Grenze International Journal of Engineering and Technology, Jan Issue

Machine Learning Approaches for Company Share Rate Prediction

Prasanna H B[1], Laxmi Pujari[2] and Akshata S Naik[3]
[1-3]K.L.E Institute of Technology/Master of Computer Application, Hubballi, India
Email: prasannahb@kleit.ac.in, laxmimundargi1990@gmail.com, akshatasnaik2018@gmail.com

***Abstract*— An investor always moves forward with well-thought-out plans that, over time, maximize earnings. The share market is where most investments are made globally. Investments in the share market carry risk, therefore investors must choose when it is best to make their decisions. Prior to investing, a significant amount of forecasting and computation is required. Using an engine that forecasts corporate share prices is crucial for making well-informed investing decisions and creating tangible financial strategies for investors. It has been suggested that the engine use suggested machine learning techniques to predict future corporate share values of corporations. The application models forecast a dependable and significant share price by incorporating market indicators, sentiment research, and previous financial data on share price. The comprehensive model's mission is to identify a significant share price for the company by using machine learning methods to forecast future share prices of companies with a high degree of dependability. The yfinance library, which gives users access to a substantial database of stock market data, is utilized by the model. This information is gathered from prior financial data in order to construct the model. Past share prices, trading activity, and other significant market indexes are included in this data. After that, the data is pre-processed to address disparities that can reduce the prediction model's precision. SVR, random forests, and LSTM machine learning systems were used in the construction of the models. After the models are trained on the prepared dataset, their performance is evaluated using appropriate metrics such as accuracy or mean squared error. Cross-validation techniques are employed to ensure the robustness of the models and prevent overfitting. To validate the model, the outcomes are compared with the share prices of the current day and the day before.**

***Index Terms*— Accuracy, Data Pre-processing, LSTM, Machine Learning, SVR.**

I. INTRODUCTION

Predicting a company's share price is crucial in the finance industry since investors are eager to make the right investments. The ability to forecast a company's future share price is crucial for traders, investors, and financial experts. Machine learning processes have aided in the forecast of corporate share rates. Informed decision-making and insights into potential share price movements are made possible by machine learning, which makes it feasible to analyse historical data to identify patterns and trends. Data-driven approaches and state-of-the-art algorithms are used to achieve this. Estimating corporate share rates has become feasible thanks to machine learning procedures, which have a major impact on the banking and financial industries. It is crucial for traders, investors, and financial analysts to be able to forecast future values of company shares. Machine learning makes it possible

Grenze ID: 01.GIJET.10.1.132

to analyse historical data to identify patterns and trends, offering insights into potential changes in share price and supporting well-informed decision-making. Data-driven methodologies and contemporary algorithms are employed to make the prediction. Using machine learning algorithms, a powerful technique for predicting company share rates has emerged in finance. The yfinance library facilitates the extraction of the firm's historical share price values. With access to vast amounts of historical stock market data, including past prices, trading volumes, company financial indicators, and market sentiment, machine learning models do assist in training to analyse these patterns and associations and generate forecasts of future share prices. The machine learning systems like LSTM, Random forests, support vector machines (SVM), and neural networks are frequently employed in predicting corporate share rates. These algorithms can learn the correlations between the input variables and the target share prices by using the finance library's historical data for training. Using the patterns found, the algorithms can forecast future share values, empowering financial analysts and investors to make well-informed decisions. The application of machine knowledge methods in conjunction with the availability of comprehensive historical stock market data via finance has revolutionized the field of business share rate prediction. By using state-of-the-art algorithms and looking at patterns in historical data, machine learning representations improve accuracy and offer invaluable insights for financial decision-making. The models do forecast future prices by using the retrieved historical share price data.

II. Literature Review

The first section of the literature review describes the importance of accurate share rate forecasting for financial planning and investment decision-making. It draws attention to the challenges that lie ahead, including the complex and dynamic financial markets.

The use of machine learning algorithms for stock market price prediction is examined in the study Stock Market Prediction Using Machine Learning Algorithms. The study examines the efficiency of several machine learning techniques, such random forest learning, support vector machine training (SVM), and neural network learning (ANN), using actual stock market information [1]. The authors started a discussion about the importance of stock market forecasting and its inherent constraints. They highlight the unpredictable nature of stock markets and the need for trustworthy prediction models to assist traders and investors in making well-informed choices.

