Stock market prediction of Bangladesh using multivariate long short-term memory with sentiment identification

4 authors:

Md Ashraful Islam
Daffodil International University
1 PUBLICATION **7** CITATIONS

SEE PROFILE

Md. Rana Sikder
Daffodil International University
1 PUBLICATION **7** CITATIONS

SEE PROFILE

Sayed Mohammed Ishtiaq
Daffodil International University
1 PUBLICATION **7** CITATIONS

SEE PROFILE

Abdus Sattar
Daffodil International University
92 PUBLICATIONS **634** CITATIONS

SEE PROFILE

Stock market prediction of Bangladesh using multivariate long short-term memory with sentiment identification

Md. Ashraful Islam, Md. Rana Sikder, Sayed Mohammed Ishtiaq, Abdus Sattar
Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

Article Info	ABSTRACT
	The prediction of stock market trends is a challenging task due to its dynamic and volatile nature. Research has shown that predicting the stock market, especially in developing nations like Bangladesh, is challenging due to the presence of multiple external factors in addition to technical ones. To address this, this study proposed a novel dataset that includes not only technical stock market data from 2014 to 2021, but also external factors such as news sentiment and other economic indicators like inflation, gross domestic product (GDP), exchange rate, interest rate, and current balance. The goal is to provide a comprehensive view of the Dhaka Stock Exchange (DSE), the largest stock market in Bangladesh. The main objective of this study is to predict the trend of DSE by taking into account both technical stock market data and relevant external factors, and to compare the predictions made with and without using external factors. The study utilized a multivariate long short-term memory (LSTM) neural network for the stock market trend prediction. The experimental results showed that the use of external factors improved the accuracy of the LSTM-based stock market trend predictions by approximately 24%.

Corresponding Author:

Md. Ashraful Islam
Department of Computer Science and Engineering, Daffodil International University
Daffodil Smart City (DSC), Ashulia, Dhaka-1341, Bangladesh
Email: ashraful15-12271@diu.edu.bd

1. INTRODUCTION

Stock market forecasting is crucial for understanding the state of a nation's economy due to its significant impact. Despite its complexity and challenges, efforts to predict stock market trends persist. However, the task is made difficult by the many external factors that influence the market [1]. The stock market in Bangladesh is no different. Advanced technology has made it easier for investors to access crucial stock market data. Analysts now use sophisticated tools and computers to predict the future trend of the stock market [2]. Every year new methods and algorithms for stock market prediction are introduced by researchers. The utilization of machine learning and data analysis as new resources is widespread in this field [3].

Stock market trends are influenced by a variety of technical and external factors. Technical factors are simple to find, but external factors are exceedingly challenging. Additionally, the stock markets in developing nations like Bangladesh may be influenced by more difficult variables, like news sentiment. As a result, relying just on the technical factors to predict the stock market will not yield reliable results. Technical parameters like open, high, low, close, and volume as well as external ones like: news sentiment score, inflation, gross domestic product (GDP), exchange rate, interest rate, and balance have all been taken into account in this research.

The main goals of this work are as follows: i) to predict the stock market price trend using technical and external factors and ii) to identify the correlation between these external factors and stock market price trend. The main objective of this paper is to predict the stock market of Bangladesh more accurately by

considering external factors like Bangla news sentiment, inflation, GDP, exchange rate, interest rate along with the stock market technical indicators.

2.   BACKGROUND AND RELATED WORKS

Technical indicator analysis has already contributed significantly to the field of stock market forecasting. Numerous publications on stock market forecasting of Bangladesh using technical factors are also available. Atsalakis and Valavanis [4] suggested a machine learning model that predicted the stock market with nearly 20% accuracy using only technical indicators as input. Kara *et al.* [5] used artificial neural networks and support vector machines with technical indicators to predict the Istanbul Stock Exchange trend. The stock market of Bangladesh is also predicted using the linear classification model [6]. This paper advised to consider logistic regression to predict Bangladesh stock market.

The GDP or economic growth of a country may also have an impact on the stock market [7]–[11]. The relationship between a nation's stock market and economic growth was discussed in a number of articles. A key external indicator for forecasting stock market patterns is economic growth or GDP [12]. The pace of inflation may also have an impact on stock market return [13]–[16]. Stock market return price is positively correlated with the inflation rate.

News sentiment has a bigger aspect on stock market trends. Positive and negative news sentiment can play a vital role in building the stock market trend [17]–[20]. Kalra and Prasad [21] used the classifier naive Bayes to detect positive and negative news sentiment and achieved up to 91.2% accuracy. In Bangladesh stock market, it depends on the Bangla financial news. So, Bangla financial news articles from some daily newspapers of Bangladesh can be used. Social media sentiment is used for stock market trend prediction, but data collection is difficult due to informal writing styles. Improved neural network algorithms were utilized in Billah *et al.* [22] to forecast the Bangladesh stock market. They used external features for prediction.

long short-term memory (LSTM)-based stock prediction is more effective than other time series-based methods. For Bangla language sentiment analysis, supervised machine learning with domain-specific Lexicon data dictionaries (LDD) provides better results. Bhowmik *et al.* [23] a rule-based Bangla text sentiment score (BTSC) algorithm that can classify the Bangla text sentiment.

