ISSN 2007-9737

Integrating Multimodal Data for Enhanced Stock Market Trend Prediction: A Deep Neural Network and Regression Approach

Swarnalata Rath[1], Nilima R. Das[1], Binod Kumar Pattanayak[1], Ayasa Kanta Mohanty[2], Bibudhendu Pati[3,*]

Siksha 'O' Anusandhan (Deemed to be University),
Department of Computer Science and Engineering,
India

Institute of Business and Computer Studies,
Faculty of Management Sciences,
Siksha 'O' Anusandhan (Deemed to be University),
India

Rama Devi Women's University,
Department of Computer Science,
India

{nilimadas, binodpattanayakayasakantamohanty}@soa.ac.in,
{swarnalata.rath, patibibudhendu}@gmail.com

Abstract. The demand for accurate stock market trend prediction models has surged among financial traders, prompting the exploration of machine learning techniques to enhance predictive performance and reduce computational complexity. Traditional models often rely solely on historical stock data, which may not fully capture the intricate dynamics of financial markets. This research addresses this limitation by developing a model that utilizes Open, High, Low, and Close (OHLC) prices, integrating both classification and regression machine learning models to predict stock market trends. However, not all models are favorable for trend prediction and therefore, we compare using different models to find the best performing model. The regression models tested include: Long Short-Term Memory (LSTM): Achieved 99.31% accuracy, Linear Regression**:** Achieved 98.85% accuracy, Decision Tree**:** Achieved 98.22% accuracy, Stochastic Gradient Descent (SGD): Achieved 97.63% accuracy, Temporal Convolutional Networks (TCN)**:** Achieved 96.95% accuracy, K-Nearest Neighbors (KNN): Achieved 80.35% accuracy, Random Forest**:** Achieved 57.12% accuracy. The classification models evaluated include: Artificial Neural Networks (ANN): Achieved 97.28% accuracy, Stochastic Gradient Descent (SGD)**:** Achieved 93.65% accuracy, K-Nearest Neighbors (KNN)**:** Achieved 89.53% accuracy, XGBoost**:** Achieved 87.01% accuracy, Decision Tree: Achieved 86.1% accuracy, Random Forest**:** Achieved 85.06% accuracy, Support Vector Machine (SVM)**:** Achieved 61.94% accuracy, AdaBoost**:** Achieved 58.38% accuracy, Naïve Bayes**:** Achieved 51.4% accuracy. Here, it is observed that the regression models are performing better in trend prediction when compared with classification models. However, it has also been noticed that ANN in classification; LSTM, and Linear regression in regression are performing better than other models in their categories.

Keywords. Stock price trend prediction, decision tree, KNN, SGD, random forest, linear regression, LSTM, TCN, ANN, SVM, naïve Bayes, XG boosting, KNN, ada boosting.

1 Introduction

Today, the volume of data is increasing exponentially, with numerous factors to consider when predicting accurate results for uncertain, continuous data. A prime example of such data is the stock market. Stock price predictions rely on one of the most unreliable sources of information—human emotions. However, in stable market conditions, human emotions have a minimal impact on price trends. As many financial traders are interested in these prediction models, it becomes highly valuable to develop a reliable model for forecasting stock market prices

Predicting stock prices depends on numerous factors. Therefore, machine learning models can be employed to forecast price trends. These models can provide traders with a reliable reference to guide their investments in stable market conditions. News related to the stock's parent company can significantly influence market dynamics. However, one reason we can rely on Open, High, Low, and Close (OHLC) prices to predict future trends is that, in stable market conditions, there is often a correlation between consecutive days. Thus, OHLC data can be effectively used in predicting current price trends

In machine learning, there are various algorithms [1] that come with their own computational parameters and methods for processing data. However, not all available models provide the necessary accuracy or reliability for real-time use. As a result, we have explored different machine learning models [2] to predict stock market trends. We employed both neural [3-5] and non-neural network models using OHLC data, ultimately identifying the most effective models.

The classification models we used include ANN (Artificial Neural Networks), SVM (Support Vector Machine) [6] [7], Naïve Bayes, Decision Tree, Random Forest, XGBoost, KNN (K-Nearest Neighbors), AdaBoost, and SGD (Stochastic Gradient Descent). For regression, the models used are Decision Tree [9], KNN, SGD, Random Forest, Linear Regression, LSTM (Long Short-Term Memory) [10,11], and TCN (Temporal Convolution Networks) [12,13].

2 Literature Review

A literature review of various models for stock price prediction is essential before proceeding with further research on stock market trend forecasting.

In their article titled *"*Development of Stock Market Trend Prediction System Using Multiple Regression*"*, Muhammad Zubair Asghar et al. utilized factors such as change, volume, return, volatility, and return on investment in their multiple regression model. The accuracy achieved by their model was nearly 95% [1]. Therefore, this model can serve as a benchmark for comparison with the use of OHLC data.

Min wen et al. [2] in their article titled “Stock market trend prediction using high-order information of time series” showed that pictorial representation of the data patterns in a neural network can give an effective database for the ML model (CNN) to use and predict. The basic idea of this paper was that the trends of a time series can be reflected in the patterns of diverse motifs. In this article, they showed that the basic neural network model which was used only as a classification model can now be used to predict time series data by analysing the trends. The underlying patterns were reconstructed and applied to CNN.

M. Ananthi and K. Vijayakumar [3] in their article titled “Stock Market analysis using candlestick regression and market trend prediction (CKRM)” stated that they used candlestick pattern recognition to predict stock trends and investment decisions predict. They used the KNN regression model was used to analyze past price movements in order to make predictions. They also used linear regression along with SVM as test samples for the same dataset. As the candlestick patterns (pictorial representation of OHLC) were used in the prediction using KNN, SVM, and linear regression we can compare ones for our OHLC data.

