International Journal of Recent Innovations in Academic Research
E-ISSN: 2635-3040; P-ISSN: 2659-1561
Homepage: https://www.ijriar.com/
Volume-5, Issue-4, April-2021: 29-40

Research Article

Analysis of Big Data for the Financial Sector Using Machine Learning Perspective on Stock Prices

[a]Hemanth Kumar Gollangi, [b]Sanjay Ramdas Bauskar, *[c]Chandrakanth Rao Madhavaram, [d]Eswar Prasad Galla, [e]Janardhana Rao Sunkara and [f]Mohit Surender Reddy

[a]KPMG, Consultant; [b]Pharmavite LLC, Senior Database Administrator; [c]Infosys, Technology Lead; [d]Infosys, Senior System Engineer; [e]Siri Info Solutions Inc., Senior Oracle Database Administrator; [f]Motorola Solutions, Senior Network Engineer
*Corresponding Author Email: chandrakanthmadhavaram@gmail.com

Received: April 04, 2021 **Accepted:** April 20, 2021 **Published:** April 28, 2021

Abstract
The stock market is one example of a segmented information economy. As a measure of both market activity and market uncertainty, price volatility is a key feature of the stock market. Investors may have a better grasp of market dynamics and create more evidence-based investing strategies with the aid of stock price volatility prediction and analysis made possible by the ever-expanding digitalization and BD (big data) technologies in the financial sector. Accurate financial stock prediction is of great interest to investors. The research focuses on key financial metrics like open, high, low, and close prices and trading volume, with data preprocessing techniques like wavelet denoising and standard scaling employed to improve model performance. This research delves into the use of ML, particularly LSTM networks, to forecast stock values with the use of Yahoo Finance's daily time series data. The LSTM model achieved an R^2 of 98.2%, demonstrating strong predictive accuracy, though the RMSE suggests room for improvement in reducing prediction errors. A comparison with linear regression showed that LSTM captures complex market trends more effectively. Future research could enhance this approach by incorporating a wider range of market indicators, exploring alternative deep learning architectures, and integrating multi-source data for more comprehensive stock price prediction and improved model robustness across varying market conditions.
Keywords: Big Data, Stock Market Forecasting, Price Prediction, Financial Markets, Risk Assessment, Yahoo Finance, Machine Learning.

I. Introduction

In recent years, the integration of big data into stock market analysis has become increasingly important for companies and industries to align their business strategies. The stock market, one of the oldest platforms for earning returns from firms, now generates massive volumes of data from trades, price fluctuations, and various economic indicators [1]. Big data analytics allows for more sophisticated and real-time analysis of this information, offering insights that were previously unattainable with traditional methods [2, 3]. An average individual could trade stocks, make investments, and earn money from companies that sold a portion of themselves on this platform. Nowadays, almost all major economic transactions take place on the stock market, where the value of stocks fluctuates as the market finds equilibrium [4]. If implemented properly, this technique has the potential to be a profitable investment strategy. The term "stock market forecasting" refers to the process of making predictions about the movement and performance of financial market indices, individual securities, or stock prices via the use of different models and methods of analysis [5, 6]. To make informed predictions about future market movements, it comprises assessing historical price and volume data in addition to other crucial elements including economic indicators, business financials, news events, and market sentiment [7, 8]. Stock market forecasting is a complex and challenging task due to the dynamic and unpredictable nature of financial markets. Nonetheless, it is an essential task for traders, analysts, investors, and financial institutions that want to properly manage risks and make informed decisions.

Time series forecasting is essential for managing portfolios, assessing risk, and making investment decisions in the stock market. Financial institutions, traders, and investors can't make informed judgements or make the most of their profits without accurate stock price and market trend forecasts. Conversely, the intricate and ever-changing nature of financial data makes stock market value forecasting a challenging task. The modern method for forecasting stock values is based on sophisticated AI algorithms that draw on fundamental or technical analysis [9]. The use of ML and DL to forecast stock prices and monitor their pattern fluctuations has also grown in popularity [10].

