Stock market index prediction based on market trend using LSTM

Ankireddy Yenireddy¹, Marimganti Srinivasa Narayana², Kalla Venkata Bangaru Ganesh³, Guvvaladinne Prasanna Kumar⁴, Madduri Venkateswarlu⁵

¹Department of Master of Business Administration, Narasaraopeta Engineering College, Narasaraopet, India
²Department of Centre for Distance and Online Education, Koneru Lakshmaiah Education Foundation University, Vijayawada, India
³Department of Master of Business Administration, Koneru Lakshmaiah Education Foundation University, Vijayawada, India
⁴Department of Electronics and Communication Engineering, Malla Reddy Engineering College, Secundrabad, Telangana, India
⁵Department of Electronics and Communication Engineering, Potti Sritamulu Chaluvadi Mallikarjuna Rao College of Engineering and Technology (PSCMRCET), Vijayawada, India

ABSTRACT

The stock market data analysis has received interest as a result of technological advancements and the investigation of new machine learning models, since these models provide a platform for traders and business people to choose gaining stocks. The business price prediction is a challenging and extremely complex process due to the impact of several factors on company prices. The numerous patterns that the stock market goes, they have been the focus of extensive research and analysis by numerous experts. There are several large data sets accessible, an artificial intelligence and machine learning techniques are developing quickly, and because of the machine’s improved computational power, complex stock price prediction algorithms can be developed. This paper presents stock market index prediction based on market trend using long short-term memory (LSTM). Using built-in application programmable interface (API), Yahoo Finance offers a simple method to programmatically retrieve any historical stock prices of an organization using the ticker name. The standard and poor’s 500 index (S&P 500 index) include the firms that have been taken into consideration here. Utilizing the selected input variable, single-layer and multi-layer LSTM models are implemented, and the measurement parameters of mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R) are used to compare each performance. Nearly all of the real closing price’s curve and the prediction curve’s closing price for test data overlap. A potential stock investor may benefit significantly from such a prediction by using it to make well-informed choices that would increase his earnings.

Keywords:
Classification
Long short-term memory
Root mean square error
Stock market index prediction
Stock prices

1. INTRODUCTION

The primary objective of the original methods, a common person may trade stocks make investments, profit from businesses that sell a portion of themselves on this platform is the stock market [1]. A developing country like India experiences rapid economic expansion mostly supported by the stock market. Therefore, the development of our country and other emerging nations may be impacted by stock
market performance [2]. A rising stock market would indicate strong economic development for the entire country. Economic development would slow significantly if the stock market fell [3]. In other words, we may argue that the success of the stock market closely correlates with both national and worldwide economic growth [4]. Due to the dynamic nature of the stock market, only 10% of people in a particular country actively participate in the stock market. A common misunderstanding regarding the stock market is that purchasing or selling shares is equivalent to gambling [5]. The fundamental idea is really simple, these companies would offer their shares as small commodities known as stocks [6]. To raise money for the business, they act in this manner. The initial public offering (IPO) occurs when a company issues its shares for a specific price. The offer price is the amount at which the firm will sell its stock and raise financing [7]. On a market like the Bombay stock exchange (BSE) after, the owner may sell the shares to a buyer at any price, at this moment it belongs to the individual [8]. Despite the fact that investors and buyers still sell these shares for their own prices, the corporation only keeps the profits from the IPO. The ongoing transfer of hares between parties in an effort to create profits are leads to an increase in the price of the specific share following each profitable transaction [9]. However, the exchange rate in the market, when the business makes with less shares, the price drops and traders lose money. In a nutshell, the increase and fall in stock values, since the worry people feel while making investments in the stock market, are all caused by this phenomena.

To forecast future profits, corporations, investors, and equity traders must develop the intricate and difficult process of stock price prediction [10]. The nature of the stock market makes it a chaotic, non-parametric, noisy, non-linear system. Making accurate and precise pricing predictions becomes difficult. The stock price changes are unpredictable, and a number of interrelated factors contribute to this behaviour. Potential explanations include things like global economic data, shifts in the unemployment rate, monetary policies of immigration policies, significant countries, natural disasters, and situational factors that affect public health [11]. Following a comprehensive market analysis, all market participants want to increase earnings and decrease risks. Integrating all the many information sources together in one place collecting them is the main problem, implementing an accurate representation to make precise forecasts.

