Hybrid Approaches To Stock Market Price Prediction Using Machine Learning And Deep Learning

Yogendra Prasad. P [1, a)], Siva Samhitha.V [2, b)], Sai Tejaswini.D [3, c)], Jnyana Prasoona.P [4, d)], Rakesh.S [5, e)], Ramu.M [6, f)]

[1] *Assistant Professor, Dept. of Data Science*
Mohan Babu University Erstwhile Sree Vidyanikethan Engineering College, Tirupathi, India [2,3,4,5]
Student, Dept. of CSSE, Sree Vidyanikethan engineering College, Tirupathi, India.
[6]*Assistant Professor,Dept. Of CSE, Annamacharya Institute of Technology,Tirupathi,India*

a) *yogendraprasad.p@gmail.com,*
b) *sivasamhitha24@gmail.com,*
c) *saitejaswini527@gmail.com,*
d) *jnyanaprasoona25@gmail.com,*
e) *rakeshyadavs705@gmail.com,*
f) *ramucsse20@gmail.com,*

Abstract— Predicting stock market prices can be really tricky There's a lot going on, like how unpredictable things are and all the different relationships in the money world. A lot of people have used traditional methods like Linear Regression (which is called LR), Support Vector Machines (SVM), and Random Forest (RF). But these often miss the tiny details in stock data.This paper explores how we might do better with newer machine learning methods. Had look at cool techniques like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Light Gradient Boosting Machine (LGBM), & XGBoost. CNN helps find complex patterns. It's usually for images but works for stock data too Then there's LGBM, which is super fast and pretty accurate. It deals with big piles of data way better than older methods. This paper check how well each of these methods—CNN, LSTM, LGBM, and XGBoost—does when looking at past stock data. The goal See which one predicts stock trends the best. This could give us some awesome insights on how machine learning & deep learning can work together in finance. With this advanced and hybrid machine learning techniques, could make even better guesses about stock prices, helping with smart investment choices.

Keywords—stock market price prediction, financial forecasting, Machine learning, Convolutional Neural Networks (CNN), Long Short Term Memory(LSTM), LightGBM(LGBM))

INTRODUCTION

The stock marketplace is a financial gadget where diverse assets such as stocks, equities, and commodities are offered and bought. It permits people to very own stocks of publicly traded groups, making it a critical a part of the global economy. buyers ordinarily aim to shop for stocks that are anticipated to upward thrust in value even as keeping off those anticipated to fall. on this fast-paced surroundings, predicting inventory expenses has become important for maximizing income and minimizing losses. The challenge lies inside the reality that inventory market information is often noisy, non-desk bound, and nonlinear. this means that inventory expenses show off frequent fluctuations (noise), their statistical properties change over the years (non-desk bound), and their relationships with external factors can be complex (nonlinear).

To deal with those demanding situations, researchers and buyers have an increasing number of became to superior predictive strategies like statistical analysis, system studying, and deep getting to know. those strategies have verified tremendous potential in forecasting stock costs by way of identifying trends, styles, and future movements inside the marketplace. however, predicting inventory costs as it should be remains difficult due to the unpredictable nature of financial markets. In this take a look at, we observe numerous fashions to expect inventory charges, which includes conventional statistical models such as ARIMA, deep studying fashions like Convolutional Neural Networks (CNN) and lengthy short-time period memory (LSTM) networks, and gradient-boosting algorithms such as LightGBM (LGBM). ARIMA is properly-acceptable for timecollection facts and helps become aware of traits with the aid of making the facts stationary.

I. LITERATURE SURVEY

Awad et al. [1] presented an innovative framework for predicting stock values by combining Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, and natural language processing (NLP) techniques with a Deep Q Network (DQN). Their research applied this combined approach to historical stock data, focusing specifically on gold stocks. Additionally, they enriched their model using social media sentiment analysis from various platforms like S&P, Nifty 50, Yahoo, NASDAQ, and various gold market-related sources. The results indicated that this hybrid model significantly improved the prediction of the next day's stock opening price, showcasing superior accuracy compared to benchmark models.

Juliana et al. [2] explored stock price forecasting using deep learning techniques optimized based on technical analysis indicators. Their work focused on the Multilayer Perceptron (MLP) algorithm in combination with a day-shifting strategy to forecast stock prices for the following week. To further enhance model accuracy, they introduced a Mean Error to Mean Price Ratio (MEMPR) metric. The experimental results demonstrated that their model achieved an R2 score of 0.995, proving its capability in accurate stock price prediction.