3.   RESEARCH METHOD

3.1. Process flow diagram

Figure 1 shows the process for stock market prediction of Bangladesh using multivariate LSTM with sentiment identification. It starts with data collection, preprocessing, and sentiment identification. Then, an LSTM model is created, trained, and evaluated, and finally, the output is generated.

3.2. Data collection

The multivariate LSTM neural network model uses 11 variables as inputs. These are: technical variables (open, high, low, close, and volume), news sentiment variable (news sentiment score), and economic variables (inflation, GDP, exchange rate, interest rate, and balance). The resulting dataset CSV file has 11 columns for these variables. The 11 variables can be grouped into three categories for data collection:
− Technical Dhaka Stock Exchange data from (3/3/2014 to 12/29/2021).
− Financial news articles from The Daily Ittefaq (3/3/2014 to 12/29/2021) [24].
− Yearly economic data from International Monetary Fund (IMF) database (2014 to 2021).

3.2.1. Technical stock data collection

Data for the historical DSE index was gathered from the Dhaka Stock Exchange's official website. The dataset includes daily numbers for open, high, low, close, and volume of DSE from 2014 to 2021. Variables were created for the daily open, high, low, close, and volume columns in the main dataset.

3.2.2. Financial news articles for sentiment analysis

The "Beautiful Soup 4.4.0" library in Python was utilized to gather 7,695 financial news articles from The Daily Ittefaq. This is an efficient web scraping tool that can be used to obtain public data from any website. All the financial Bangla news articles from 3/3/2014 to 12/29/2021 was collected [24]. There were several news reports on the same day. The news articles were gathered and saved in CSV format. After collecting all the news articles, a sentiment identification algorithm model was applied. The results were combined to create a sentiment score for each date [23]. These sentiment scores served as the foundation for the news sentiment column in the dataset.
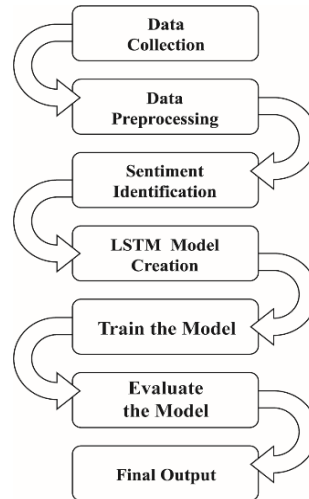
Figure 1. Process flow diagram

3.2.3. Yearly economic data

Economic data for Bangladesh was obtained from the IMF's global economic database for the period 2014 to 2021. The dataset contained information on various economic indicators, including inflation, GDP, exchange rate, interest rate, and current balance. We used this data to populate the economic variable column in our main dataset.

4. NEWS SENTIMENT IDENTIFICATION

The methodology for news sentiment identification came from Bhowmik *et al.* [23]. This work proposed the BTSC algorithm, a rule-based system for Bangla sentence-level news sentiment analysis. Some words may behave differently in sentiment analysis of the Bangla language depending on the relevant domain.

For example, in the phrase "শেয়ার ইস্যু করা" [issued shares], the word "ইস্যু" [issue] has a different meaning in business articles domain than our everyday life. Therefore, a business domain-specific extended sentimental lexicon dataset was created in order to obtain an accurate result in sentiment analysis on financial Bangla news items. Steps are described below for Bangla news sentiment identification: i) a finance domain-specific weighted sentiment lexicon dictionary was constructed in the Bangla language; ii) used modified BTSC algorithm to analyze sentimental scores from Bangla financial news articles [23]; iii) 7,695 financial Bangla news articles were collected from the year 2014 to 2021; iv) after collecting the data, a sentiment analysis process was used to gather sentiment scores for each day from 2014 to 2021 [24]; and v) these date-wise news sentiment data was merged with our existing technical indicator dataset.

4.1. Creation of financial lexicon data dictionary

The financial lexicon data dictionary is a list of words used to calculate the sentiment of financial news articles. Bangla words were collected from an online Bangla dictionary API and manually categorized into 6 weighted groups. To accurately determine the sentiment of sentences, a lexicon data dictionary is required. This project's lexicon data dictionary only contains Bangla words and includes words with positive sentiment and words with negative sentiment. Table 1 includes some examples from the financial LDD.

4.1.1. Bull words

This word collection is called bull words because, from a financial standpoint, they are considered to have positive connotations. These words are typically associated with upward market trends, increasing stock prices, and overall economic growth. In this sense, bull words are viewed as desirable and are often used by financial analysts and investors to convey optimism about the state of the economy.

4.1.2. Bear words

Bear word list is the opposite of positive sentimental words in financial sentiment analysis. For the purpose of evaluating the sentiment around business news, every phrase on this list is regarded as having a contradictory sentiment. Bear word lists typically consist of words that are associated with downward trends in the stock market, such as recession, inflation, unemployment, and bankruptcy.