Jingyi et al. [4], titled "Short-term stock market price trend prediction using a comprehensive deep learning system," presented a high-level architecture divided into three main components: First being “Feature selection”, the second being “Dimensionality Reduction” and the third being “Creating the prediction model”. In their research SVM, Random Forest, LSTM, Naïve Bayes, Logistic Regression, and Multi-Layer perceptron were compared. Their paper showed that the min-max scaling of data gave more accuracy. Hence Min-Max scaling can be used in the regression models in our model.

M. Nabipour et al. [5] in their article titled “Deep Learning for Stock Market Prediction” used several models, including Decision Tree, XG Boosting, Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM), to predict stock prices. As a result, these same models can be applied in regression tasks. However, there is one issue with their paper: comparing RNN and LSTM with ANN, which is

ISSN 2007-9737

Table 1. Sample of TATA Motors Stock Price Dataset

Date	Open	High	Low	Close	Adj_close	Volume
2000-01-03 00:00:00	43.5	43.5	43.5	43.5	16.11473655	0
2000-01-04 00:00:00	43.5	43.5	43.5	43.5	16.11473655	0
2000-01-05 00:00:00	43.5	43.5	43.5	43.5	16.11473655	0
2000-01-06 00:00:00	43.5	43.5	43.5	43.5	16.11473655	0
2000-01-07 00:00:00	43.5	43.5	43.5	43.5	16.11473655	0

Table 2 Sample of Mahindra & Mahindra Ltd historical stock data with monthly prices, volumes, open/close values

Date	Open	High	low	Close	Volume	Change percent
09-12-2023	3,619.20	3,658.00	3,670.20	3,606.90	2.11M	-1.70%
08-12-2023	3,681.70	3,720.00	3,725.00	3,671.60	1.50M	-0.95%
05-12-2023	3,717.10	3,659.30	3,721.50	3,642.50	2.39M	+1.24%
04-12-2023	3,671.60	3,675.00	3,678.20	3,620.20	1.98M	+0.61%
03-12-2023	3,649.40	3,736.00	3,740.00	3,640.00	2.00M	-1.81%

generally more accurate for classification tasks, is not advisable. It is recommended to use a comparison model that can predict as closely as an LSTM, but for a more basic model.

Chen et al. [7] in their article titled "A Deep Fusion Model for Stock Market Prediction with News Headlines and Time Series Data" introduced a multimodal deep learning approach for forecasting stock trends. This model combines three key data sources: Daily Stock Price, Technical Indicators, and Sentiment derived from News Headlines. The authors evaluated the model on 12 different stock datasets, showing that it outperforms traditional methods in terms of both prediction accuracy and trading outcomes.

N. Jamaludin et al. [16] proposed a multiple regression model to examine the relationship between inflation rate, exchange rate, interest rate, and stock prices. Their findings suggested that the exchange rate is a significant factor influencing stock price movements, while the interest rate has no substantial impact. The study concluded that approximately 54.2% of stock price movements can be explained by the considered external variables. However, the limited number of variables used in the model raises concerns about its overall reliability.

Kamley et al. [17] applied multiple regression using open price, close price, and high price to predict stock market prices. They performed necessary data pre-processing and transformations to prepare the data effectively. Their model achieved an accuracy rate of 89%. This approach motivates us to explore the use of basic OHLC (Open, High, Low, Close) data for making predictions. Additionally, classification models and neural network models, which incorporate multiple variables, can also be considered for prediction.

Yuan et al. [18] proposed a model which considers price, volume and position of the stocks and their movements. Their dataset consists of open price, lowest price, volume and holdings of the stocks. The implemented their model using decision tree which gave an accuracy of about 70%. However, they also showed that volume and position play fewer roles in price prediction. Hence from their model we need not consider volume data for our model building as volume is also considered as basic available data.

Javaid [19] used dividend prices, KIBOR, GDP, earnings, and inflation in multiple regression model for the stock market price prediction. His model gave an accuracy of about 62%. However, we can use other factors to gain more accuracy. With this, we can be certain that even though earnings might show some effect on stock price movement, we do not get good accuracy as humans as a whole do not usually consider looking into a company's current earnings and GDP. But the price depends mostly on the actions companies promise to take up in the near future and their reputation of the past events.

3 Proposed Methodology

OHLC data is an essential aspect of newly developed models, as the need for continuous development and modification is increasing.

To build an effective model, we utilize various machine learning techniques on pure OHLC data. For the computational implementation, we used Jupyter Notebook with Python 3, running on an Intel Core i5 8th Gen processor.

ISSN 2007-9737

The datasets used for training the model are from Tata Motors and MM(Mahindra and Mahindra), sourced from Kaggle [20].

The TATA Motors Stock Price Dataset provides historical stock price and trading data for TATA Motors Limited from January 3, 2000, to September 2, 2023.

It includes daily records of open, high, low, close prices, adjusted close prices, and trading volumes.

The Mahindra & Mahindra Ltd stock dataset contains monthly prices, volumes, open/close values and change percentages. Samples of both the data sets are given in Table 1 and Table 2 respectively.

3.1 Classification Models

We compared the following classification models to identify the best-performing model for OHLC data [32-35].

3.1.1 Artificial Neural Network (ANN)

Neural networks, by adjusting their own weights, are particularly helpful for making accurate stock market predictions and allow for efficient computation.

3.1.2 Stochastic Gradient Descent (SGD)

SGD is commonly used as an optimization technique for other algorithms to improve performance. We applied the SGD classifier directly to make predictions based on the OHLC data.

3.1.3 K Nearest Neighbours (KNN)

KNN makes predictions by identifying similarities in the given data. It proves to be a reliable method for predicting when new and recent data is provided.

3.1.4 XG Boosting

XG Boosting, a variant of decision trees, is effective for establishing relationships between OHLC data points and making accurate predictions based on those relationships.

3.1.5 Decision Tree

Decision Trees make predictions by establishing relationships between the attributes of the input data. They can be a reliable method for identifying patterns between OHLC (Open, High, Low, Close) data and making predictions based on that.