A. Motivation and Contribution of Study

Big data analytics' increasing importance in the financial industry, especially in stock market forecasting, is what inspired this research. Stock prices exhibit complex patterns influenced by various macroeconomic and market-driven factors, and traditional analytical methods often fall short of capturing these intricate dynamics. With the rapid advancements in machine learning and the availability of large-scale financial data, there is an opportunity to develop more accurate and data-driven prediction models. For investors, financial institutions, and governments to make wise choices and reduce risks, accurate stock price forecasts are essential. This study makes several key contributions to the field of financial analytics. The following key contributions are:

- ✓ Utilizes real-time stock market data sourced from the Yahoo Finance API, providing daily time series data.
- ✓ Uses standard scaling to normalise the dataset, ensuring that all features are scaled to zero mean and unit variance, which improves the performance of ML models.
- ✓ Compares the performance of ML models LIR and LSTM, providing a comprehensive analysis of their effectiveness in stock price prediction.
- ✓ Evaluate model performance using key metrics such as MAE, RMSE, and R^2. This ensures a thorough and multi-dimensional assessment of the models' accuracy and predictive capabilities.

B. Organization of the Paper

This research is structured as follows for parts that follow: Section II presents the background research on stock price prediction in the financial sector. Section III provides the research approach that is utilised for this study. Section IV covers the outcomes and assessments of the study. Our research study findings and plans for the future form Section V.

II. Related Work

In the literature, there is a lot of research on predicting stock prices. Some of them are listed below. This study, Kalra and Prasad, [11] focusses on tracking changes in stock prices in relation to pertinent corporate news pieces. This research proposes a daily prediction model to forecast the movements of the Indian stock market using historical data and news items. The NBC is used to classify news texts that have a positive or negative sentiment. The number of news stories with positive and negative sentiment for each day, the variation of the closing prices of the previous days, and historical data are utilised to make predictions. Using a variety of machine learning approaches, an accuracy ranging from 65.30 to 91.2% is attained [11].

In this study, Gumelar et al. [12] carried out a test to forecast the closing stock prices of 25 companies. These chosen businesses are formally listed on the Indonesia Stock Exchange (IDX) to guarantee data accuracy and regional concept. Extreme Gradient Boosting (XGBoost) and LSTM, two ML algorithms renowned for their excellent prediction accuracy from a variety of sample data, were used in this experiment. We were able to provide a trading strategy by establishing two thresholds: when to purchase and when to sell. There are several advantages to this forecast result from the ML algorithm used in the subsequent trading strategy. XGBoost performed best in this trial, with a prediction accuracy of 99% [12].

In this paper, Mootha et al. [13] provide a system that uses a Bidirectional LSTM based Sequence to Sequence Modelling approach to forecast a stock's future open, high, low and close (OHLC) value. Multitask learning aids in mapping the relationships between each OHLC price, which is a separate series. Additionally, a multitasking system that models pricing via shared tasks and subtasks is suggested. The NSE of India's Tata Consumer Products Limited stock prices are utilised. The suggested solutions are examined against different machine learning algorithms in order to assess their effectiveness. The suggested Seq2Seq and multitask systems perform noticeably better than the current techniques, with corresponding RMSE values of 3.98 and 7.87 [13].

In this work, Majumder et al. [14] have forecasted Bangladeshi stock indices using a variety of stock prediction algorithms, including Holt-Winter, Linear model, ARIMA, and FFNN, and evaluated the algorithms' performance across 35 Bangladeshi stocks. This study takes time series analysis into account, and the algorithms' performance is calculated by measuring the proportion of correct predictions. Based on study, the best algorithm for stock index forecasting is FFNN, whereas ARIMA (1,0,0) has the highest prediction accuracy (82.1%) on average. Maximum accuracy is provided by FFNN in 14 of 35 stocks [14].

In this paper, Song et al. [15] stock prices are gathered from a financial website, and internet comments are also mined and used to assess investor sentiment towards certain equities. The price of a stock fluctuates and trends, respectively, and investor feelings towards a particular stock are reflected in two types of time series. To assess the relationship between stock price and investor sentiment time series, we suggest Multiple Dimensional DCCA, which takes use of DCCA's strengths in time series analysis. Then, to improve price prediction, we provide a method to assess the range of effects of investor mood on stock price trends. Using the affecting period to forecast stock prices improves accuracy to around 85%, according to experiments [15].

In this work, Aradi and Hewahi [16] an approach is suggested for forecasting the direction and value of a market's movement based on a number of datasets, such as public opinion, business earnings reports, social media, and technical indicators. The use of LSTM and DNN in a case study of Apple Inc. shares made use of AI. The results demonstrated that the LSTM model outperformed the DNN model, which had prediction lag and relied on trend indicators to achieve a classification accuracy of 53.1%, in predicting the direction of the stock and in predicting the value of the stock (75.4 MSE, +-2.52 PE) [16].

Table 1 highlights key information from each of the studies regarding stock price prediction using various machine-learning techniques.