Table 1. Financial lexicon data dictionary

Types	Examples	Weight	Word count
Bull words	সচ্ছল [well off], উন্নত [developed], শীর্ষ [peak], লাভবান [profitable], সমাধান [solution], কার্যকর [effective]	+1	3385
Bear words	অনুপস্থিত [absent], অকার্যকর [ineffective], বাতিল [canceled], দুর্ঘটনা [accident], দেউলিয়া [bankrupt]	-1	3966
Negative words	না [no], নেই [no], নাই [it's not], নয় [not], নিষেধ [forbidden]	-1	35
Coordinating conjunction	বরং [rather], কিন্তু [but], এবং [and], অথবা [or]	+2	21
Subordinating conjunctions	অধিকন্তু [furthermore], এমনকি [even], এসত্ত্বেও [even so]	+1.5	14
Adjectives and Adverbs	সবচাইতে [the most], সর্বাধিক [maximum], যথেষ্ট [enough]	+3	23
	অধিক [more], বেশি [more], অনেক [a lot]	+2	13
	সামান্য [a little], প্রায় [almost]	+0.5	25

4.1.3. Negative words

Negative word list has words like "না", "নয়", and "নেই" which can make a full sentence negative in the Bangla language. These negative words can have a significant impact on the overall sentiment of a sentence, even if the other words in the sentence are positive. The negative word list is a crucial tool for sentiment analysis in the Bangla language.

4.1.4. Coordinating conjunction words (Co con.)

In the Bangla language conjunctions like "কিন্তু", "আদপে", "এবং", "অথবা" plays an important role in sentence making. They should have their own weighted effect value in sentiment analysis [23]. By assigning weighted effect values to conjunctions in Bangla language, resulting in more accurate sentiment analysis.

4.1.5. Subordinating conjunctions (Sub con.)

Another kind of conjunctions list with words like "অধিকন্তু", "এমনকি", "বিশেষত". These conjunctions are often used to indicate a shift in tone or emphasis in a sentence and can play a significant role in shaping the overall sentiment. By assigning weighted effect values to these conjunctions, financial analysts can further refine their sentiment analysis, providing even more accurate insights into the sentiment of financial news and information.

4.1.6. Adjectives and adverbs (Adj.)

We listed some adjectives and adverbs like "সবচাইতে", "অধিক", "সর্বাধিক" as they are used to glorify the sentence sentiment more than other simple words. We categorized them into 3 weighted categories: high, medium, and low. Words with high weight have the greatest impact, words with medium weight have a moderate impact, and words with low weight have the least impact.

4.2. Data tokenization

To generate the news sentiment analysis, data preprocessing was performed on the collected data. The PyPI package Bengali natural language processing toolkit (BLTK), an open-source licenses (OSI)-approved Python library, was used for this preprocessing [25], which is licensed under the Massachusetts Institute of Technology (MIT) license. Data preprocessing was done to make the data more suitable for the news sentiment identification model. The first step in data preparation was tokenization, which involved splitting the news into individual sentences and then words. Sentiment analysis was performed at the sentence level for all financial news articles.

4.3. Data normalization

Data normalization is the process of removing the characters which are not necessary for sentiment analysis. Characters like ",", "।", ";", "#", "!", "@", "%", "$" have no meaning in sentiment counting. All special characters were removed from the tokenized datasets and question marks were also excluded. Unlike the BTSC algorithm, question marks were not taken into account in sentiment scores [23].

4.4. Stop words

Various stop words were removed from normalized data. When analyzing the sentiment dataset, stop words are regarded as having no impact. For example, words like "জন্য" [for], "পর্যন্ত" [up to], "নাগাদ" [by], "নিতে" [to take], "হত" [would have been] have no impact on sentiment analysis. The BLTK stop word list was used to remove stop words from the dataset [25]. For stop word removal, the BLTK package's level hard was utilized.

4.5. Data stemming

Data stemming means to remove any suffix/prefix of a word and convert the word to its root form. Suffixes and prefixes like "টি," "টার," "গুলো," and "গুলি" were removed from the words. For example: "শেয়ারবাজারে" will be "শেয়ারবাজার", "বছরের" will be "বছর" after stemming the word. Stemming the words helped to match the input data with the LDD dataset words for sentiment analysis.

4.6. Parts of speech tagger (POS tagger)

A POS tagger categorizes data into its parts of speech form. Given that the LDD consists of lists of adjectives, adverbs, and conjunctions, it is important to accurately identify the components of speech. Detecting them is crucial for implementing the algorithm of news sentiment analysis. In the Bangla language, there are 5 basic parts of speeches - বিশেষ্য, বিশেষণ, সর্বনাম, ক্রিয়াপদ, অব্যয় [23]. Among them সর্বনাম has no impact on sentiment score. Stop words were removed during the normalization stage.

4.7. Sentiment score counting

The BTSC algorithm has been modified for use in the financial Bangla news sentiment identification domain [23]. In the first step, a loop is run to read CSV news files from the dataset folder, which contains 7,095 news articles in CSV format. A variable "News_score" is declared and initialized to zero.

In the third step, a second loop is executed to tokenize each news article at the sentence level, splitting the entire news into sentences. The sentences are then normalized and stop words are removed. A variable "Sentence_score" is declared and initialized to zero. The sentences are further tokenized at the word level, with each word subjected to stemming and POS tagging functions [25].