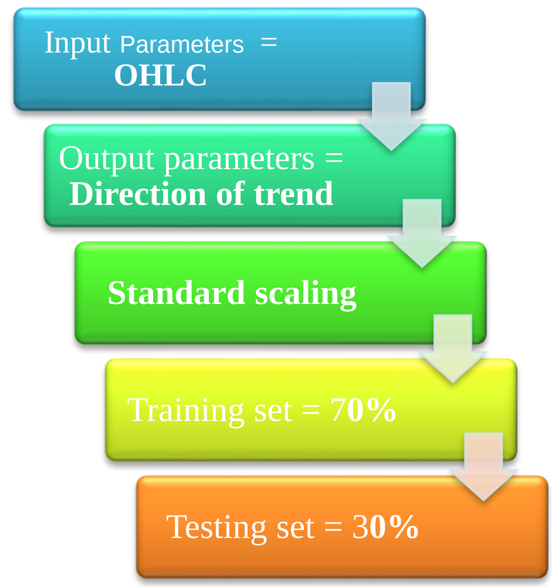

Fig. 1. Pre-Processing for Classification.

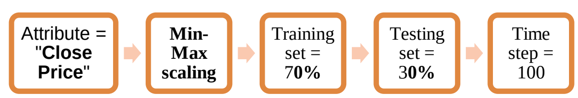

Fig. 2. Pre-Processing for regression

3.1.6 Random Forest

Random Forests predict outcomes by considering multiple decision trees formed from subsets of the input data. This method is useful for identifying relationships within OHLC data and making predictions based on those patterns.

3.1.7 Support Vector Machine (SVM)

SVM creates a boundary from the results of training data, which can be particularly useful in the stock market for accurate predictions and easy computation, helping to separate data points for classification.

3.1.8 AdaBoostin

AdaBoosting predicts outcomes by using binary data as output. This technique can form reliable relationships between OHLC data and forecast

ISSN 2007-9737

trends accordingly by focusing on the misclassified data in previous iterations.

3.1.9 Naïve Bayes

Naïve Bayes relies on Bayes' theorem, which uses probability. Since probability is a powerful tool for general data analysis, applying it in the stock market can lead to accurate predictions based on the likelihood of various events.

The following flow chart shows the pre-processing steps used for classification models.

The input data used for processing and model building consists of the OHLC data, while the output is the trend for the day, indicating whether it is up or down.

To make the data easier to compute, we applied Standard Scaling. Additionally, dividing the dataset into training and testing sets is crucial, so we split the data into 70% for training and 30% for testing.

3.2 Regression Models

We compared the following regression models to identify the best-performing model for OHLC data.

3.2.1 Long-Short Term Memory (LSTM)

LSTM is a variation of Recurrent Neural Networks used for time series prediction. It can adapt to sudden changes in the data, making it effective for achieving accurate results in predicting stock market trends.

3.2.2 Linear Regression

Linear Regression makes predictions by linearly adjusting the trained data. It is particularly useful for predicting stock prices in stable market conditions.

3.2.3 Decision Tree

A Decision Tree predicts outcomes by forming relationships between the attributes of the input data. It can be a reliable method for establishing connections between each closing price and previous patterns, making predictions based on that.

3.2.4 Stochastic Gradient Descent (SGD)

SGD is an optimization technique used to improve the performance of other algorithms. By applying SGD regression, we can optimize predictions using historical price data.

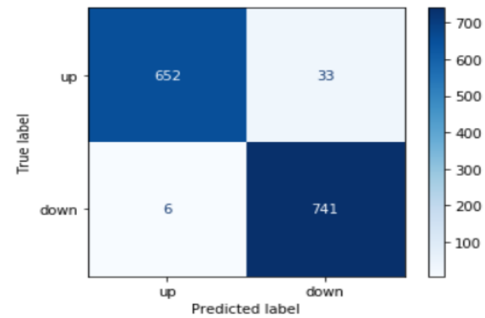

Fig. 2. Confusion matrix for Artificial Neural Networks

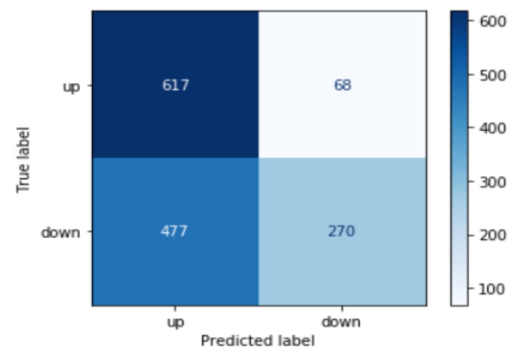

Fig. 3. Confusion matrix for Support Vector Machine

3.2.5 Temporal Convolutional Networks (TCN)

Temporal Convolutional Networks (TCN) are a newer variation of convolutional neural networks that show promise in accurately predicting stock price trends. They may yield strong performance when applied to our OHLC data.

3.2.6 K Nearest Neighbours (KNN)

KNN makes predictions by clustering the nearest neighbors in the data. It can be a reliable method for predicting trends when recent data is provided.

3.2.7 Random Forest

Random Forest makes predictions by forming relationships between decision trees built from subsets of the data. It can be effective in creating relationships between continuous closing data points to make predictions accordingly.

The following flow chart shows the pre-processing steps used for regression models.

ISSN 2007-9737

Here, Input data for processing and model building is 'Close Price' of the day. And output is the scaled data trend for the upcoming days. For scaling the data for easy computation, we used Min-Max Scaling. However, dividing the dataset for training and testing is important and therefore we divided the data into 70% of training data and 30% of testing data. Another parameter considered for regression is the time step. Time step is the number of days the current day trend depends on and it is default set to 100.

The dataset used for classification contains an attribute that indicates the direction of the trend for a specific day based on the trade. This helps the model draw conclusions about the trend for that day when the OHLC data is provided during classification training. For regression, the attribute used from the datasets is the closing price for all training, testing, and prediction.