An additional difficult issue in stock prediction is the identification of characteristics from the financial data, and numerous strategies have been proposed [12]. Machine learning has the potential to revolutionize the field of stock price prediction. Unerringly accurate predictions may be made using machine learning approaches, which have the power to reveal patterns and insights we hadn’t previously seen [13]. The development of machine learning is accelerating dramatically in the modern world. To address diverse issues and the inherent structure of datasets, there are a number of deep learning architectures available. In a basic feedforward neural network design, information just travels forward. It does not maintain data from the prior stage since each input is handled individually [14]. As a result, when dealing with sequential data, these models are unimportant since it takes a series of past occurrences to accurately forecast the future.

Predicting the behaviour of stock market prices is quite difficult due to a number of elements, there is a significant amount of data and a very low signal to noise ratio that must be considered. This study investigated the effects of machine learning models. While earlier studies have explored the impact of stock price prediction, they have not explicitly addressed its influence on efficiency of the model. Whenever the model asks to analyze time series data or natural language, recurrent neural networks (RNNs) are utilized [15]. Due to the RNN architecture’s use of loops, the relevant data may be preserved throughout time. Internally, the network is communicating information from one time step to the next. As a result, for time series applications like sequential data modelling, the RNN is more appropriate, language translation, message/email auto-completion, and stock market predictions. The ability to give each sample different weights which makes long short-term memory (LSTM) is one of the most effective RNN systems. It can handle extended input sequences more effectively than other RNNs. Therefore, a very high level of accuracy may be attained when forecasting future trends and price forecasts for various equities using the LSTM network. Following is an organization of the remaining paper: section 2 discusses the research on stock price prediction in the literature. In section 3 explains the described framework of stock market index prediction based on market trend using LSTM, section 4 of the analysis contains the results and discussions, and section 5 brings it all to a conclusion.

2. LITERATURE SURVEY

Bouktif et al. [16] an improved approach of sentiment analysis to objectively investigate the prediction of stock market movement direction, the author contributes positively to this discussion. For a more detailed analysis, they specifically work with the history of stock prices, sentiment polarity, subjectivity, N-grams, and customized text-based characteristics. They have gathered and examined the NASDAQ stocks of 10 significant corporations that fall under various stock categories. Described suggested
model is operating as expected and forecasting stock movements with a greater accuracy of 60% when compared to existing sentiment-based stock market prediction techniques, such as deep learning.

Chen et al. [17] rate and direction of the movement of the stock price is predicted using an adaptive stock price trend predicting model (TPM) built on an encoder-decoder architecture. There are two steps to this concept. In the first, it would be helpful to extract more information from the market data using a dual feature extraction technique based on several time periods. Then, during the TPM's second stage of development, to choose, integrate relevant dual features predict the stock price trend, this encoder-decoder architecture based on the dual attention mechanism is used. They gathered high-frequency market data to assess described suggested TPM. The results of the experiment show that the proposed TPM performs the more accurately in terms of prediction than the present advanced methods.

Wang et al. [18] proposed hybrid time-series predictive neural network (HTPNN) for calculating impact on news. As distributed word vectors, news headline properties are expressed, in order to improve the model’s performance is dimensionally decreased by sparse automated encoders. The news is then combined taking into consideration the way stocks change with the daily K-line data. By recognizing the way that news and time series combine, HTPNN is able to grasp the prospective law of stock price fluctuation. Described approach combines a wider range of stock features and provides greater benefits in terms of running speed when compared to advanced models. In addition, accuracy often increases by over 5%.