In step 10, the LDD word lists are scanned. In step 11, if a word is identified as a bull word, a score of +1 is added to the sentence score. Bear words, conjunctions, adjectives, and adverbs are also checked against the LDD lists, and a score is assigned if a match is found. The coordinating conjunction score is +2, the subordination conjunction score is +1.5, while the scores for adjectives and adverbs (high, medium, and low) are +3, +2, and +0.5, respectively. In step 25, if a word is found to be in the negative words list, its existing sentence score is checked. The pseudocode is given in Table 2. If the score is positive, it is multiplied by -1. This is because if the sentence score is already negative and multiplied by -1, the result would be a positive sentiment score, which is incorrect. The loop ends in step 28. The full news score is calculated by summing the sentiment scores of each sentence in step 29. The remaining for loops end in steps 30 and 31.

Table 2. Modified BTSC algorithm

Steps: Algorithm Pseudocode
1.   **for** each News_articles.csv in datafolder **do**
2.   News_score = 0
3.   **for** each Tokenize (Sentence) in News **do**
4.   Normalizing (Sentence)
5.   Stop word (Sentence)
6.   Sentence_score = 0
7.   **for** each Tokenize (Word) in Sentence **do**
8.   Stemming (Word)
9.   POS tagger (word)
10.   Scanning LDD
11.   **if** word is Bull word in LDD **then**
12.   Sentence_score =+1
13.   **else if** word is Bear word in LDD **then**
14.   Sentence_score =-1
15.   **else if** word is Co con. in LDD **then**
16.   Sentence_score =+2
17.   **else if** word is Sub con. in LDD **then**
18.   Sentence_score =+1.5
19.   **else if** word is Adj. (High) in LDD **then**
20.   Sentence_score =+3
21.   **else if** word is Adj. (Med) in LDD **then**
22.   Sentence_score =+2
23.   **else if** word is Adj. (Low) in LDD **then**
24.   Sentence_score =+0.5
25.   **else if** word is Negative in LDD **then**
26.   **if** Sentence_score >=0
27.   Sentence_score *-1
28.   **end for**
29.   News_score =+ Sentence_score
30.   **end for**
31.   **end for**

4.8. Simulation of sentiment score counting

Now, the modified BTSC algorithm will be demonstrated with an example to showcase its simulation. The input consists of one of the news articles collected during the data collection stage. This example will provide a clear illustration of how the modified BTSC algorithm can be used in practice for sentiment analysis tasks.

First let's see the news: টানা দরপতন আর লেনদেন খরায় চলছে শেয়ারবাজার। ধারাবাহিকভাবে কমতে কমতে গতকাল শেয়ারবাজার ৪০০ কোটি টাকার নিচে নেমে গেছে। এমন লেনদেন খরার বাজারে প্রায় ১০০ প্রতিষ্ঠান ক্রয়াদেশ দিতে পারছে না । ফলে সর্বাধিক পতন হয়েছে মূল্যসূচকের।

[The stock market is running in continuous price decline and transaction drought. Yesterday, the main stock market of the country, Dhaka Stock Exchange (DSE), which has been continuously decreasing for quite some time, fell below Tk 400 crore. After April 5 last year, less than Tk 400 crore was traded in DSE. In the dry market of such transactions, the purchase order room of nearly 100 companies became empty. As a result, there has been a big fall in the price]. After sentence level tokenization: ['টানা দরপতন আর লেনদেন খরায় চলছে শেয়ারবাজার', 'ধারাবাহিকভাবে কমতে কমতে গতকাল শেয়ারবাজার ৪০০ কোটি টাকার নিচে নেমে গেছে', 'এমন লেনদেন খরার বাজারে প্রায় ১০০ প্রতিষ্ঠান ক্রয়াদেশ দিতে পারছে না ', 'ফলে সর্বাধিক পতন হয়েছে মূল্যসূচকের']. After word tokenization, normalization, removing stop words & stemming:

- 1st sentence: ['টানা', 'দরপতন', 'লেনদেন', 'খরা', 'চলছে', 'শেয়ারবাজার']
- 2nd sentence: ['ধারাবাহিকভাবে', 'কমতে', 'কমতে', 'শেয়ারবাজার', 'কোটি', 'টাকা', 'নিচে', 'নেমে']
- 3rd sentence: ['লেনদেন', 'খরা', 'বাজার', 'প্রতিষ্ঠান', 'ক্রয়াদেশ', 'পারছে','না']
- 4th sentence: ['সর্বাধিক', 'পতন', 'হয়েছে', 'মূল্যসূচক']

The simulation of the modified BTSC algorithm is presented in Table 3. Here, the news has a negative score, indicating that it has a negative sentiment. Here, (*-1) indicates multiplying the existing *sentence_score* by (-1).