Peivandizadeh et al. [3] proposed a model for predicting stock market trends incorporating Pattern-based Long Short-Term Memory (LSTM) networks and analysis of sentiment from social media. Their research addressed the challenges of class imbalance in sentiment data by employing the Off-policy Proximal Policy Optimization (PPO) method, which adjusted the reward mechanism to favor the correct classification of minority classes. The model excelled in capturing temporal patterns by integrating sentiment analysis data with historical stock prices, significantly improving prediction accuracy.

Chavan et al. [4] conducted an extensive comparative study of stock market forecasting models that employ a range of machine learning and deep learning methods. Their research focused on traditional econometric models such as linear regression and modern machine learning algorithms, including support vector machines (SVM), random forests, and deep learning architectures. The study also analyzed the influence of feature selection, meta-parameter tuning, and ensemble methods on the predictive outcome of these models. Their findings revealed that deep learning models outperformed traditional approaches, particularly in terms of directional accuracy and mean squared error.

II. PROPOSED SYSTEM

Data Collection

The first step is collecting stock market data from available sources such as financial APIs, publicly available datasets, or stock market data providers (e.g., Yahoo Finance, Quandl, TATA group dataset). The data usually includes several features such as:

1. **Open, High, Low, Close Prices**: These represent the stock price movements.
2. **Volume**: The number of shares traded.
3. **Additional Features**: Some datasets may include technical indicators or macroeconomic data.

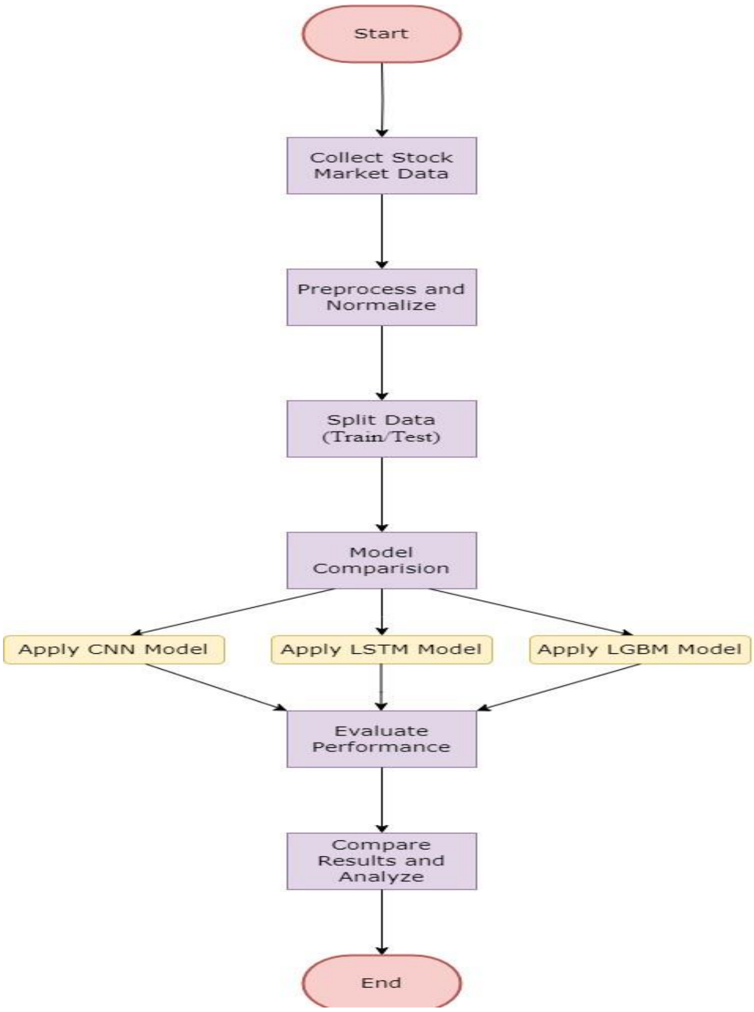


FIGURE 1 : Methodology

Preprocessing and Normalization

Preliminary processing is essential to filter and prepare the raw stock data for modeling. This step consists of multiple sub-processes:

1. **Handling Missing Data**: Stock data may have missing values due to market holidays or incomplete records. Approaches like forward filling, backward filling, or interpolation are used to fill gaps. Alternatively, rows with missing data may be dropped if data integrity is ensured.
2. **Data Transformation**: Time-series data is transformed by adding useful financial indicators. These indicators help enhance the predictive power of models. Examples include moving averages (SMA, EMA), price change percentage, and volume change rate.
3. **Normalization**: Since stock prices can vary across different scales, feature scaling is applied. Min-max normalization ensures that each feature value is scaled to a range between 0 and 1:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

This step is crucial because it prevents features with large values from dominating model training and helps the model converge more efficiently.

