science 175(2020) 603–608.
- [12] D. Karmiani, R. Kazi, A. Nambisan, A. Shah, V. Kamble, Comparison of predictive algorithms: backpropagation, svm, lstm and kalman filter for stock market, in: 2019 Amity International Conference on Artificial Intelligence (AICAI), IEEE, 2019, pp. 228–234.
- [13] P. Yu, X. Yan, Stock price prediction based on deep neural networks, Neural Computing and Applications 32 (2020) 1609–1628.
- [14] A. Yadav, C. Jha, A. Sharan, Optimizing LSTM for time series prediction in Indian stock market, Procedia Computer Science 167 (2020) 2091–2100.
- [15] J. Kalyani, P. Bharathi, P. Jyothi, et al., Stock trend prediction using news sentiment analysis, arXiv preprint arXiv:1607.01958 (2016).
- [16] S. Gite, H. Khatavkar, K. Kotecha, S. Srivastava, P. Maheshwari, N. Pandey, Explainable stock prices prediction from financial news articles using sentiment analysis, PeerJ Computer Science 7 (2021) e340.
- [17] D. Duong, T. Nguyen, M. Dang, Stock market prediction using financial news articles on Ho Chi Minh stock exchange, in: Proceedings of the 10<sup>th</sup> International Conference on Ubiquitous Information Management and Communication, 2016, pp. 1–6.
- [18] C.-R. Ko, H.-T. Chang, Lstm-based sentiment analysis for stock price forecast, PeerJ Computer Science 7 (2021) e408.
- [19] M. Li, L. Chen, J. Zhao, Q. Li, Sentiment analysis of Chinese stock reviews based on BERT model, Applied Intelligence 51 (2021) 5016–5024.
- [20] A. Jabeen, M. Yasir, Y. Ansari, S. Yasmin, J. Moon, S. Rho, An empirical study of macroeconomic factors and stock returns in the context of economic uncertainty news sentiment using machine learning, Complexity 2022 (2022).
- [21] Novianti, Analysis on the influence of selected macroeconomic indicators (consumer price index, trade balance, non-farm payroll, housing starts, and s&p 500) towards us index (period 2010-2015), Ph.D. thesis, President University (2016).
- [22] L. Ruan, Research on sustainable development of the stock market based on vix index, Sustainability 10 (11) (2018) 4113.
- [23] W.F. Pan, Does the stock market really cause unemployment? a cross-country analysis, The North American Journal of Economics and Finance 44 (2018) 34–43.
- [24] F. Farsio, S. Fazel, The stock market/unemployment relationship in USA, China and Japan, International Journal of Economics and Finance 5 (3)(2013) 24–29.
- [25] B. S. Bernanke, K. N. Kuttner, What explains the stock market's reaction to federal reserve policy?, The Journal of finance 60 (3) (2005) 1221–1257.
- [26] Unemployment rate (unrate) , FRED, St. Louis Fed, <https://fred.stlouisfed.org/series/UNRATE>, accessed on 06/19/2023
- [27] K. J. Lansing, M. Tubbs, et al., Using sentiment and momentum to predict stock returns, FRBSF Economic Letter 2018 (2018) 29.
- [28] A. Rodríguez-González, Á. García-Crespo, R. Colomo-Palacios, F. G. Iglesias, J. M. Gómez-Berbís, Cast: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator, Expert systems with Applications 38 (9) (2011) 11489–11500.
- [29] G. D. I. Anghel, Stock market efficiency and the MACD. evidence from countries around the world, Procedia economics and finance 32 (2015) 1414–1431.
- [30] L. Ortega, K. Khashanah, A neuro-wavelet model for the short-term forecasting of high-frequency time series of stock returns, Journal of Forecasting 33 (2) (2014) 134–146.
- [31] US dollar index (dx-y.nyb) - Yahoo Finance, <https://finance.yahoo.com/quote/DX-Y.NYB/>, Accessed on 06/19/2023
- [32] Cboe volatility index (VIX)-Yahoo Finance, <https://finance.yahoo.com/quote/%5EVIX/history?p=%5EVIX>, Accessed 06/19/2023