III. Methodology

The research design for the analysis of big data for the financial sector using a machine learning perspective on stock prices focuses on the systematic collection, pre-processing, modelling, and evaluation of stock market data. The dataset is obtained through the Yahoo Finance API, leveraging daily time series data that includes key financial metrics like open, high, low, and close prices and trading volume. A data flow in various steps and phases that shown in Figure 1.

After data collection, proceed to preprocessing. Preprocessing involves wavelet denoising for noise reduction and standard scaling for feature normalisation. Then, feature extraction is performed to derive essential signals, while a train-test split is used to evaluate model generalisation. The study compares ML models like LR and deep learning models like LSTM to forecast stock prices, with performance evaluated employing metrics like MAE, RMSE, and R^2. This design ensures a robust framework for analysing and predicting stock price movements by leveraging advanced machine-learning techniques.

Table 1. Background summary of financial sector using ML perspective on stock prices.

Author	Source	Techniques	Key findings	Limitation/future work
Kalra and Prasad [11]	Indian stock market data and news articles	Naïve Bayes (for sentiment analysis)	Accuracy ranged from 65.3% to 91.2% using different techniques	Future work: Could explore other sentiment analysis techniques or incorporate more diverse datasets
Gumelar et al. [12]	Stock data of 25 companies from Indonesia Stock Exchange (IDX)	LSTM, XGBoost	XGBoost achieved 99% prediction accuracy	Limitation: Focused only on 25 companies Future work: Could apply to more companies or markets
Mootha et al. [13]	Tata Consumer Products Limited (NSE, India) stock prices (OHLC)	Bidirectional LSTM, Seq2Seq modeling, Multitask learning	RMSE: 3.98 (Seq2Seq) and 7.87 (Multitask) Outperformed other ML algorithms	Limitation: Applied to a single stock Future work: Could extend the model to multiple stocks or other markets
Majumder et al. [14]	Bangladeshi stock market indices (35 stocks)	FFNN, ARIMA, Linear model, Holt-Winter approaches	ARIMA (1,0,0) gave 82.1% accuracy (average) FFNN performed best for 14 out of 35 stocks	Limitation: ARIMA accuracy varies by stock Future work: Could test more advanced models or use external factors for prediction
Song et al. [15]	Stock prices from a financial website and online investor comments	DCCA, Multiple dimensional DCCA (time series analysis)	Improved prediction accuracy to 85% by incorporating investor sentiment	Limitation: Dependency on investor sentiment Future work: Explore more complex sentiment measures or expand the dataset
Aradi and Hewahi [16]	Apple Inc. stock data (news sentiment, social sentiment, earnings, technical indicators)	LSTM, Deep neural networks (DNN)	LSTM: MSE of 75.4, classification accuracy of 70.1% for direction prediction DNN: Accuracy of 53.1% for movement direction	Limitation: DNN struggled with prediction lag Future work: Explore hybrid models or feature engineering for better results

Each step of the data flowchart is briefly explained below:

A. Data Collection

An essential first step in each project is data collecting, which is why this module is so fundamental. The selection of an appropriate dataset is a common theme. Prediction dataset sourced from Yahoo Finance API. Time series data is accessible via the API on an intraday, daily, weekly, and monthly basis. We choose to use daily time series data, which contains the following: daily volume, daily high price, daily low price, daily closing price, and daily open price (Figures 2 and 3), as our domain is short-term prediction.