The regression dataset includes OHLC data along with volume; however, the volume is deliberately ignored to achieve the desired results, as shown in Section 4. As a spoiler alert, the dataset must be large enough to generate proper results from any models used. As highlighted in Section 4.2, it is crucial to avoid falling into the trap of abnormal results. The same principle applies to humans: we learn better with enough examples. Having sufficient training data is far more beneficial than having limited knowledge of patterns. When the dataset used for training and testing is small, this is known as overfitting.

We consider our experience as a tool. In our calculations, we used **dataset1**, which contains 5306 records of data, and **dataset2**, which has about 3000 records for regression.

4 Results and Analysis

4.1 Classification Models

4.1.1 Artificial Neural Network (ANN)

The figure below shows the confusion matrix for the predicted accuracy of the Artificial Neural Network (ANN). The true label represents the actual values of the trend, while the predicted label shows the predicted trend values. The accuracy achieved by the ANN on the dataset is 97.28%. Out of the total

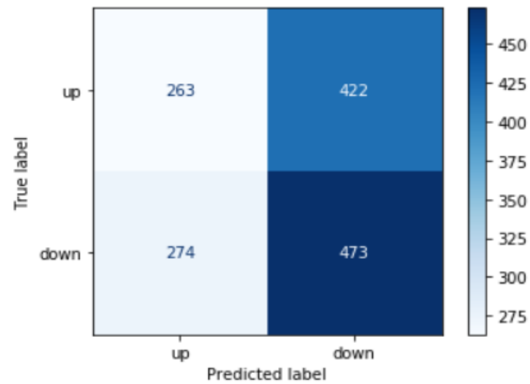

Fig. 4. Confusion matrix for Naïve Bayes

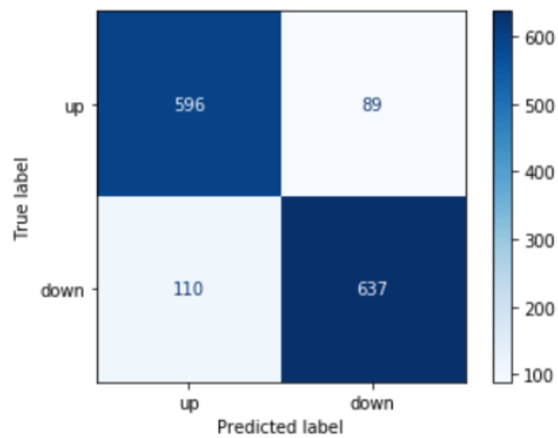

Fig. 5. Confusion matrix for Decision Tree Classification

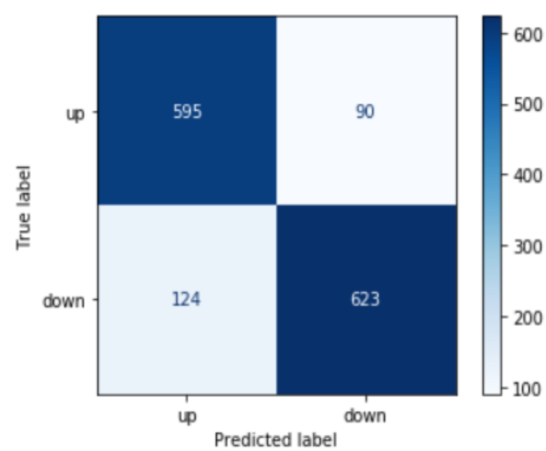

Fig. 6. Confusion matrix for Random Forest Classification

ISSN 2007-9737

1432 testing set, 1393 results were predicted accurately, and 39 results were inaccurate.

4.1.2 Naïve Bayes

The following figure is the confusion matrix of the predicted accuracy of Naïve Bayes. The true label gives the exact values of trend and the predicted label gives the predicted values of the trend.

The accuracy achieved by Naïve Bayes in the dataset is 51.4%. Out of the whole (1432) testing set 736 results are predicted accurately and 696 results are inaccurate.

4.1.3 Decision Tree

The confusion matrix of Decision Tree Classification below shows that 1233 out of 1432 results are predicted accurately and 199 results are wrongly calculated.

The true label gives the exact values of trend and the predicted label gives the predicted values of the trend. The accuracy achieved by Decision tree classification in the dataset is 86.1%.

4.1.4 Random Forest

The following figure is the confusion matrix of the, predicted accuracy of, Random Forest Classification. The true label gives the exact values of trend and the predicted label gives the predicted values of the trend.

The accuracy achieved by Random Forest Classification in the dataset is 85.06%. Out of the whole (1432) testing set 1218 results are predicted accurately and 214 results are inaccurate.

4.1.5 XG-Boosting (XGB)

The figure shown below is the confusion matrix of XGB Classification. The true label gives the exact values of trend and the predicted label gives the predicted values of the trend. The accuracy achieved by XGB in the dataset is 87.01%. Out of the whole (1432) testing set 1246 results are predicted accurately and 186 results are inaccurate.

4.1.6 K Nearest Neighbours

The accuracy achieved by KNN in the dataset is 89.53. The confusion matrix of SVM below shows that out of the whole (1432) testing set 1282 results are predicted accurately and 150 results are inaccurate. The true label gives the exact values of