J. Liang et al. research paper explores the application of Long Short-Term Memory (LSTM) recurrent neural networks for predicting stock market prices [2]. The paper focuses on investigating the effectiveness of LSTM models in capturing the temporal dependencies of stock market data and compares different input features for prediction. In their summary of the difficulties and significance of stock market forecasting, the authors emphasize the need for models that are able to represent the intricate dynamics and temporal patterns found in financial time series data. The design and operation of LSTM networks, including the use of input, output, forget gates, and memory cells, are proposed by the authors.

Deep Learning for Stock Prediction Using Numerical and Textual Information by Y. Zhang et al. paper explores the application of deep learning techniques for stock market prediction. The paper focuses on the integration of numerical and textual information to improve the accuracy of stock market predictions [3]. The writers start off the conversation by outlining the difficulties in predicting the stock market and the drawbacks of relying just on conventional numerical data. They emphasize how crucial it is to include textual data, like news stories or financial reports, which can offer insightful analysis and sentiment data pertinent to changes in stock prices. In order to predict stock market movements, the paper presents a deep learning approach that combines textual and numerical data.

Mohammed Hamzah Abed, Muntasir Al-Asfoor, Zahir Hussain proposed the model that focuses on the classification of photographs of architectural history using deep learning methods, notably Convolutional Neural Networks [4]. The authors of this work, who come from different institutions, worked together to explore the potential applications of deep learning techniques in the field of cultural heritage understanding. The authors may examine the significance of automated classification techniques for the effective examination and conservation of architectural heritage.

A Survey on Stock Market Prediction Using Machine Learning Techniques by N. Singh et al., article emphasis on comprehensive survey of the application of machine learning techniques in stock market prediction. The paper reviews various machine learning methods and approaches used in this domain, along with their strengths, limitations, and evaluation metrics [5]. These strategies cover more sophisticated techniques like ensemble methods, deep learning, and hybrid models in addition to more conventional ones like support vector machines, decision trees, and linear regression.

III. DATA AND METHODOLOGY

A. Data

In company share rate prediction, the availability and quality of the dataset play a crucial role in developing accurate and reliable prediction models. One widely used source of financial data for stock market analysis is the yfinance library. By leveraging yfinance capabilities the researchers and analysts can access a comprehensive dataset encompassing ancient stock prices, trading volumes, financial indicators, and other market-related information pertaining to firm.

The data obtained from yfinance library is updated information dataset till the previous day share price and provides a rich source of information for training and evaluating prediction models. It includes a wide range of features that capture the dynamics of stock market movements over time. Historical stock prices provide the foundation for analysing price trends and patterns, which are essential for predicting future share rates. Trading volumes help assess market liquidity and investor activity, providing insights into market sentiment and price movements. Furthermore, yfinance provides access to various financial gauges that are commonly used in company share rate prediction. These gauges include moving averages, relative strength index (RSI), moving average convergence divergence (MACD), and many others. These structures detention different aspects of market behaviour and can be used to identify key turning points or market trends that influence share prices, Sentiment analysis is a valuable aspect of company share rate prediction, and finance can facilitate the incorporation of textual data for sentiment analysis. By integrating newscast trainings, societal media feeds, or other textual sources through finance, sentiment analysis algorithms can extract sentiment information and gauge market sentiment dynamics. This adds an additional dimension to the dataset, enabling the models to consider the impact of public perception and sentiment on share rates. The dataset from finance is dynamic and can be updated regularly to incorporate the latest market information. This allows researchers to adapt their models to changing market conditions and ensure that the predictions remain up-to-date and relevant.

B. Methodology

The methodology adopted to access the company share rate prediction involves a series of steps to ensure accurate and reliable predictions. It begins with data collection, where historical stock prices, trading volumes, financial indicators, and other relevant variables are gathered from sources such as yfinance or financial databases. The collected data is then pre-processed to handle missing values, outliers, and inconsistencies. Feature selection techniques are applied to identify the most informative variables that impact share rates.