Table 3. Simulation of modified BTSC algorithm

Sentence	Scoring				Sentence score
1	টানা [continuous] null 0	দরপতন [fall in price] bear -1	লেনদেন [transaction] bull +1	খরা [drought] bear -1	-1
	চলছে [happening] null 0	শেয়ারবাজার [stock market] null 0			
2	ধারাবাহিকভাবে [consistently] bull +1	কমতে [decrease] bear -1	কমতে bear -1	শেয়ারবাজার [stock market] null 0	-2
	কোটি [crore] null 0	টাকা [taka] null 0	নিচে [down] bear -1	নেমে [fall] null 0	
3	লেনদেন [transaction] bull +1	খরা [drought] bear -1	বাজার [market] null 0	প্রতিষ্ঠান [institution] null 0	-1
	ক্রয়াদেশ [purchase order] bull +1	পারছে [can] null 0	না [not] neg *-1		
4	সর্বাধিক [maximum] adj. (high) +3	পতন [downfall] neg *-1	হয়েছে [happened] null 0	মূল্যসূচক [value] null 0	-3
				News_score	-7

4.9. Date wise news sentiment score counting

The scores of all the news for each day were tallied, and the total score was divided by the number of news articles on that day to obtain the average news sentiment score for each date. This allowed us to obtain the average news sentiment score for each date, which was useful in identifying trends and patterns in news sentiment over time. We incorporated the sentiment results into our main dataset, organizing them by date.

5.   MULTIVARIATE LSTM MODEL

Recurrent neural networks (RNNs) excel in learning patterns in time series data, making them effective for time series prediction. LSTM neural networks have an architecture that allows for more accurate

prediction of time series patterns, both in the short and long term. A multivariate LSTM was chosen over a univariate LSTM, as it takes into account the correlation of multiple variables for more accurate predictions.

5.1. Time series analysis using LSTM in python

The LSTM neural network for stock market prediction used the Dhaka Stock Exchange dataset from 2014 to 2021 for training. The trend was then predicted. The last date in the stock market data was December 29, 2021, and the stock price for December 30, 2021 was predicted using the LSTM model. The steps for the prediction with the multivariate LSTM neural network model are as follows: i) reading the dataset, ii) feature selection and scaling, iii) data cleaning and transforming; iv) training the LSTM neural network, and v) predict next day's price.

5.2. Prerequisites

The model building and prediction process utilized a Python 3 environment and necessary Python packages. The following standard packages were installed and used: pandas, NumPy, math, and matplotlib. Pandas was used for data frame operations and matplotlib for plotting graphs and results.

5.3. Dhaka Stock Exchange dataset

The CSV file contains the stock market data of the Dhaka Stock Exchange from 2014 to 2021. Technical variables, such as open, high, low, close, and volume, are considered, along with external variables, such as news sentiment score, inflation, GDP, exchange rate, interest rate, and balance. The "Open" variable serves as the output, as it allows for easy evaluation of the stock market trend through the daily opening price of the stock market. The remaining variables act as input variables. Preview of the dataset is shown in Table 4.

Table 4. Preview of DSE stock market dataset

Date	Open	High	Low	Close	Volume	Sentiment	Inflation	GDP	Exchange	Interest	Balance
03-03-2014	4697.3	4732.25	4685.53	4687.19	50542994	0.5	7.34	172.88	77.64	9.08	1.406
03-04-2014	4687.19	4728.12	4687.19	4703.88	46282176	-1.5	7.34	172.88	77.64	9.08	1.406
03-05-2014	4703.88	4717.24	4684.67	4697.54	35469576	5.0	7.34	172.88	77.64	9.08	1.406
03-06-2014	4697.54	4724.82	4690.5	4699.63	41292218	-0.6	7.34	172.88	77.64	9.08	1.406
03-03-2014	4699.63	4715.65	4687.2	4687.2	51064991	4.6	7.34	172.88	77.64	9.08	1.406

5.4. Feature selection and scaling data tokenization

The data must be cleaned and scaled to achieve accurate predictions from the model. 11 input features (open, high, low, close, volume, news sentiment score, inflation, GDP, exchange rate, interest rate, and balance) are selected for feature selection and the opening price serves as the output prediction feature. To improve accuracy and speed up model training time for prediction, the dataset must be scaled. The MinMaxScaler approach from the scikit-learn package is used to standardize the input data to a range of 0 to 1. Finally, the data is unscaled to retrieve the original values from the dataset.

5.5. Data cleaning and transforming

Data cleaning involves finding and filling missing values in the dataset. In this case, there are no missing values in the stock market dataset. The LSTM sliding windows approach was used to train and validate the dataset through a neural network. The data was split into two parts: 80% for training the model and 20% for testing.

5.6. Training the LSTM neural network

The multivariate LSTM neural network model consists of 4 layers: The first layer is an LSTM layer. It takes our mini-batches from the sliding window process as input and returns the whole sequence. The second LSTM layer again takes the returned sequence from the first layer as input. The third layer is a dense layer that consists of 5 layers. The last dense layer returns the predicted value.

The sequence length of the model is 50, resulting in a mini-batch matrix of 50 steps and 11 features from the sliding window process. The size of the first layer's input neurons is equal to the size of the mini-batch input data. The total input layer of the LSTM neural model comprises 550 neurons, calculated as (50*11). The epoch count is set at 50.

5.7. Predict next day's price

After training the multivariate LSTM neural network, the next day's opening price can be forecasted by processing the necessary data from the dataset and running the model. The minimum required input for the model is 50 steps of the dataset from the sliding window process, due to the sequence length of 50-time steps. The model outputs the predicted opening price for the next day, following the end of the input dataset.