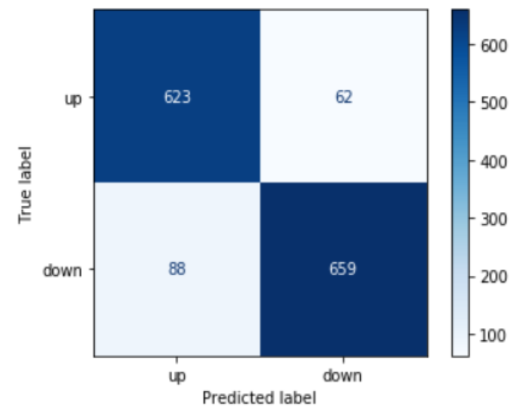

Fig. 7. Confusion matrix for K Nearest Neighbours Classification

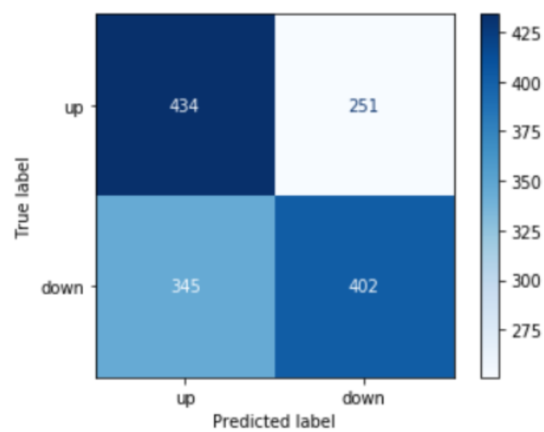

Fig. 8. Confusion matrix for Ada Boosting Classification

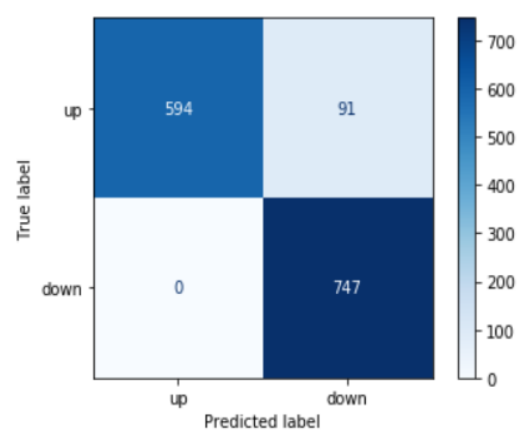

Fig. 9. Confusion matrix for Stochastic Gradient Classification

trend and the predicted label gives the predicted values of the trend.

4.1.7 Ada-Boosting

The following figure is the confusion matrix of the predicted accuracy of Ada-Boosting Classification. The true label gives the exact values of trend and the predicted label gives the predicted values of the trend. The accuracy achieved by Ada-boosting in the dataset is 58.38%. Out of the whole (1432) testing set 836 results are predicted accurately and 596 results are inaccurate.

4.1.8 Stochastic Gradient Descent Classification

The confusion matrix of SGD Classification below shows that 1341 out of 1432 are predicted accurately and 91 are wrongly predicted. The true label gives the exact values of trend and the predicted label gives the predicted values of the trend. The accuracy achieved by SGD classification in the dataset is 93.65%.

The following table gives the accuracy achieved by individual classification models. In our proposed model building strategy, we achieved the highest accuracy in Artificial neural networks which is 97.28%, the second is stochastic gradient descent which is 93.65%, third is K nearest neighbours which is 89.53%, fourth is XG-boosting which is 87.01%, fifth is Decision tree classification which is 86.1%, sixth is Random forest classification which is 85.06%, seventh is support vector machines which is 61.94%, eigth is Ada-Boosting which is 58.38%, and ninth is Naïve Bayes which is 51.4%. By this we can conclude that if classification models are being used for prediction we better rely on ANN and SGD for more accuracy.

4.2 Regression Models

represents the predicted transformed data. As the regression models demonstrated higher accuracy in predictions, considering another dataset with fewer data points could potentially result in even better-performing models.

Therefore, we considered this other dataset, and it revealed that LSTM, TCN, Linear Regression, and SGD provided more accurate results compared to the other models.

Table 3. Accuracy of the classification models

Classification Prediction Model	Accuracy Score
ANN	97.28%
SGD	93.65%
KNN	89.53%
XG Boosting	87.01%
Decision Tree	86.1%
Random Forest	85.06%
SVM	61.94%
Ada Boosting	58.38%
Naïve Bayes	51.4%

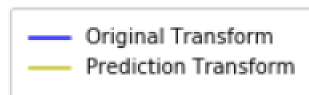

Fig.10. Regression Graphs Legend

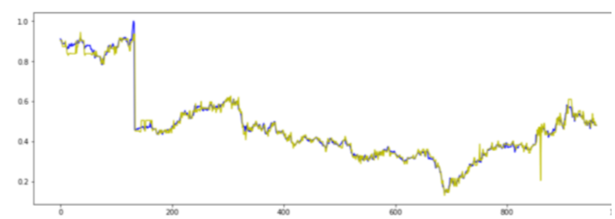

Fig. 11. Graph for Decision Tree

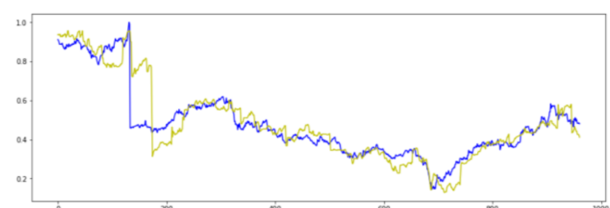

Fig. 12. Graph for K Nearest Neighbors Regression

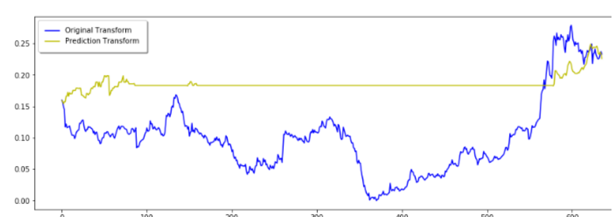

Fig. 15. Graph for K Nearest Neighbors Regression Dataset2

Table 4. Accuracy in Regression

Regression Prediction Model	R2 Score dataset1	R2 Score dataset2
LSTM	99.31%	98.01%
Linear Regression	98.85%	98.94%
Decision tree	98.22%	-54.63%
SGD	97.63%	97.46%
TCN	96.95%	72.86%
KNN	80.35%	-180.69%
Random Forest	57.12%	-134.77%

Table 5. Comparison Study with other proposed models for stock market predictions

Prediction model	ML Models used	Accuracy Achieved
Kamley [17] [14]	Multiple Regression	54%
Ladan [16] [14]	Multiple Regression	89%
Muhammad [14]	Multiple Regression	97%
Proposed Model	Classification: ANN SGD KNN Regression: LSTM Linear Regression SGD TCN	 97.28% 93.65% 89.53% **99.31%** **98.85%** 97.63% 96.95%

4.2.1 Decision Tree Regression

The following figure predicts accuracy, of Decision Tree Regression. X-axis is the time parameter and Y-axis is the scaled data price. The r2 score of the decision tree regression is 98.22% for dataset1 and -54.63% for dataset2.