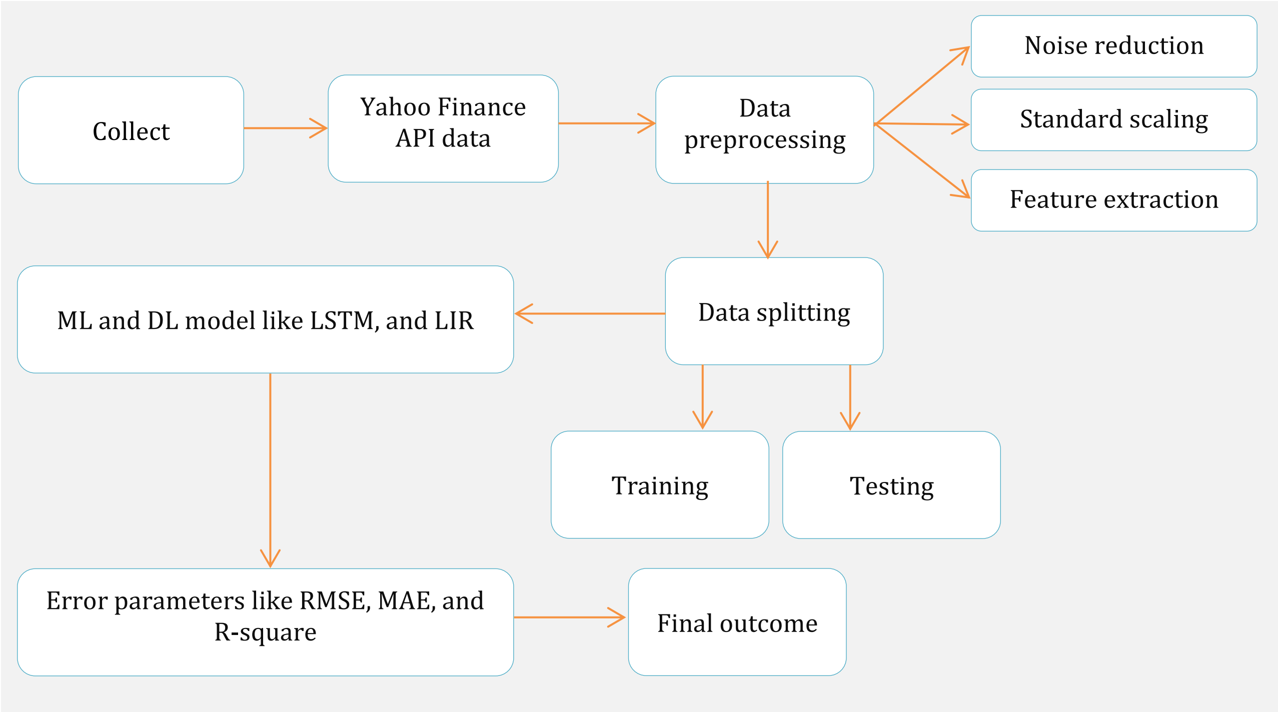

Figure 1. Data flow diagram for stock price prediction in the financial sector.

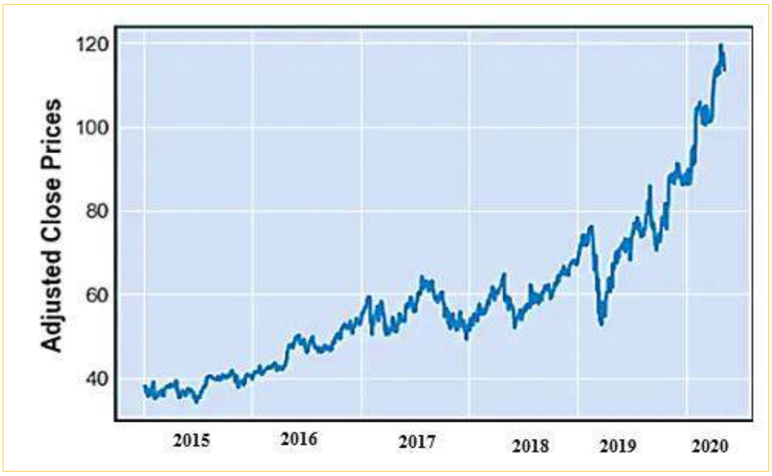

Figure 2. Line graph for adjusted close prices.

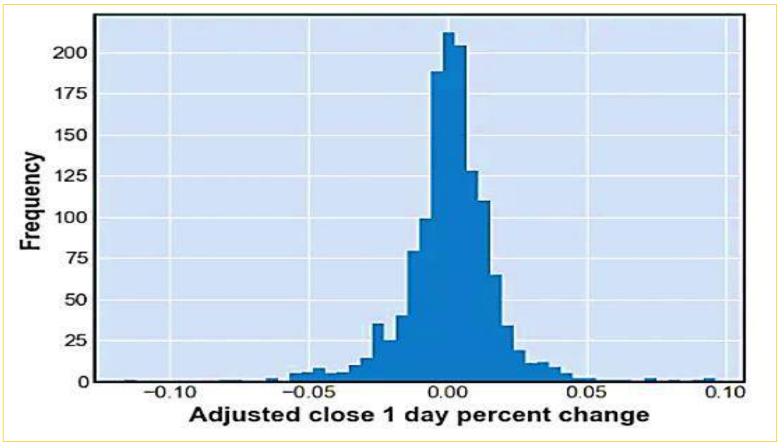

Figure 3. Histogram for 1-day per cent change.