Wang et al. [19] a clustering technique is described that uses morphological similarity distance (MSD) with k-means clustering to mining related stocks, then using an online learning model called hierarchical temporal memory (HTM) predictions are made using clustering-HTM after learning patterns from similar stocks. The results of the price prediction experiments indicate that; i) C-HTM predicts prices more accurately than HTM, which has not been trained on identical stock patterns, and ii) C-HTM performs better in terms of short-term predicts than the baseline models.

Kim et al. [20] a number of machine learning approaches are combined with time-varying effective transfer entropy (ETE) to predict the development of US stock prices. By analyzing the relationship between stock market crises and Granger-cause links, initially, they decide whether the ETE may be utilized as a market explanatory variable based on 3- and 6-months moving frames. With ETE’s practical application, the direction of the stock price is being predicted for the first time in this analysis.

Cao et al. [21] creates a robust evolutionary framework for fuzzy rough neural networks by integrating the best aspects of the two systems outlined above. Prior to improving the outcome nodes, they initially add rough neurons, and then combine the current fuzzy rough neural network model with the interval type-2 fuzzy set. Compared to other models, the modified fuzzy rough neural network model shows considerable gains with optimization strategies.

Wen et al. [22] a novel technique is designed to rebuild sequences from noisy financial temporal series by exploiting motifs (regular patterns), and then to infer the spatial arrangement of the time series using a convolutional neural network (CNN). The results of the experiment demonstrate the effectiveness of described recommended approach for feature learning and it’s outperformed over a 4%-7% accuracy gain over the frequency trading patterns modelling approach and deep learning-based signal processing techniques.

Lee et al. [23] applied deep Q-network with a CNN function approximator, which takes stock chart images as input for making global stock market predictions. We trained our model only on US stock market data and tested it on the stock market data of 31 different countries over 12 years. The results demonstrate that artificial intelligence-based stock price forecasting models can be used in relatively small markets (emerging countries) even though small markets do not have a sufficient amount of data for training.

Zhang et al. [24] to improve the prediction for stock market composite index movements, we exploit the consistencies among different data sources, and develop a multi-source multiple instance model that can effectively combine events, sentiments, as well as the quantitative data into a comprehensive framework. To effectively capture the news events, we successfully apply a novel event extraction and representation method. Evaluations on the data from the year 2015 and 2016 demonstrate the effectiveness of our model. In addition, our approach is able to automatically determine the importance of each data source and identify the crucial input information that is considered to drive the movements, making the predictions interpretable.

Alsulmi [25] an automatic labeling approach is described that exploits a metaheuristic search to perform the labeling task for stock market data. The results of empirical experiments demonstrate that this approach is very promising as it outperforms the current manual approaches for stock data labeling and achieves higher labeling effectiveness.

3. STOCK MARKET INDEX PREDICTION

The block diagram of stock market index prediction based on market trend using LSTM is represented in below Figure 1. Using built-in application programmable interfaces (APIs), Yahoo Finance

Stock market index prediction based on market trend using LSTM (Ankireddy Yenireddy)
offers a simple method to programmaticaly retrieve any historical stock prices of an organization using the ticker name. It offers the option to obtain costs with a beginning and final date. The standard and poor’s 500 index (S&P 500 index) include the companies that have been taken into consideration. The index is designed to illustrate the big 500 businesses stocks are performing and moving in price that are listed on the New York Stock Exchange (NYSE) and national association of securities dealers automated quotations (NASDAQ). Products based on the S&P500 are some of the most liquid, actively traded, and invested in alternatives in the majority of the markets. Additionally, the S&P 500 accounts for 75% of the market capitalization of the stock market. Additionally, given that the S&P500 is a market capitalization weighted index, organizations with higher stock values have a greater impact on the index.

Since the data is in raw form, analysis is not possible. The information includes the traded stock’s highest value, lowest value, starting price, closing price, and volume for a specific date. They place the highest attention on the stock’s closing date and price. For each companies, sector, and index, “Momentum” and “Volatility” are two additional factors that are calculated using the closing price of a stocks. Sector momentum, stock momentum, and index momentum are taken into consideration for every business in the dataset. Additionally taken into consideration, is the index, sector, and business volatility. This is carried out for each and every business.