4.2.2 KNN Regression

The graphs of the predicted accuracy of KNN Regression are shown below. X-axis is the time parameter and Y-axis is the scaled data price. The r2 score of the decision tree regression is 80.35% for dataset1 and -180.69% for dataset2.

In all the graphs, the blue line represents the original transformed data, while the yellow line

4.2.3 SGD-Regression

The following figures are the graphs of the predicted accuracy of SGD Regression. X-axis is the time parameter and Y-axis is the scaled data price. The r2 score of the decision tree regression is 97.63% for dataset1 and 97.46% for dataset2.

4.2.4 Random Forest Regression

The graphs of the predicted accuracy of Random Forest Regression are shown below. X-axis is the time parameter and Y-axis is the scaled data price. The r2 score of the decision tree regression is 57.12% for dataset1 and -134.77% for dataset2.

4.2.5 Linear Regression

The following figures are the graphs, of the predicted accuracy, of Linear Regression. X-axis is the time parameter and Y-axis is the scaled data price. The r2 score of the decision tree regression is 98.85% for dataset1 and 98.94% for dataset2.

4.2.6 LSTM

The figures below are the graphs, of the predicted accuracy, of LSTM. X-axis is the time parameter and Y-axis is the scaled data price. The r2 score of the decision tree regression is 98.22% for dataset1 and 98.01% for dataset2.

4.2.7 TCN

The following figures are the graphs, of the predicted accuracy, of TCN model. X-axis is the time parameter and Y-axis is the scaled data price. The r2 score of the decision tree regression is 96.95% for dataset1 and 72.86% for dataset2.

The following figure shows the comparison of the 2 best performing models in classification and regression. It depicts that in classification ANN and SGD have shown better performance than the remaining classification models. It also depicts that in regression LSTM and Linear Regression have shown better performance than the remaining regression models used.

5 Comparative Study

Our research model including classification and regression models which used OHLC data for computation and model building have shown better performance than the other researchers ’project.