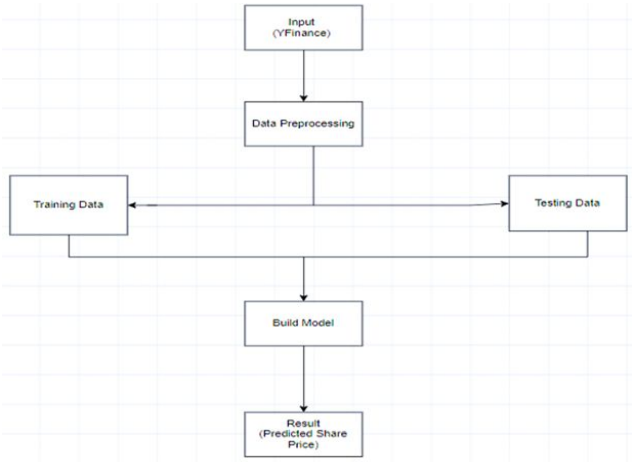

Figure 1. Steps involved in prediction of Share price rate

Following the data collection process with YFinace, appropriate machine learning models are implemented in accordance with the problem domain and the available dataset. These models can be more complex, like random forests, support vector machines, or neural networks, or they can be more conventional, like decision trees and linear regression. The preprocessed data is then used to train the selected models, which iteratively modify their parameters to discover underlying patterns and correlations between the target variable and the features. Following training, the models' performance is assessed using relevant metrics like accuracy, root mean square error (RMSE), or mean squared error (MSE). This assessment aids in determining how well the models generalize to new data

and are predictive. Furthermore, methods such as out-of-sample testing and cross-validation can be used to make sure the models are robust and avoid overfitting.The final step in the methodology is the deployment of the trained models for making predictions on new, unseen data. This involves applying the models to new instances of historical data or real-time data to forecast future share rates. Regular monitoring and updating of the models may be necessary to account for changing market conditions and ensure the predictions remain accurate and relevant.

a. Long Short-Term Memory (LSTM)

Recurrent neural networks (RNNs) with specialized architectures, such as Long Short-Term Memory (LSTM) networks, are particularly good at recognizing and analyzing long-term dependencies in sequential data. Three essential parts make up the architecture: the input gate, forget gate, and output gate. Information can move through networks in a controlled manner thanks to LSTMs. The input gate determines the amount of new information to allow into the LSTM cell, considering both the current input and the information from the previous time step. The forget gate regulates which information should be discarded from the previous cell state, allowing the LSTM to selectively forget irrelevant or outdated information.By the conceptual support with these gates, LSTMs can effectively model and remember long-range dependencies, overcoming the vanishing gradient problem typically associated with traditional RNNs.

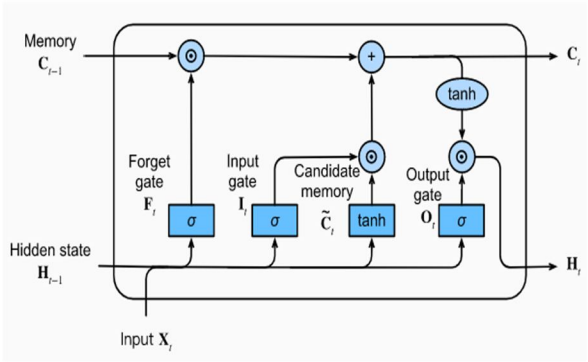

Figure 2. LSTM Architecture

The inclusion of a cell state, which acts as a memory unit, enables LSTMs to retain information over extended sequences. This architecture has proven to be highly effective in various tasks such as time series forecasting, natural language processing, and other sequential data analysis applications.

b. Support Vector Regression (SVM)

Support Vector Regression (SVR) is a powerful machine learning algorithm designed for regression tasks. It identifies support vectors, a subset of training data points that are crucial for defining the hyperplane and decision boundary. The introduction of an epsilon-tube around the hyperplane allows for a margin of error in predictions. SVR aims to minimize a loss function, such as the epsilon-insensitive loss, to balance the trade-off between accuracy and model complexity.