Each of the above-discussed input factors makes some kind of contribution to predicting the closing price. In the heatmap, the correlation between the variables on the horizontal and vertical axes is shown as a numerical number. For instance, the correlation between the variable and itself is 1 when the matrix’s diagonal value is 1. As a result, they will ignore the data on the diagonal in our analysis. The feature selection procedure uses the entries on the off-diagonal. These numbers are displayed depending on the color’s intensity, it additionally operates as an estimation of the way the relevant elements are associated. The graph’s adjacent vertical bar displays a scale from 0 to 1 for the color’s intensity. There may or may not be a strong association between the closing price and the other factors. It conveys the degree of the relationship’s intensity. Duplicate characteristics are indicated by a significant connection between open price and closing price. One of the characteristics can be eliminated because it won’t add much to the predictions are understanding.

![Figure 1. Block diagram of stock market index prediction](image-url)

The sequential discrete format of stock price data is noisy. It is standard practice to denoise time-series data using the discrete wavelet transformation. There is a lot of variation since the input variables values change from one to the next step. In this case, the difference between the close price of the stock index
and the interest rate is significant. 695.33 is the close price’s standard deviation, this exceeds the interest rate’s standard deviation, which is significantly greater than 1.664. Whenever the range of one characteristic swing more wildly than the others, a large percentage of ML algorithms might not work well. To address this concern, they have used a min-max normalization method to the feature scaling. In the (1) represents the min-max normalization technique:

$$Z = \frac{x - x_{min}}{x_{max} - x_{min}}$$ \hspace{1cm} (1)

In (1), where the original input is represented by $x$ and scaled by $z$. Similar to this, where the input's minimum and maximum values are represented by $x_{min}$ and $x_{max}$, respectively. The LSTM model’s input sequence is then made using a given time step. For predicting time series, RNNs frequently use the deep learning method LSTM. For instance, classification and regression issues are both addressed by LSTM. The cell state is crucial to the LSTM architecture, it maintains the information flow by moving linearly through the chain. Information about the cell state is deleted or modified by the LSTM’s gate mechanism. It is a method of selective information transmission that combines the sigmoid layer, hyperbolic tangent layer, and point-wise multiplication.

The models has been modified to account for the hyperparameters, learning rate, neurons, batch size, epochs, and time step are a few examples of these parameters. Regularization approaches have been used to address the overfitting difficulties (number of observations, size of time step, and number of input characteristics) all three-dimensional input data that the LSTM model predicts. Making the input data consistent with the models, thus the relevant procedures are considered. In order to build the train-test datasets, time series data are different in this regard, they divided the whole amount of data into two parts, 80% and 20%. The set of validations is re-included in the training data once the hyperparameters have been tuned. The final models are then fitted using optimized hyperparameters to the whole training set of data. Finally, the test results are provided together with the performance ratings.

4. RESULT ANALYSIS

The main objective is to accurately predict the S&P 500 index’s closing price. Using built-in APIs, Yahoo Finance offers a simple method to programatically retrieve any historical stock prices of an organization using the ticker name. In order to forecast the closing price, they utilize LSTM architecture. By computing three separate performance measures (root mean square error (RMSE), correlation coefficient (R), and mean absolute error (MAE)), prediction accuracy and dependability of these models are evaluated. The (2) defines RMSE, calculates the difference between actual and estimated values.

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

Between actual and expected values, R calculates the linear correlation as shown in (3). A higher R value denotes a similarity between the expected and actual series. In addition, performance ratings are determined once the inverse transformation is applied to the predictions made from the normalized data.

$$R = \frac{\sum_{i=1}^{N} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}(\hat{y}_i - \bar{\hat{y}})^2}$$ \hspace{1cm} (3)

In (4) shows MAE calculation, calculated is the average variation between estimated and actual data. Due to the fact that it also is known as scale-dependent accuracy because it establishes the accuracy of observations taken at the same scale. For regression models in machine learning, it provides an assessment metric. It computes the differences in values between the model’s predicted values and the actual values.

$$MAE = \frac{1}{N} \sum |y_i - \hat{y}_i|$$ \hspace{1cm} (4)

Where, $y_i$: unique time-series, $\bar{y}$: the typical value of the initial time series, $\hat{y}_i$: this time series predictions calculated using the model, $\bar{\hat{y}}$: Value on average of the predicted time series, $N$: the quantity of observations.