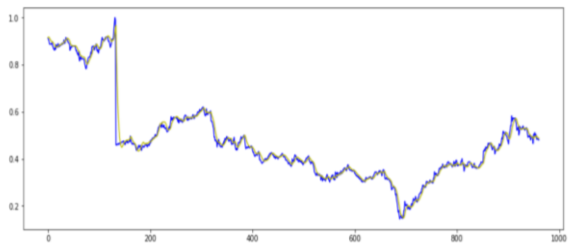

Fig. 16. Graph for Stochastic Gradient Regression

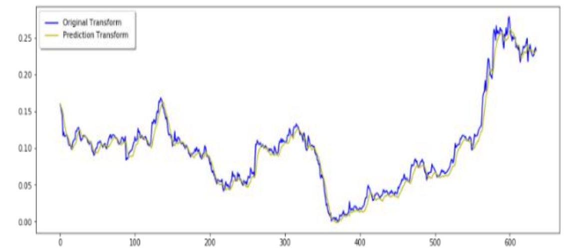

Fig. 17. Graph for Stochastic Gradient Regression Dataset2

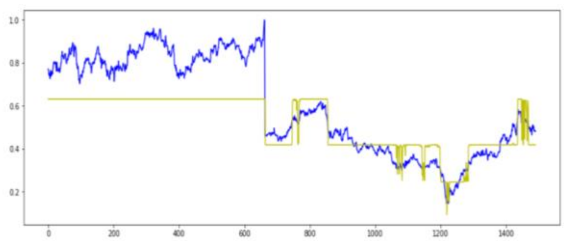

Fig. 18. Graph for Random Forest Regression

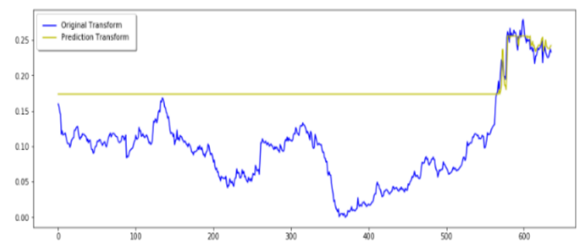

Fig. 19. Graph for Random Forest Regression Dataset2

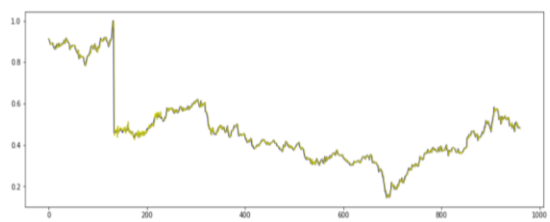

Fig. 20. Graph for Linear Regression

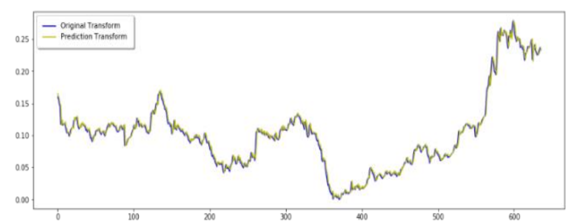

Fig. 21. Graph for Linear Regression Dataset2

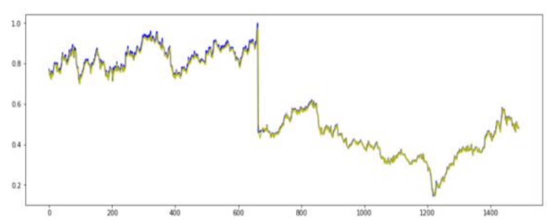

Fig. 22. Graph for Long-Short Term Memory

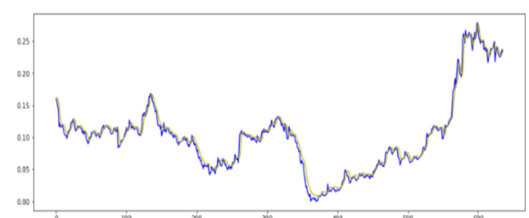

Fig. 23. Graph for Long-Short Term Memory Dataset2

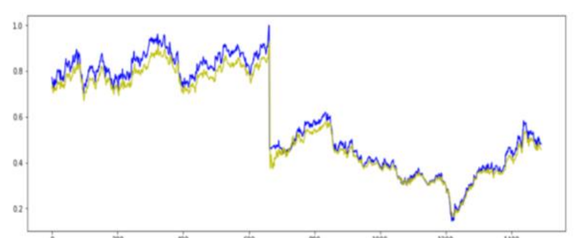

Fig. 24. Graph for Temporal Convolution Networks

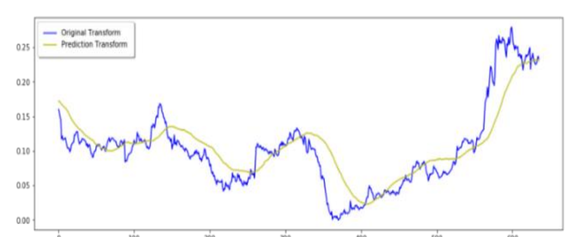

Fig. 25. Graph for Temporal Convolution Network Dataset2

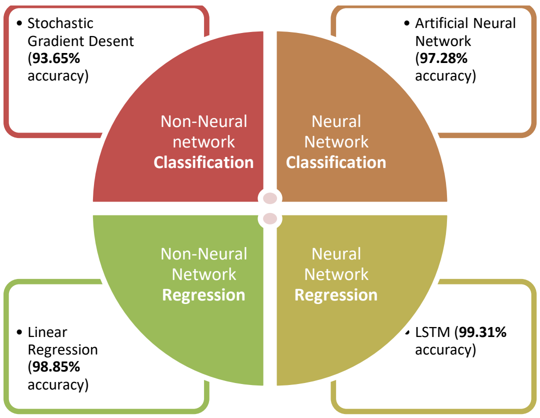

Fig. 26. Best performing Models

Our maximum accuracy achieved is 99% and 98.85% which is higher than the other individuals' models which are 89% [16] and 97% [14].

The following table illustrates the differences in accuracies of different models:

6 Conclusion and Future Work

The research paper 'Stock Price Trend Prediction Models Using Neural and Non-Neural Network Models' focuses on using OHLC data in machine learning models to predict stock price trends. When models incorporate modified OHLC prices as data for more accurate predictions, various methods are applied to solve the same problem. However, with the advancements in machine learning, we can directly use OHLC data for reliable trend prediction, not just for price forecasting. We can conclude that both LSTM and Linear Regression are reliable models for predicting stock market trends, as long as the market is stable.

The goal of the research presented in this paper was to find the best-fitting model for stock price trend prediction.

Additionally, real-time external factors/news influence human emotions and investment decisions. Ultimately, stock prices are determined by the people (individuals or organizations) investing in the company. Based on prediction accuracies, regression models and ANNs can be more reliable when incorporating real-time news analysis.

Therefore, the next phase of the project will involve integrating real-time market news into LSTM, TCN, SGD, and Linear Regression models, as these models outperformed others."

Acknowledgments

The authors would like to thank Siksha 'O' Anusandhan University for providing facilities. We also thank the participant schools for giving us opportunities to explore the cognitive response of students in different pollution levels. The author(s) also acknowledge the support of the State Pollution Control Board, Odisha, India.

References

1. **Asghar, M.Z., Rahman, F., Kundi, F.M. (2019).** Development of Stock Market Trend Prediction System Using Multiple Regression. Computational and Mathematical Organization Theory, Vol. 25, No. 3, pp. 271–301. doi: 10.1007/s10588-019-09292-7.
2. **Wen, M., Li, P., Zhang, L. (2019).** Stock Market Trend Prediction Using High-Order Information of Time Series. IEEE Access, Vol. 7, pp. 28299–28308. doi: 10.1109/ACCESS.2019.2901842.
3. **Ananthi, M., Vijayakumar, K. (2020).** Stock Market Analysis Using Candlestick Regression and Market Trend Prediction (CKRM). Journal of Ambient Intelligence and Humanized Computing. doi: 10.1007/s12652-020-01892-5.
4. **Shen, J., Shafiq, M.O. (2020).** Short-Term Stock Market Price Trend Prediction Using a Comprehensive Deep Learning System. Journal of Big Data, Vol. 7, No. 66. doi: 10.1186/s40537-020-00333-6.
5. **Nabipour, M., Nayyeri, P., Jabani, H. (2020).** Deep Learning for Stock Market Prediction. Entropy, Vol. 22, No. 8, pp. 840. doi: 10.3390/e22080840.
6. **Dai, W., An, Y., Long, W. (2021).** Price Change Prediction of Ultra High Frequency Financial Data Based on Temporal Convolutional Network. arXiv. doi: 10.48550/arXiv.2107.00261.