The line graph shows the adjusted close prices of a financial asset from 2016 to 2020, with an overall upward trend. With growth peaking around 2017, followed by a decline in 2018-2019 and a recovery in 2020. The prices exhibit significant volatility throughout the period (Figure 2). The histogram shows a roughly bell-shaped distribution of daily percentage changes in the adjusted close price of a financial asset centred around 0. It is slightly skewed to the right, indicating occasional large positive changes, and displays leptokurtosis, suggesting a higher likelihood of extreme price changes. However, the relatively narrow range of changes indicates that the asset's price volatility is contained, with some downside risk but overall moderate fluctuations (Figure 3).

B. Data Preprocessing

The preparation of a dataset for ML tasks requires preprocessing. Several processes are involved in this process to clean, transform, and format the data so the model can learn successfully. Denoising is the first stage of preprocessing; it seeks to eliminate noise and incorrect or unnecessary data points that can obstruct the patterns necessary for efficient model learning. Further preprocessing steps are as:

i) Noise Reduction

To mitigate this issue, implemented wavelet denoising. The method's proficiency in both noise reduction and feature extraction led to its selection [17].

ii) Standard Scaling

Machine learning also makes use of the standard scaler, often known as standardisation, to scale features. This technique standardises all features by transforming them to a mean with zero variation. While this approach does not limit the data to a certain period or change its spread, it does ensure that most data points will be located around 0. This indicates that no matter how much data is scaled, outliers will remain.

As seen in equation 1, standard scaling is defined.

$$x_{scaled} = \frac{x-\bar{x}}{\sigma} \quad (1)$$

Where: xscaled = scaled sample point; x = sample point; $\bar{x}$ = mean of the training samples; σ = standard deviation of the training samples.

iii) Feature Extraction

The goal of feature extraction is to fulfill human intent by retrieving task-specific information from signals. A number of formats are available for feature extraction, including amplitude measurement, peak power, spectral density, Hjorth parameters, and others. Both univariate and multivariate feature extraction need substantial processing resources and mathematical analysis.

C. Train-Test Split

To evaluate the model's performance and make sure it generalises effectively to unknown data, it is essential to split the dataset into training and test sets.

D. Model Selection

Various methods can be used to predict the stock prices. In this work, utilise machine and deep learning models: LR, and LSTM explained below:

i) Long Short-Term Memory (LSTM)

The learning of distant nodes' data parameters is challenging in conventional RNNs because of issues with disappearing and expanding gradients. LSTM is an upgraded model that this research uses. Figure 4 shows that LSTM's memory function allows it to learn over the long term, identify characteristics, and correlate data from time series.

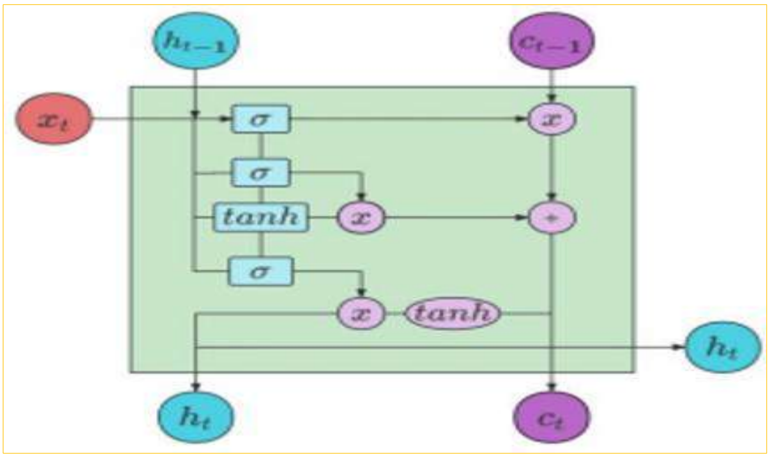

Figure 4. LSTM structure.

The structure of any recurrent neural network is a series of modules that repeat themselves. While RNNs in the past have had relatively basic structures for their modules, LSTM networks have four layers of complicated architecture that interact in unique ways. An LSTM repeat module computation involving a single neuron consists of two steps: updating the state of the neural network and calculating the output value. The input gate, the forgetting gate, and the output gate are the three gate functions found in a neuron. The gate function regulates the values of the input, memory, and output [17].

The quantity of data that is now being overlooked by the neural network is controlled by the forgetting gate. Below is the computation technique for the forgetting gate (2):

$$f_t = \sigma(W_f . [h_{t-1}\ x_t] + b_f \quad (2)$$

f_t represents the forgetting gate's output, ht-1 denotes the hidden state at the final second, and the fraction of information forgotten is modulated by the information fusion Wf, bf, and the sigmoid function of σ.