Data Splitting

The dataset is divided into training, validation, and test sets, ensuring a fair assessment of the prototype. The Split Data process follows a time-based chronological split, which is critical for time-series forecasting problems like stock price prediction:

1. **Training Set**: This set is used for model training and contains the greater part of the data. The stereotype learns the underlying patterns and relationships from this set. Typically, 70% of the data is granted for training.
2. **Validation Set**: The validation set, often 15% of the total data, is used for tuning hyperparameters and preventing overfitting during training. It is used to evaluate the model performance at every epoch of training.
3. **Test Set**: The remaining 15% is reserved for final model evaluation. The test set is not seen by the model during training or validation, establishing an objective assessment of the model's potentiality to popularize to new, unseen data.

The chronological split respects the temporal dependencies inherent in time-series data, certifying that the model does not have entry to later information during training.

Training Set →Validation Set →Testing Set

Model Comparison

The core part of the methodology involves applying different models to predict future stock prices. Here, we compare the performance of three models:

Model 1: Convolutional Neural Network (CNN)

CNN is used to take captive short-term patterns and localized trends in time-series data. In stock price prediction, CNN can extract useful features by applying convolutional filters over time windows of stock prices.

1 **. Convolution**: Extracts feature maps from time windows.

$$y(t) = (x * w)(t) = \sum_{K=0}^{K-1} x(t-k)w(k) \quad (2)$$

2. **Pooling**: Reduces dimensionality and highlights the most important features.

Model 2: Long Short-Term Memory (LSTM)

LSTM is constructive in apprehending long-term dependencies in successional data. Stock price movements can often depend on patterns from both short and long-term historical data, making LSTM a suitable choice.

1. **LSTM Cells**: Use memory cells and gates to retain or forget particulars.

$$f_t = \sigma(W_f . [h_{t-1} , x_t] + b_f) \quad (3)$$

Model 3: Light Gradient Boosting Machine (LGBM)

LGBM is a quick, efficient, and Supercharged decision tree-based ensemble learning model. LGBM handles large datasets and provides interpretability, making it a good candidate for structured data like stock prices.

$$L(y , \hat{y}) = \sum_{i=1}^{N} l(y_i, \hat{y}_i)+\lambda \sum_{\;=1}^{T_j} ||\theta_j||^2 \quad (4)$$

Evaluate Performance

After instructing, the functioning of the models is evaluated using various metrics. Common evaluation metrics for stock price prediction are:

1. Mean Absolute Error (MAE):

Calculates the mean immensity of errors between forecasted and true stock prices.

$$MAE = 1/n \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (5)$$

2. Root Mean Squared Error (RMSE):

Measures the square root of the average squared differences among predicted and real values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

3.R-squared (R^2):

Determines how efficiently the model accounts for the variance in stock prices.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{7}$$

The goal is to minimize MAE and RMSE while maximizing R^2.

Hybrid Model: CNN + LSTM

At this stage , a cross model merging CNN and LSTM is applied. CNN is used to extract flash patterns from the time-series data, and the production of CNN is fed into LSTM to capture long-term dependencies.

1. **CNN + LSTM Architecture**: Combines the strengths of both CNN and LSTM for sequential data like stock prices. CNN helps with extracting local features, while LSTM retains long-term trends.

Compare Results and Analyze

Finally, the consequences of the CNN, LSTM, LGBM, and CNN+LSTM models are compared using the performance metrics (MAE, RMSE, R^2). The comparison helps in determining which model is most effective in predicting stock prices.

III. DATASETS

The Tata Group Stock Dataset:

The Tata Group stock dataset is significant for this research as it includes historical data from companies including Tata Motors, Tata Steel, and TCS. Using this dataset will help you create prediction models. The goal of the project is to forecast future stock movements utilizing sophisticated deep learning models such as CNN, LSTM, and LGBM. Finding significant market trends is the primary objective in order to assist investors in making data-driven decisions.