The optimal model would be one that has the lowest RMSE and MAE together with the highest R. Comparative analysis of described stock market index prediction based on market trend using LSTM and other techniques as linear regression (LR), the support vector machine (SVM) is stated in Table 1. Figure 2
shows the comparative performance analysis of different techniques-based stock market index prediction. From results are clear that Figure 2(a) low RMSE and Figure 2(b) MAE with highest Figure 2(c) linear correlation are obtained for LSTM model. The objective is to project with high accuracy the closing price of the S&P 500 index, which displays complicated, noisy, and volatile behaviour as illustrated in Figure 3. Comparative analysis of Figure 3(a) is S&P 500 closing price, Figure 3(b) is 100 day moving averages, and Figure 3(c) is 200 day average price.

<table>
<thead>
<tr>
<th>Technique used</th>
<th>RMSE</th>
<th>MAE</th>
<th>Linear correlation (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>1.12</td>
<td>0.9</td>
<td>0.34</td>
</tr>
<tr>
<td>SVM</td>
<td>0.97</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.42</td>
<td>0.09</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 1. Comparative analysis

Figure 2. Comparative performance analysis; (a) RMSE, (b) MAE, and (c) linear correlation

Figure 3. Comparative analysis; (a) S&P 500 closing price, (b) 100 day moving averages, and (c) 200 day average price
Stock market index prediction based on market trend using LSTM (Ankireddy Yenireddy)

This horizontal axis and vertical axis of the blue curve represent the original closing price time series from 2010 to 2022. The short- and long-term patterns of the closing price, the 100-day and 200-day moving averages. Despite a number of anomalies, the closing price is often moving upward as shown in Figure 3. The initial closing price and forecasts from the replication using the top LSTM model with the lowest RMSE score are shown in Figure 4. Nearly all of the real closing price’s curve in Figure 4(a) and the prediction curve’s closing price in Figure 4(b) for test data overlap. The best model may be able to almost accurately discover the initial closing price moved up and down. The even these unique market conditions, when there is a rapid substantial market collapse followed by a strong V-shaped rebound, the model is well fitted.

![Figure 4](image)

Figure 4. Comparative graph (a) true closing price and (b) predicted closing price

From above results it is stated that, described stock market index prediction based on market trend using LSTM is efficient than other models in terms of performance parameters. We found that stock market index prediction based on market trend using LSTM correlates with other techniques as LR, the SVM. The proposed method in this study is efficient than other models in terms of performance parameters. Our study demonstrates that stock market index prediction based on market trend using LSTM are more resilient than machine learning model-based prediction.

5. CONCLUSION

In this paper, stock market index prediction based on market trend using LSTM is described. For equities traders and private investors, the topic of stock price predicting is really interesting. It is difficult to predict prices accurately and consistently due to the noisy and nonlinear character of stock price movement. Using a combination of distributed input variables that cover all market and economic dimensions, this analysis focuses on developing LSTM-based models to predict the S&P 500 index closing price. The preprocessing of the input dataset from Yahoo Finance was successful. From comparative performance analysis, it is clear that low RMSE and MAE with highest linear correlation are obtained for LSTM model. Nearly all of the real closing price’s curve and the prediction curve’s closing price for test data overlap. Our findings provide conclusive evidence that this phenomenon is associated with less RMSE and MAE with highest linear correlation, as a result, the given model fits the data well even under typical market conditions, such as when there is an unexpectedly large decrease in the market. Future studies may explore combination of deep learning models with feasible ways of producing efficient predictive model.