There are two components to the input gate: the original input value and the new input (3, 4):

$$i_t = \sigma(W_i . [h_t, x_t] + \ b_i) \quad (3)$$

$$\tilde{c}_t = \tanh\ (W_c . [h_{t-1}, x_t] + b_c) \quad (4)$$

The calculating procedure of the input gate is as follows (5), and it filters the information from the input layer when the output value of the tanh function is between -1 and 1.

$$C_t = f_t . C_{t-1} + i_t . C_t \quad (5)$$

Where the current state of the neural network is represented by Ct. Here is the procedure for calculating the hidden state and output gate (6, 7):

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t . \tanh\ (C_{t)} \quad (7)$$

ii) Linear Regression (LIR)

Fitting a collection of characteristics with their corresponding variables employing a linear equation is the goal of linear regression. The following equation (8, 9) describes the linear connection between a set of variables X and a set of responses/labels y:

$$X = \begin{bmatrix} x_1^1 & \dots & x_1^m \\ \vdots & \ddots & \vdots \\ x_1^m & \dots & x_n^m \end{bmatrix}; Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (8)$$

$$y = \alpha + \beta X \quad (9)$$

In this case, α stands for the slope and β for the intercept. A good match between the variables is achieved by minimising these parameters.

E. Model Evaluation

To evaluate the model performance, use some performance matrix. The ML models' predicting performance is evaluated using MAE, RMSE, and R^2.

i) MAE: The most typical applications of MAE are in regression problems using loss functions and error measurements, but it is also used to transform learning issues into optimisation problems. The following Eq. (10) of MAE is:

$$MAE = \frac{\sum_{1=1}^{n} |Y_i - X_i|}{n} \quad (10)$$

Here, Yi stands for the forecast, Xi for the actual value, and n for the overall count of records or samples.

ii) RMSE: The RMSE is one of the most popular ways to measure how well a forecasting model performs. It uses the distance from the real value to the observed value to display how much. A discrepancy between the predicted and observed values for every given sample is taken into account. The following Eq. (11) of RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{X}_i)^2}{n}} \quad (11)$$

iii) R-square: R^2 indicates the goodness-of-fit for linear regression models. The R^2 statistic shows how well your model takes into account the dependent variable. Calculating the amount of variation in y that can be explained by x-variables is done using the R^2 formula. You may find values between zero and one on the scale. The following Eq. (12) represent the R^2.

$$R^2 = 1 - \frac{SR}{TR} \quad (12)$$

Where, SR stands for sum of square residuals and TR for total square sum.

IV. Result Analysis and Discussion

The experiment results of ML models that utilised for stock price prediction are presented in this section. The following results are implemented on the Yahoo finance dataset across the performance matrix, including RMSE, MAE, and R-square. Table 2 shows the results of the LSTM model.

Table 2. Results of the LSTM model for stock price forecasting on Yahoo finance data.

Matrix	LSTM
R^2	98.2
RMSE	2.02
MAE	32.283

The performance metrics for the LSTM model show a high R^2 value of 98.2%, indicating strong predictive accuracy. The RMSE is 2.02, suggesting that the model's predictions deviate from the actual values by an average of approximately 20 units. The MAE is 32.283, reflecting the average absolute difference between predicted and actual values. In sum, these measures point to the LSTM model's strong performance, with very few prediction errors and a high degree of accuracy.

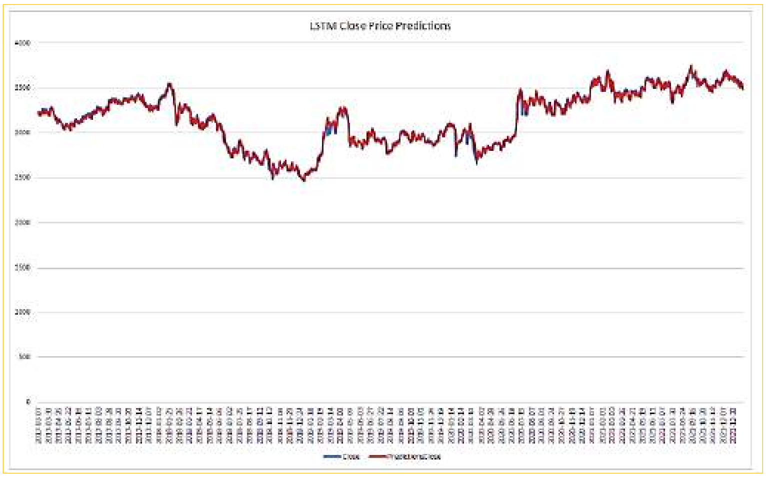

Figure 5. LSTM model close price prediction.