IV. RESULT AND DISCUSSION

The stock market prediction model was created using a mix of machine learning & deep learning methods. These included CNN, LSTM, LGBM, XGBoost, plus a hybrid version combining CNN + LSTM and CNN + LSTM+LGBM+XGBoost.The main goal to see which model could best predict stock prices & find the most reliable one out there.

After training & testing all these models, the results were interesting:

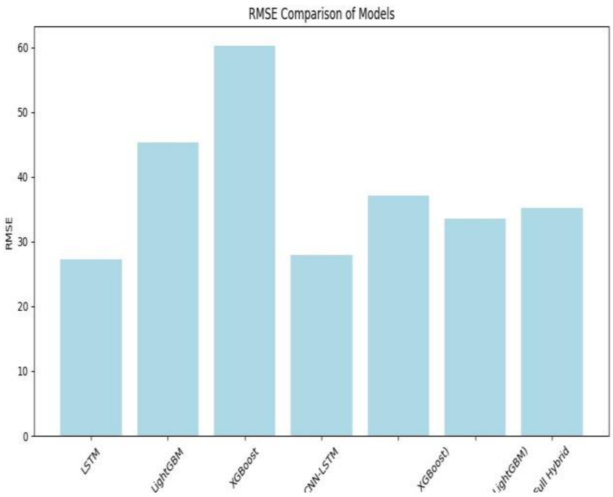
LSTM (Long Short-Term Memory): This one did really well! It captured patterns in stock price changes over time. Because it can remember information from long sequences, its accuracy was impressive. So, it’s a strong choice for predicting stock prices.

CNN+LSTM (Convolutional Neural Network + Long Short-Term Memory): Here’s where things got exciting! This model showed the best accuracy and was super close to the actual stock prices. By using CNN’s skills for picking up important features along with LSTM's talent for sequence prediction, this hybrid model nailed both short-term details and long-term trends in stock features. Its predictions stood out as the most consistent across various datasets.

LGBM and XGBoost: Now, while LightGBM (LGBM) and XGBoost gave decent predictions, they weren't as strong as the LSTM models. These tree-based models work great with tabular data but had a tougher time understanding the sequence of stock market prices like the deep learning models did.

Hybrid Model (CNN+LSTM+LGBM+XGBoost): Finally, this combined approach tried to take advantage of all the different models. Even though it performed okay, it didn't beat the CNN+LSTM model. The extra complexity didn't really help accuracy much. It seems that simpler models focusing on time patterns, like CNN+LSTM, are a better fit for predicting stock prices.

So there you have it! Each model has its strengths & weaknesses in predicting what might happen in the stock market.