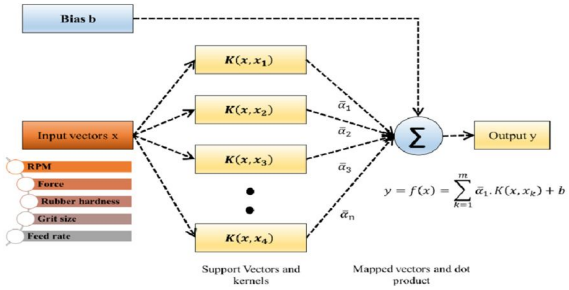

Figure 3. SVR Architecture

The regularization parameter (C) controls the width of the margin and influences the trade-off between fitting the training data and preventing over fitting. By solving an optimization problem, SVR finds the optimal hyper plane that maximizes the margin while minimizing errors within the epsilon-tube. The combination of these components enables SVR to handle nonlinear regression problems effectively and make accurate predictions.

c. Random Forest

The Random Forest architecture is based on an ensemble of decision trees that work collectively to make predictions. In this architecture, a large number of decision trees are constructed, each trained on a random subset of the training data. Each decision tree independently makes predictions, and the final prediction of the Random Forest is determined through a voting or averaging mechanism. The architecture starts by randomly selecting subsets of the training data, known as bootstrap samples, to train individual decision trees. Additionally, at each split in the tree, a random subset of features is considered, limiting the influence of any single feature.

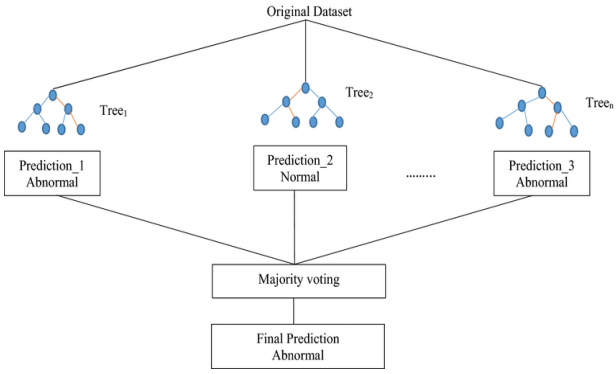

Figure 4. Random Forest Architecture

To determine the final prediction for regression tasks, the average of each tree's predictions is used. There are numerous benefits to the Random Forest architecture. It can handle high-dimensional data, capture nonlinear relationships, and handle both categorical and continuous features effectively. Moreover, Random Forests are robust against overfitting and can handle missing data or outliers.

IV. IMPLEMENTATION

A user interface (UI) for predicting company share rate is shown in Fig. 5. Users can choose the stock ticker of a particular company from a dropdown menu. After selecting a business, the user interface (UI) computes and shows the Root Mean Squared Error (RMSE), a metric that assesses the prediction model's accuracy. This provides users with an understanding of how well the model performs in estimating share rates. Below the RMSE display, a graph is presented, illustrating both the predicted and actual values of the company's share rates over time.

Figure 5. User Interface for the Share price prediction

By comparing the predicted values with the actual data, users can visualize the information presented in this graph. The time period is indicated by the x-axis, and the share rate values are shown by the y-axis. The incorporation of interactive elements may facilitate users' exploration of the graph by allowing them to enlarge it, obtain comprehensive details about individual data points, or obtain supplementary perspectives. All things considered, this user interface gives users an extensive tool for analyzing and assessing company share rate projections, enabling them to make defensible decisions based on the information presented.

The Figure 6. features a combination of key elements for stock market analysis. Firstly, a table presents a side-by-side comparison of predicted and actual stock prices for a selected company on various dates. This allows users to assess the accuracy of the prediction model and understand the variance between forecasted and real values. Secondly, the interface includes a section where users can input a specific date or interact with the graph to view

the corresponding stock price. This function helps with in-the-moment decision-making by providing rapid access to accurate stock price information. The stock's performance trend is finally shown on a graph, which shows the actual and predicted stock prices over time. Users are assisted in recognizing trends and anomalies by this visual depiction, which provides insightful information about the behaviour of the stock. Traders and investors can use the UI as a comprehensive tool to analyse stock market data effectively and make wise financial decisions.

Figure 6. Features a combination of key elements

V. Results And Discussion

This section contains the experimental data and research conclusions. The investigation involves comparing various train and test sizes to see how well various algorithms or models performed. The aim of the study was to determine the effects of data distribution between training and testing on the predicted accuracy of individual algorithms or models. We measured performance using the Root Mean Squared Error (RMSE) metric, which expresses the average deviation between the values that were predicted and those that were observed. Approximately 3000 share prices per business are collected over a ten-year period with the help of YFinance Library. This data is obtained for 8 firms, and the root mean squared error is computed and averaged for each firm. The results are excellent for the Random Forest model. This table 1 summaries the results obtained.