REFERENCES


**BIographies of Authors**

Dr. Ankireddy Yenireddy received B.com degree from Acharya Nagarjuna University, MBA degree from JITUK, and Ph.D. degree from Acharya Nagarjuna University. He worked as Assistant Professor in AM Reddy memorial college of engineering and technology, Narasaraopet. He worked as Professor and Head of the Department in Loyola institute of technology and management at Sattenapalli, Guntur District, Andhra Pradesh, India from 2010 to 2023. He is currently working as Assistant Professor in department of MBA, Narasaraopeta Engineering college (Autonomous), Narasaraopet. He has supervised and co-supervised more than 100 PG Scholars. He has authored 2 text books in the area of Management and MIR with reputed publication houses. He has several National, International and SCOPUS publications. He has participated and organized conferences, workshops, and faculty development programs. He can be contacted at email: ankiyreddy.yenireddy74@gmail.com
Stock market index prediction based on market trend using LSTM (Ankireddy Yenireddy)

Dr. Marimganti Srinivasa Narayana received B.com degree from Acharya Nagarjuna University and M.com degree from Acharya Nagarjuna University and the Ph.D. degree from Acharya Nagarjuna University. He has done MBA (Marketing) from DR B. R. Ambedkar Open University and MBA (Human Resource Management) from Acharya Nagarjuna University. He has also done MSC (Psychology) from Dr BR Ambedkar Open University. He worked as Lecturer in TJPS College (PG courses), Guntur and worked as Principal, Director (PG Courses) and Head of the Department in various management departments in Guntur and Krishna Districts, Andhra Pradesh, India from 2004 to 2023. He is currently working as Professor in CDOE department (Centre for Distance and Online Learning of KL University, Vijayawada. He has supervised and co-supervised more than 20 Ph.D. Scholars. He has authored 3 text books and contributed lessons to various subjects of Management of different Universities in Andhra Pradesh and Telangana. He is a member of Board of Studies of 5 autonomous colleges for their MBA programs. Few of his papers received best papers award in international conferences and national conferences. He has 17 Scopus publications and few of them are H-indexed. His areas of research include marketing management and human resource management. He has chaired many sessions in National conferences. He can be contacted at email: dr.srinivasanarayana@gmail.com.

Dr. Kalla Venkata Bangaru Ganesh has done his Ph.D. in Commerce and Management from Krishna University, A.P. Master in Business Administration with Human Resource and Marketing as a dual specialization A.P. He has more than 15 years of teaching experience at PG level in various prestigious institutes and Universities. He is presently working as Assistant Prof. in KL Business School, Koneru Lakshmaiah Education Foundation (Deemed to be university), Vijayawada, Andhra Pradesh (Category1 University by UGCA++ Accreditation by NAAC). He has published several research papers in International and National reputed Journals. He has attended and presented papers in various International and National seminars and conferences. He also attended several FDP’s and STTP’s organized by reputed Institutions. He has authored two books and editor for two books in the area of Management and BRM with reputed publication houses. He is also associate with several international journals as editorial board member and reviewer. He can be contacted at email: kvbganesh@kluniversity.in.

Dr. Guvvaladinne Prasanna Kumar is working as a professor in the department of Electronics and Communication Engineering in Malla Reddy Engineering College (Autonomous) Hyderabad, Telangana, India. He has 22 years of teaching experience. He holds a Ph.D. Degree in ECE with specialization in Low Power VLSI from Sri Satya Sai University of Technology and Medical Science, Madhya Pradesh. He has authored and co-authored more than 18 publications. He received 10 different awards. He received 30 NPTEL certificates. He filed 4 patents. He published 2 book chapters. He can be contacted at email: prasanna4600@gmail.com.

Dr. Madduri Venkateswarlu working as associate professor in the Department of Electronics and Communication Engineering at Potti Sriramulu Chalavadi Mallikarjuna Rao College of Engineering and Technology (PSCMRCET), Vijayawada, Andhra Pradesh, India. He has 15 years of Teaching experience and 15 years of Industrial experience. He can be contacted at email: pragna2004@gmail.com.