RMSE Comparison of Models

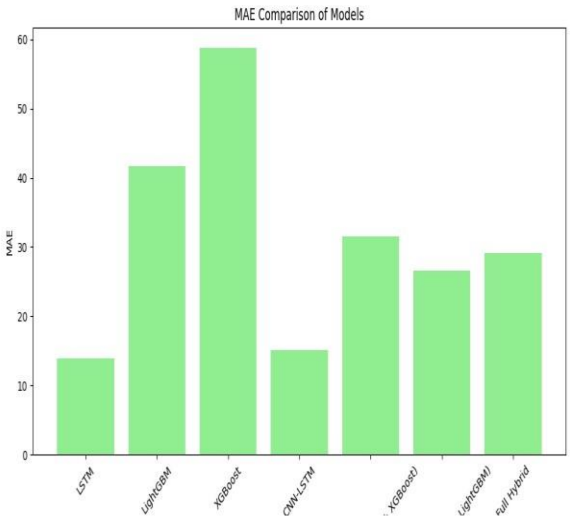


FIGURE 3 : MAE Comparison of Models

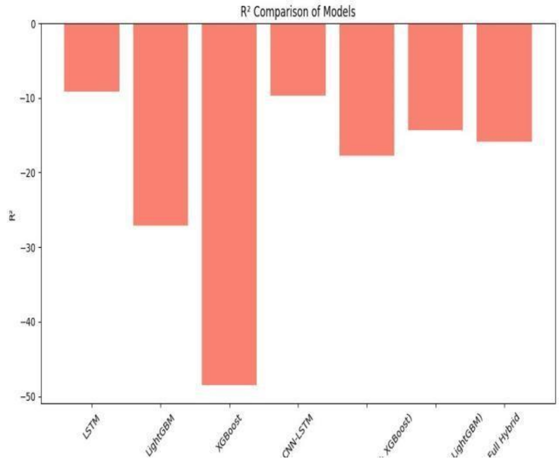


FIGURE 4 : R^2 Comparison of Models

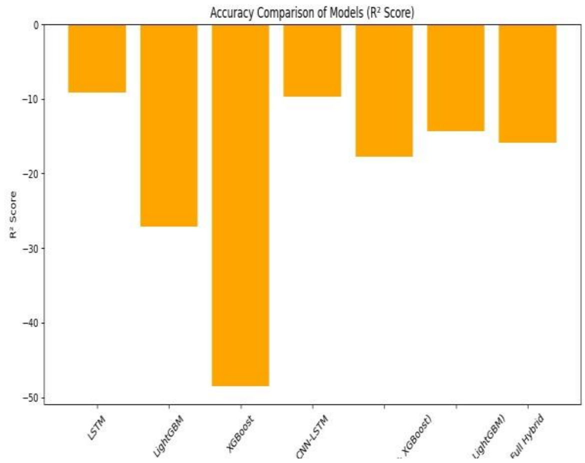


FIGURE 5 : Accuracy Comparison of Models

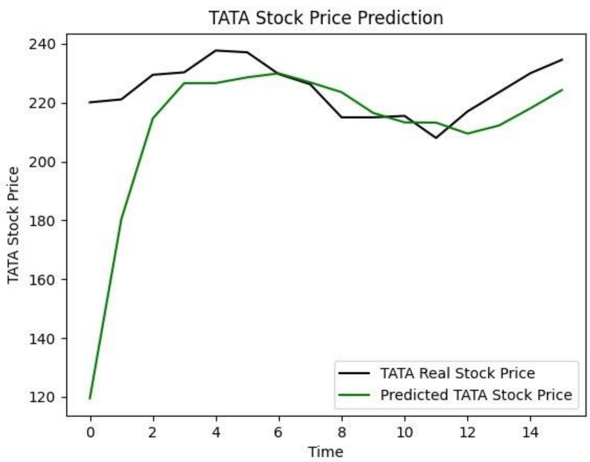


FIGURE 6 : TATA Stock Price Prediction

In FIGURE 2, shows the RMSE values of different stock market prediction models. Lower RMSE indicates better execution, with XGBoost having the highest error, while LSTM and CNN perform better. The CNN+LSTM model also shows competitive accuracy.

In FIGURE 3, illustrates the Mean Absolute Error (MAE) for various stock value estimation models. XGBoost has the highest error, while LSTM and CNN show better accuracy, with CNN+LSTM also performing well. Lower MAE values indicate better model performance.

In FIGURE 4, displays the R^2 values for different models. Negative R^2 values indicate poor model performance, with XGBoost showing the lowest score. In contrast, LSTM and CNN exhibit better predictive capability, reflecting higher R^2 values.

In FIGURE 5, shows the R^2 scores of various models, all of which are negative. Hybrid models like CNNLSTM and Full Hybrid perform slightly better than XGBoost, which has the lowest score.

In FIGURE 6, displays the real (black line) and projecting (green line) TATA stock prices over time. While the predicted values follow the overall trend, noticeable deviations occur after the 6th time step, suggesting room for improvement in model accuracy

The findings highlight how critical model selection is for predicting stock prices especially when it comes to time series data traditional machine learning models like LGBM, XGBoost are quite strong with structured data yet they struggle to grasp the subtle details in sequential data on the flip side deep learning models particularly LSTM and CNN+LSTM really shine here they have a unique ability to learn from sequences over time pick up complex patterns in the data LSTM stands out for its skill in tackling the vanishing gradient issue it keeps important information over long periods making it especially good for financial data that shows recurring trends CNN+LSTM is even better because it combines two different methods CNN looks for features finds local patterns while LSTM uses those patterns to make time-based predictions this blend enables the hybrid model to learn effectively from both past data local trends which results in top-notch prediction accuracy from a practical viewpoint this research shows that CNN+LSTM models provide more precise predictions for the stock market this can be very helpful for traders investors looking for insights grounded in data with greater accuracy in predicting stock prices decisions can be made with more confidence potentially boosting profits lowering risks in a shaky market there is room for more improvement too fine-tuning hyperparameters or adding extra datalike sentiment analysis or macroeconomic indicatorscould help enhance prediction accuracy even more still this study points out that sometimes complex hybrid models don’t always beat simpler wellmatched architectures aligning the model appropriately with the nature of the data is crucial.