5.8. Model performance

The train and validation dataset were input into the LSTM model, and the model was run with both the training and validation data for 50 epochs. A comparable run was performed without external variables. The resulting training and validation loss curve without the use of external factors is displayed in Figure 2. The training and validation loss curve using all 11 variables is displayed in Figure 3. We used different patterns to represent the curves of both the training and validation loss.
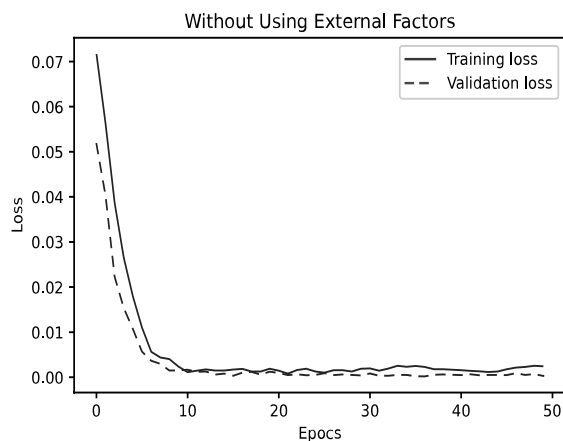
Figure 2. Loss curve without using external factors   Figure 3. Loss curve using all the 11 factors
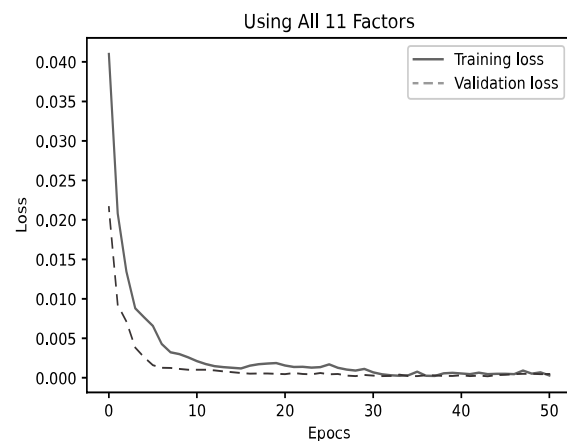
6.   RESULT AND ANALYSIS

Comparison of the accuracy is shown in Table 5. The first model, which only employs technical indicators achieved an accuracy of 62% using multivariate LSTM. In contrast, the second model utilized both technical and external features. This model achieved a significantly higher accuracy of 86% using multivariate LSTM. External factors enhanced the accuracy of LSTM-based stock market trend predictions by around 24%.

Figure 4 graph displays the results of a multivariate LSTM neural network model that utilizes historical technical data variables (open, high, low, close, and volume) as input data, without the incorporation of external factors. The predictions of the LSTM model are plotted against the original opening price in Figure 4. The results indicate that the model, relying solely on technical variables, can capture a long-term upward trend, but fails to accurately predict short-term trends, which is a crucial aspect in daily trading. Only the last few days are plotted in the figure for clearer visualization.

Figure 5 graph presents the results of an LSTM model that includes both technical data variables and external factors. A total of 11 variables were utilized as input. The predictions of the LSTM model are again plotted against the original opening price in Figure 5. The results demonstrate that the incorporation of external factors improves the accuracy of the model's predictions of the stock market trend.

Accuracy of the prediction was only measured graphically. Error calculation methods, such as mean percentage absolute error (MAPE), mean absolute error (MAE), and root mean square error (RMSE), were intentionally avoided as they would not accurately detect the accuracy of the result. The model predicted the opening price of the next day after the end of the training dataset, which was 30 December 2021, as 6721.46 taka. In reality, the opening price at the Dhaka Stock Exchange was 6731.15 taka.

Table 5. Comparison of accuracy (%)

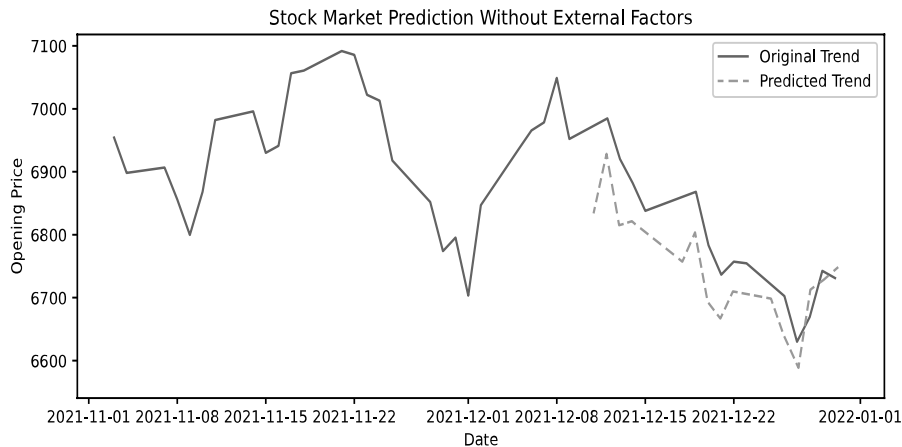| Method	Variables	Accuracy
Multivariate LSTM only using technical indicators	Date, Open, High, Low, Close, Volume	62%
Multivariate LSTM using both technical and external features	Date, Open, High, Low, Close, Volume, News sentiment, Inflation rate, GDP, Exchange rate, Interest rate, Current balance	86%