Figure 5 displays LSTM close price predictions comparing actual close prices with the close prices predicted by the LSTM model. The actual close prices are shown in blue, and the predicted prices in red. Indicating that the LSTM model does a good job of following the overall trend of the actual closing prices, the two lines closely track each other. There are minor deviations at certain points, but overall, the predictions seem to align well with the actual data. This suggests the LSTM model performs reasonably in forecasting close prices, though further analysis with evaluation metrics would help confirm the accuracy.

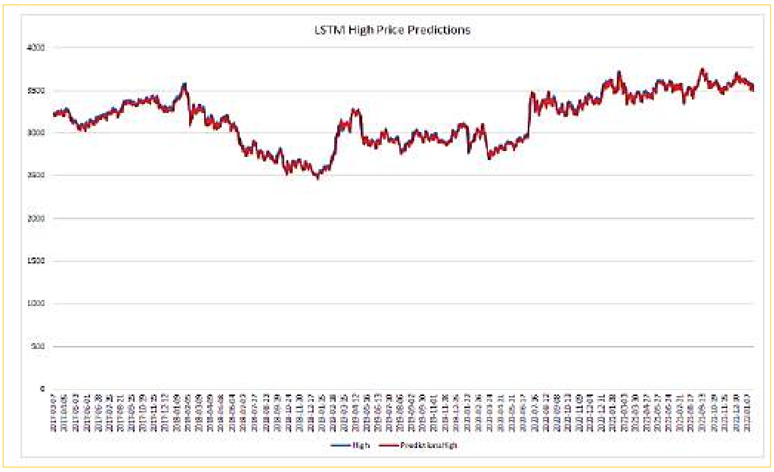

Figure 6. LSTM model high price prediction.

Figure 6 shows LSTM high price predictions comparing actual high prices with LSTM model-predicted high prices. The actual high prices are represented by a blue line, while the predicted values are in red. The two lines appear closely aligned indicating that the LSTM model captures the general trend of the high prices with reasonable accuracy. Some slight deviations are visible, but overall, the model seems to track the price movements well over the displayed time period. Further evaluation metrics would be helpful for a more detailed analysis of model performance.

Table 3. Comparative analysis of model performance on Yahoo finance data.

Matrix	LSTM	LIR [18]
R^2	98.2	91.43
RMSE	2.02	1.145

Table 3 shows the comparative analysis of the model's performance. The performance metrics comparison between the LSTM model and LIR indicates that the LSTM model demonstrates superior predictive accuracy with an R^2 of 98.2%, significantly higher than LIR's 91.43%. Additionally, the LSTM has an RMSE of 2.02, which is higher than LIR's RMSE of 1.145. This implies that although the LSTM successfully captures the data's volatility.

V. Conclusion and Future Work

Investors rely heavily on stock price forecast when formulating a trading strategy. Investors may boost their profit margins via accurate stock price predictions. The interconnected nature of stock prices with variables beyond of investors' control, such as headlines, the state of the economy, public opinion, and other confidential financial data, makes accurate trend forecasting in the stock market very challenging. This research employs a unique approach that combines deep interest in historical market data with the latest DNN technology for linear regression and time series prediction LSTM.

The LSTM model demonstrated superior predictive accuracy for stock price forecasting, with a high R^2 of 98.2%, indicating that it effectively captures stock price trends based on the Yahoo finance data. Despite its strong performance, the model's relatively higher RMSE compared to linear regression suggests that while LSTM captures complex patterns, there is room for improvement in reducing prediction errors.

One limitation of this study is the reliance on a single dataset and specific time intervals, which may not generalise well across different market conditions or financial sectors. Future work could explore the inclusion of more diverse datasets, additional market indicators, and alternative deep learning models like transformer-based architectures to enhance prediction robustness and accuracy.

Declarations

Acknowledgments: We gratefully acknowledge all of the people who have contributed to this paper.

Author Contributions: All authors have contributed equally to the work.

Conflict of Interest: The authors declare no conflict of interest.

Consent to Publish: The authors agree to publish the paper in International Journal of Recent Innovations in Academic Research.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Research Content: The research content of manuscript is original and has not been published elsewhere.

References

1. Nichante, V. and Patil, S. 2016. A review: Analysis of stock market by using big data analytic technology. International Journal on Recent and Innovation Trends in Computing and Communication, 4(1): 305-306.

2. Shastri, M., Roy, S. and Mittal, M. 2019. Stock price prediction using artificial neural model: An application of big data. EAI Endorsed Transactions on Scalable Information Systems, 6(20): 1-8.