V. CONCLUSION

This study is a helpful piece in the ongoing quest to understand stock market trends better. Have took a close look at different machine learning & deep learning methods. The main goal here is to see how these fancy techniques work with stock data, so we can make good predictions and figure out how well they really do. By checking out the models, we saw what they did well and where they needed some work. This helped us show off what each method could do to boost our guesses about stock prices. Our research, came up with a better way to predict stocks by mixing the trends from different algorithms. This way, hoped to get even more accurate forecasts. Have changed up the models a bit and looked at how those changes affected our predictions. Throughout the study, tried out many algorithms but really zoomed in on this cool hybrid model called CNN+LSTM to guess stock prices.Looking ahead, this paper would be neat to broaden this analysis to other stock markets & types of assets. This would help people tweak the prediction models even more and make them perform better in all kinds of market situations.

VI. REFERENCES

1. Mohamed el Mahjouby , Mohamed Taj Bennani , Mohamed Lamrini, Badre Bossoufi , Thamer a. h. Alghamdi and Mohamed el Far, Predicting Market Performance Using Machine and Deep Learning Techniques.
2. Stock Market Prediction Using Deep Reinforcement Learning by Alamir Labib Awad *ORCID,Saleh Mesbah ElkaffasORCID andMohammed Waleed Fakhr .Submission received: 5 September 2023 / Revised: 24 October 2023 / Accepted: 3 November 2023 / Published: 10 November 2023
3. Ren, Y.; Duan, J.; Li, S.E.; Guan, Y.; Sun, Q. Improving Generalization of Reinforcement Learning with Minimax Distributional Soft Actor-Critic. In Proceedings of the IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020.
4. X. Sun, M. Liu, and Z. Sima, ''A novel cryptocurrency price trend forecasting model based on LightGBM,'' Finance Res. Lett., vol. 32, Jan. 2020, Art. no. 101084
5. G. Rekha, D. B. Sravanthi, S. Ramasubbareddy, and K. Govinda, ''Prediction of stock market using neural network strategies,'' J. Comput. Theor. Nanoscience, vol. 16, no. 5, pp. 2333–2336, May 2019.
6. Y.-P. Wu, J.-M. Mao, and W.-F. Li, ''Predication of futures market by using boosting algorithm,'' in Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET), Mar. 2018, pp. 1–4.
7. FIKRY MANSOUR M. ALSHEEBAH AND BELAL A. AL-FUHAIDI, Emerging Stock Market Prediction Using GRU Algorithm: Incorporating Endogenous and Exogenous Variables.
8. Q. Liu, Z. Tao, Y. Tse, and C. Wang, ''Stock market prediction with deep learning: The case of China,'' Finance Res. Lett., vol. 46, May 2022, Art. no. 102209.
9. Saranya K1 *, Vijayashaarathi S2 , Sasirekha N3 , Koushikrajaa M4 , Lohith Raksha S5, Journal of Population Therapeutics & Clinical Pharmacology, 06 March 2023; Accepted: 19 April 2023; Published: 15 May 2023.
10. STOCK PREDICTION USING MACHINE LEARNING TECHNIQUES, Ms.N.VivekaPriya,Dr.S.Geetha.
11. Siddharth Chavan , Prathmesh Bondre, Pandurang Dahibhat, Harshali Devkar, Prof. Poonam Patekar, Stock Price Prediction Using Deep Learning Models.
12. Ali Peivandizadeh,Sima Hatami, Amirhossein Nakhjavani , Lida Khoshsima , Mohammad Reza Chalak Qazani , Muhammad Haleem , and Roohallah Alizadehsani, Stock Market Prediction With Transductive Long Short-Term Memory and Social Media Sentiment Analysis.
13. Timothy Juliana , Theodorus Devrisona , Varian Anoraa , Kristien Margi Suryaningruma, Stock Price Prediction Model Using Deep Learning Optimization Based on Technical Analysis Indicators.
14. Alamir Labib Awad * , Saleh Mesbah Elkaffas and Mohammed Waleed Fakhr, Stock Market Prediction Using Deep Reinforcement Learning.
15. Duan, G.; Lin, M.; Wang, H.; Xu, Z. Deep Neural Networks for Stock Price Prediction. In Proceedings of the 14th International Conference on Computer Research and Development (ICCRD), Shenzhen, China, 7–9 January 2022.