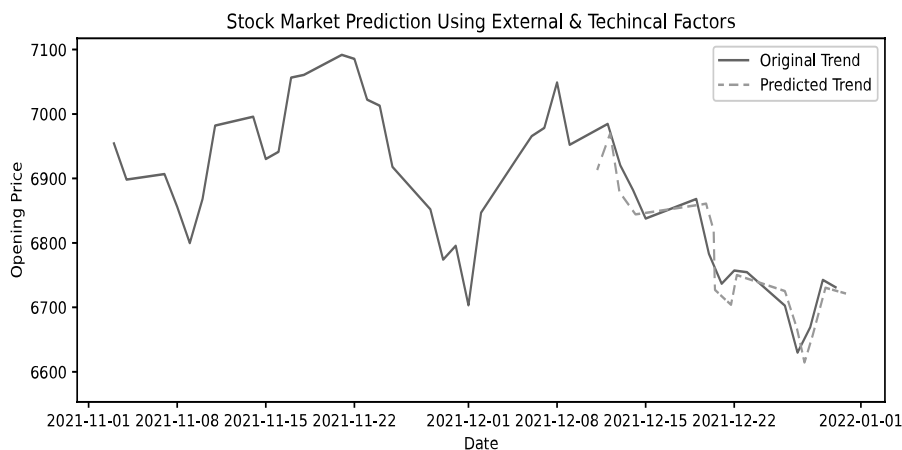Figure 4. Opening price trend without using external factors

Figure 5. Opening price trend using technical and external factors

7. CONCLUSION AND FUTURE WORK

The aim of this paper was to predict the stock market in Bangladesh and determine the correlation between external factors such as news sentiment and other economic aspects for improved prediction. News sentiment alone has a significant ability to detect many socially affecting factors on the stock market trend. In the future, a multivariate multilayer LSTM or bi-LSTM neural network may be used for better prediction. Public mood-driven asset allocation is also a possibility. To gather sentiment data, multiple news sources can be utilized and social media sentiment can also play a crucial role. The combination of these external factors with financial news sentiment has the potential to produce more accurate results in multivariate LSTM stock market prediction.

REFERENCES

[1] M. Qi and G. S. Maddala, "Economic factors and the stock market: a new perspective," *Journal of Forecasting*, vol. 18, no. 3, pp. 151–166, May 1999, doi: 10.1002/(SICI)1099-131X(199905)18:3<151::AID-FOR716>3.0.CO;2-V.

[2] I. Svalina, V. Galzina, R. Lujić, and G. Šimunović, "An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices," *Expert Systems with Applications*, vol. 40, no. 15, pp. 6055–6063, Nov. 2013, doi: 10.1016/j.eswa.2013.05.029.

[3] D. K. Sharma, H. S. Hota, K. Brown, and R. Handa, "Integration of genetic algorithm with artificial neural network for stock market forecasting," *International Journal of System Assurance Engineering and Management*, vol. 13, no. S2, pp. 828–841, Jun. 2022, doi: 10.1007/s13198-021-01209-5.

[4] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques – part II: Soft computing methods," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5932–5941, Apr. 2009, doi: 10.1016/j.eswa.2008.07.006.

[5] Y. Kara, M. Acar Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5311–5319, May 2011, doi: 10.1016/j.eswa.2010.10.027.

[6]   M. Karimuzzaman, N. Islam, S. Afroz, and M. M. Hossain, "Predicting stock market price of Bangladesh: A comparative study of linear classification models," *Annals of Data Science*, vol. 8, no. 1, pp. 21–38, Mar. 2021, doi: 10.1007/s40745-020-00318-5.

[7]   L. Carp, "Can stock market development boost economic growth? empirical evidence from emerging markets in central and Eastern Europe," *Procedia Economics and Finance*, vol. 3, pp. 438–444, 2012, doi: 10.1016/S2212-5671(12)00177-3.

[8]   A. A. Enisan and A. O. Olufisayo, "Stock market development and economic growth: Evidence from seven sub-Sahara African countries," *Journal of Economics and Business*, vol. 61, no. 2, pp. 162–171, Mar. 2009, doi: 10.1016/j.jeconbus.2008.05.001.

[9]   C. Ball and J. French, "Exploring what stock markets tell us about GDP in theory and practice," *Research in Economics*, vol. 75, no. 4, pp. 330–344, Dec. 2021, doi: 10.1016/j.rie.2021.09.002.

[10]   J. E. Stiglitz, "Capital market liberalization, economic growth, and instability," *World Development*, vol. 28, no. 6, pp. 1075–1086, Jun. 2000, doi: 10.1016/S0305-750X(00)00006-1.

[11]   C. Florackis, G. Giorgioni, A. Kostakis, and C. Milas, "On stock market illiquidity and real-time GDP growth," *Journal of International Money and Finance*, vol. 44, pp. 210–229, Jun. 2014, doi: 10.1016/j.jimonfin.2014.02.006.