Table I: Machine Learning Model Performance

Models	RMSEScore
LSTM	0.4864
SVR	0.8895
Random Forest	0.0060

The results of the company share rate prediction using different algorithms revealed distinct levels of accuracy as listed in Table 1. The LSTM model achieved an RMSE of 0.486, indicating moderate accuracy in predicting the share rates. On the other hand, the Random Forest model exhibited a significantly lower RMSE of 0.060, suggesting superior performance compared to the LSTM model. The Random Forest model demonstrated a high level of accuracy in predicting the company's share rates. However, the SVR model had the highest RMSE value of 0.889, indicating relatively poor performance in comparison to the other models. The another matric is adopted to measure the performance is accuracy. In this matric also the random forest model is efficient with 98%. The table 2 summaries the results for the accuracy and it prove the evidence that the random forest model is better option to predict the share price.

Table II: Machine Learning Model Performance

Models	Accuracy in %
LSTM	67%
SVR	48%
Random Forest	98%

VI. Conclusions

Finally, while predicting a company's share rate is a difficult undertaking, financial planning and investment decision-making greatly depend on it. In this field, a lot of work has been done with machine learning algorithms like Random Forest, SVR, and LSTM. The accuracy of the prediction models varied, according to their results.

The Random Forest algorithm was most successful in capturing and modelling complex relationships in the data, as evidenced by its lowest RMSE value and highest accuracy. The SVR model performed comparatively worse than the LSTM model, which attained a moderate level of accuracy. The significance of choosing the right algorithm for share rate prediction is highlighted by these results. When selecting a model, it is critical to take the desired level of accuracy, the intricate nature of the relationships, and the dataset's characteristics into account. Moreover, data preprocessing strategies, model evaluation methodologies, and the quality and applicability of the input feature set all play a role in the success of share rate prediction, in addition to the algorithm itself. Predictive accuracy and reliability can be further improved by giving careful thought to and fine-tuning these factors. The investor is advised to use a prediction model based on random forests and with an error rate of 6 to 8%. With a benchmark of the random forest model's outcomes, the investor can base their decision-making on past performance.

REFERENCES

[1] Stock Price Prediction Using LSTM on Indian Share Market by Achyut Ghosh, Soumik Bose1, GiridharMaji, Narayan C. Debnath, Soumya Sen

[2] S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman - Stock price predictionusing LSTM, RNN and CNN-sliding window model - 2017.

[3] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, “Predicting Stock Prices Using LSTM” in Undergraduate Engineering Students, Department of Information Technology, Mumbai University, 2015.

[4] Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, “An innovative neural network approach for stock market prediction”, 2018

[5] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan Department of Computer Science and Engineering National Institute of Technology, Hamirpur – 177005, INDIA - Stock Market Prediction Using Machine Learning.

[6] Pranav Bhat Electronics and Telecommunication Department, Maharashtra Institute of Technology, Pune. Savitribai Phule Pune University - A Machine Learning Model for Stock Market Prediction.

[7] Anurag Sinha Department of computer science, Student, Amity University Jharkhand Ranchi, Jharkhand (India), 834001 - Stock Market Prediction Using Machine Learning.

[8] V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India - Stock Market Prediction Using Machine Learning.

[9] Asset Durmagambetov currently works at the mathematics, CNTFI- 'The Riemann HypothesisMillennium Prize Problems' - stock market predictions.

[10] Mariam Moukalled Wassim El-Hajj Mohamad Jaber Computer Science Department American University of Beirut - Automated Stock Price Prediction Using Machine Learning.

[11] Manh Ha Duong Boriss Siliverstovs June 2006 - The Stock Market and Investment.

[12] Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India - Indian stock market prediction using artificial neural networks on tick data.

[13] Lufuno Ronald Marwala A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering - Forecasting the Stock Market Index Using Artificial Intelligence Techniques.

[14] Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, Junwen Duan Research Centre for Social Computing and Information Retrieval Harbin Institute of Technology, China - Event Representation Learning Enhanced with External Common-sense Knowledge.

[15] Huicheng Liu Department of Electrical and Computer Engineering Queen’s University, Canada - Leveraging Financial News for Stock Trend Prediction with Attention-Based Recurrent Neural Network.

[16] Hyeong Kyu Choi, B.A Student Dept. of Business Administration Korea University Seoul, Korea = Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model.