7. **Chen, P., Boukouvalas, Z., Corizzo, R. (2024).** A Deep Fusion Model for Stock Market Prediction with News Headlines and Time Series Data. Neural Computing and Applications, Vol. 36, pp. 21229–21271. doi: 10.1007/s00521-024-10303-1.
8. **Farimani, S.A., Jahan, M.V., Fard, A.M. (2022).** Investigating the Informativeness of Technical Indicators and News Sentiment in Financial Market Price Prediction. Knowledge-Based Systems, Vol. 247. doi: 10.1016/j.knosys.2022.108742.
9. **Song, H., Choi, H. (2023).** Forecasting Stock Market Indices Using the Recurrent Neural Network Based Hybrid Models: CNN-LSTM, GRU-CNN, and Ensemble Models. Applied Sciences, Vol. 13, No. 7, pp. 4644. doi: 10.3390/app13074644.
10. **Qasem, M., Thulasiram, R., Thulasiram, P. (2015).** Twitter Sentiment Classification Using Machine Learning Techniques for Stock Markets. International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 834–840. doi: 10.1109/ICACCI. 2015.7275714.
11. **Ariyo, A., Adewumi, A.O., Ayo, C.K. (2014).** Stock Price Prediction Using the ARIMA Model. UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 106–112. doi: 10.1109/UKSim.2014.67.
12. **Devi, B.U., Sundar, D., Alli, P. (2013).** An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50. International Journal of Data Mining and Knowledge Management Process, Vol. 3, pp. 65. doi: 10.5121/ijdkp.2013.3106.
13. **Hochreiter, S., Schmidhuber, J. (1997).** Long Short-Term Memory. Neural Computation, Vol. 9, No. 8, pp. 1735–1780.
14. **Tay, F.E.H., Cao, L. (2001).** Application of Support Vector Machines in Financial Time Series Forecasting. Neural Computing and Applications, Vol. 10, No. 2, pp. 184–192. doi: 10.1007/s00 5210170010.
15. **Kim, K.J. (2003).** Financial Time Series Forecasting Using Support Vector Machines. Neurocomputing, Vol. 55, No. 1–2, pp. 307–319.
16. **Reddy, B.S. (2010).** Prediction of Stock Market Indices-Using SAS. 2nd IEEE International Conference on Information and Financial Engineering. doi: 10.1109/ICIFE.2010.5609262.
17. **Mondal, P., Shit, L., Goswami, S. (2014).** Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. International Journal of Computer Science, Engineering and Applications. doi: 10.5121/IJCSEA.2014.4202.
18. **Nichols, E.P., Shen, Y., Xiong, R. (2015).** Deep Learning Stock Volatility with Google Domestic Trends. arXiv.
19. **Jamaludin, N., Ismail, S., Manaf, S.A. (2017).** Macroeconomic Variables and Stock Market Returns: Panel Analysis from Selected ASEAN Countries. International Journal of Economics and Financial Issues, Vol. 7, No. 1, pp. 37–45.
20. **Kamley, S., Jaloree, S., Thakur, R.S. (2013).** Multiple Regression: A Data Mining Approach for Predicting the Stock Market Trends Based on Open, Close and High Price of the Month. International Journal of Computer Science Engineering and Information Technology Research, Vol. 3, No. 4, pp. 173–180.
21. **Yaun, L.Y., J. (2014).** Test on Validity of Futures Market's High Frequency Volume and Price on Forecast. International Conference on Management of e-Commerce & e-Government.
22. **Javaid, M.U. (2010).** Determinants of Equity Prices in the Stock Market. ISSN1996-2800, Vol. 4, No. 1, pp. 98–114.
23. **Fischer, T., Krauss, C. (2017).** Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. FAU Discussion Papers in Economics.
24. **M, H., Gopalakrishnan, E.A., Menon, V.K. (2018).** NSE Stock Market Prediction Using Deep-Learning Models. Procedia Computer Science, Vol. 132, pp. 1351–1362.
25. **LeCun, Y., Bengio, Y. (1995).** Convolutional Networks for Images, Speech, and Time-Series. Handbook of Brain Theory and Neural Networks, Vol. 3361.
26. **Zhang, L., Zhang, L.L., Chen, Y.B. (2015).** Based on Information Fusion Technique with Data Mining in the Application of Finance Early-

Warning. Chinese Journal of Management Science, Vol. 23, No. 10, pp. 170–176.

27. **Soni, S., Shrivastava, S. (2010).** Classification of Indian Stock Market Data Using Machine Learning Algorithms. International Journal on Computer Science and Engineering, Vol. 2, No. 9, pp. 2942–2949.
28. **Thakkar, A., Chaudhari, K. (2021).** Fusion in Stock Market Prediction: A Decade Survey on the Necessity, Recent Developments, and Potential Future Directions. Information Fusion, Vol. 65, pp. 95–107.
29. **Pang, X., Zhou, Y., Wang, P. (2020).** An Innovative Neural Network Approach for Stock Market Prediction. The Journal of Supercomputing, Vol. 76, No. 3, pp. 2098–2118.
30. **Nti, I.K., Adekoya, A.F., Weyori, B.A. (2021).** A Novel Multi-Source Information-Fusion Predictive Framework Based on Deep Neural Networks for Accuracy Enhancement in Stock Market Prediction. Journal of Big Data, Vol. 8, No. 1, pp. 1–28.
31. **Farimani, S.A., Jahan, M.V., Fard, A.M. (2024).** An Adaptive Multimodal Learning Model for Financial Market Price Prediction. IEEE Access, Vol. 12, pp. 121846–121863.
32. **Sahoo, S., Rani-Panigrahi, C., Pati, B. (2024).** IoMT-Enabled Smart Healthcare Models to Monitor Critical Patients Using Deep Learning Algorithms: A Review. Computación y Sistemas, Vol. 28, No. 4, pp. 1823–1832.
33. **Cowlessur, S.K., Basava, A., Pati, B. (2024).** PIONEER: An Interest-Aware POI Recommendation Engine. Computación y Sistemas, Vol. 28, No. 1, pp. 179–188.
34. **Nanda, S., Panigrahi, C.R., Pati, B. (2023).** Flood Prediction with Optimized Attributes and Clustering. Computación y Sistemas, Vol. 27, No. 4, pp. 1157–1167.
35. **Mansingh, P., Pattanayak, B.K., Pati, B. (2023).** Deep Learning-Based Sentiment Analysis for the Prediction of Alzheimer's Drugs. Computación y Sistemas, Vol. 27, No. 4, pp. 979–989.

Article received on 21/06/2025; accepted on 08/12/2025.
**Corresponding author is Bibudhendu Pati.*