[12]   C.-K. Choong, A. Z. Baharumshah, Z. Yusop, and M. S. Habibullah, "Private capital flows, stock market and economic growth in developed and developing countries: A comparative analysis," *Japan and the World Economy*, vol. 22, no. 2, pp. 107–117, Mar. 2010, doi: 10.1016/j.japwor.2009.07.001.

[13]   M. Omran and J. Pointon, "Does the inflation rate affect the performance of the stock market? The case of Egypt," *Emerging Markets Review*, vol. 2, no. 3, pp. 263–279, Sep. 2001, doi: 10.1016/S1566-0141(01)00020-6.

[14]   A. Jain and P. C. Biswal, "Dynamic linkages among oil price, gold price, exchange rate, and stock market in India," *Resources Policy*, vol. 49, pp. 179–185, Sep. 2016, doi: 10.1016/j.resourpol.2016.06.001.

[15]   G. Bekaert and E. Engstrom, "Inflation and the stock market: Understanding the 'fed model,'" *Journal of Monetary Economics*, vol. 57, no. 3, pp. 278–294, Apr. 2010, doi: 10.1016/j.jmoneco.2010.02.004.

[16]   R. P. Pradhan, M. B. Arvin, and S. Bahmani, "Causal nexus between economic growth, inflation, and stock market development: The case of OECD countries," *Global Finance Journal*, vol. 27, pp. 98–111, Jul. 2015, doi: 10.1016/j.gfj.2015.04.006.

[17]   Y. Jiang, "Investor sentiment and stock market response to earnings announcement," in *2011 International Conference on Management and Service Science*, Aug. 2011, pp. 1–4, doi: 10.1109/ICMSS.2011.5998465.

[18]   X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Information Processing & Management*, vol. 57, no. 5, Sep. 2020, doi: 10.1016/j.ipm.2020.102212.

[19]   D. E. Allen, M. McAleer, and A. K. Singh, "Daily market news sentiment and stock prices," *Applied Economics*, vol. 51, no. 30, pp. 3212–3235, Jun. 2019, doi: 10.1080/00036846.2018.1564115.

[20]   M. W. Uhl, "Reuters sentiment and stock returns," *Journal of Behavioral Finance*, vol. 15, no. 4, pp. 287–298, Oct. 2014, doi: 10.1080/15427560.2014.967852.

[21]   S. Kalra and J. S. Prasad, "Efficacy of news sentiment for stock market prediction," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Feb. 2019, pp. 491–496, doi: 10.1109/COMITCon.2019.8862265.

[22]   M. Billah, S. Waheed, and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," in *2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, Dec. 2016, pp. 1–4, doi: 10.1109/ICECTE.2016.7879611.

[23]   N. R. Bhowmik, M. Arifuzzaman, M. R. H. Mondal, and M. S. Islam, "Bangla text sentiment analysis using supervised machine learning with extended lexicon dictionary," *Natural Language Processing Research*, vol. 1, no. 3–4, pp. 34–45, 2021, doi: 10.2991/nlpr.d.210316.001.

[24]   H. Tasmima, "The Daily Ittefaq," Ittefaq. https://www.ittefaq.com.bd/business (accessed Jun. 20, 2022).

[25]   S. Hossain, "BLTK: The Bengali natural language processing toolkit," Python Package Index. https://pypi.org/project/bltk/ (accessed Jun. 15, 2022).

BIOGRAPHIES OF AUTHORS

Md. Ashraful Islam is currently pursuing his B.Sc. in Computer Science in the Department of Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh. His research interests include data mining, artificial intelligence, digital image processing, machine learning, data Science, web and information retrieval, human-computer interaction and information visualization. He can be contacted at email: ashraful15-12271@diu.edu.bd.

Md. Rana Sikder is currently pursuing his B.Sc. in Computer Science in the Department of Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh. His research interests include natural language processing, computer vision, data mining, embedded and real time systems, virtual reality and intelligent system. He can be contacted at email: rana15-12511@diu.edu.bd.

Sayed Mohammed Ishtiaq is currently pursuing his B.Sc. in Computer Science in the Department of Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh. His research interests include computer graphics, computer vision, data mining, large scale networks, block chain technology and knowledge processing. He can be contacted at email: sayed15-12530@diu.edu.bd.

Abdus Sattar is currently doing a Doctor of Philosophy (Ph.D.) in Computer Science & Engineering at Bangladesh University of Professionals (BUP). Previously, He received a Bachelor of Science in Computer Science and Engineering (CSE) from Ahsanullah University of Science and Technology (AUST) and a Master's Program of Interactive Systems Engineering (ISE) from KTH-Royal Institute of Technology, Sweden. During the Master's thesis research and development project, He was employed as a Research Assistant on a Research Project. This Research Project work was carried out in collaboration with Södertörn University, Stockholm University, and Karolinska Institute Innovations, Sweden. Currently, He is employed as an Assistant Professor in the Department of Computer Science and Engineering (CSE) at Daffodil International University (DIU). Previously, He was employed as an Assistant Professor in the Department of Computer Science and Engineering (CSE) at Britannia University, Comilla. He can be contacted at email: abdus.cse@diu.edu.bd.