3. Pandey, S. 2020. Transforming performance management through AI: Advanced feedback mechanisms, predictive analytics, and bias mitigation in the age of workforce optimization. International Journal of Business Quantitative Economics and Applied Management Research, 6(7): 1-10.

4. Yarlagadda, V.K., Maddula, S.S., Sachani, D.K., Mullangi, K., Anumandla, S.K.R. and Patel, B. 2020. Unlocking business insights with XBRL: Leveraging digital tools for financial transparency and efficiency. Asian Accounting and Auditing Advancement, 11(1): 101-116.

5. Mullangi, K., Yarlagadda, V.K., Dhameliya, N. and Rodriguez, M. 2018. Integrating AI and reciprocal symmetry in financial management: A pathway to enhanced decision-making. International Journal of Reciprocal Symmetry and Theoretical Physics, 5(1): 42-52.

6. Chen, Y.J., Chen, Y.M. and Lu, C.L. 2017. Enhancement of stock market forecasting using an improved fundamental analysis-based approach. Soft Computing, 21: 3735-3757.

7. Pal, S.S. and Kar, S. 2019. Time series forecasting for stock market prediction through data discretization by fuzzistics and rule generation by rough set theory. Mathematics and Computers in Simulation, 162: 18-30.

8. Goyal, R. 2020. The role of business analysts in information management projects. International Journal of Core Engineering and Management, 6(9): 76–86.

9. Anumandla, S.K.R., Yarlagadda, V.K., Vennapusa, S.C.R. and Kothapalli, K.R.V. 2020. Unveiling the influence of artificial intelligence on resource management and sustainable development: A comprehensive investigation. Technology and Management Review, 5(1): 45-65.

10. Ghania, M.U., Awaisa, M. and Muzammula, M. 2019. Stock market prediction using machine learning (ML) algorithms. The Advances in Distributed Computing and Artificial Intelligence, 8(4): 97-116.

11. Kalra, S. and Prasad, J.S. 2019. Efficacy of news sentiment for stock market prediction. In: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 491-496). doi: 10.1109/COMITCon.2019.8862265

12. Gumelar, A.B., Setyorini, H., Adi, D.P., Nilowardono, S., Widodo, A., Wibowo, A.T., et al. 2020. Boosting the accuracy of stock market prediction using XgBoost and long short-term memory. In: 2020 international seminar on application for technology of information and communication (iSemantic) (pp. 609-613). doi: 10.1109/iSemantic50169.2020.9234256

13. Mootha, S., Sridhar, S., Seetharaman, R. and Chitrakala, S. 2020. Stock price prediction using bi-directional LSTM based sequence to sequence modeling and multitask learning. In: 2020 11th IEEE annual ubiquitous computing, electronics and mobile communication conference (UEMCON) (pp. 0078-0086). doi: 10.1109/UEMCON51285.2020.9298066

14. Majumder, M.M.R., Hossain, M.I. and Hasan, M.K. 2019. Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis. In: 2019 international conference on electrical, computer and communication engineering (ECCE) (pp. 1-5). doi: 10.1109/ECACE.2019.8679480

15. Song, C., Chen, W., Fu, L. and Arshad, A. 2020. Improving stock price prediction based on investing sentiments. In: 2020 IEEE 6th international conference on computer and communications (ICCC) (pp. 1639-1644). doi: 10.1109/ICCC51575.2020.9345303

16. Al Aradi, M. and Hewahi, N. 2020. Prediction of stock price and direction using neural networks: Datasets hybrid modeling approach. In: 2020 international conference on data analytics for business and industry: way towards a sustainable economy (ICDABI) (pp. 1-6). doi: 10.1109/ICDABI51230.2020.9325697

17. Agarwal, N., Gupta, S. and Gupta, S. 2016. A comparative study on discrete wavelet transform with different methods. In: 2016 symposium on colossal data analysis and networking (CDAN) (pp. 1-6). doi: 10.1109/CDAN.2016.7570878

18. Shakhla, S., Shah, B., Shah, N., Unadkat, V. and Kanani, P. 2018. Stock price trend prediction using multiple linear regression. International Journal of Engineering and Science Invention, 7(10): 29-33.

Citation: Hemanth Kumar Gollangi, Sanjay Ramdas Bauskar, Chandrakanth Rao Madhavaram, Eswar Prasad Galla, Janardhana Rao Sunkara and Mohit Surender Reddy. 2021. Analysis of Big Data for the Financial Sector Using Machine Learning Perspective on Stock Prices. International Journal of Recent Innovations in Academic Research, 5(4): 29-40.