

Stock Price Prediction using Machine Learning

Dr. T. Senthil Kumaran¹, Chethan V², Abhishek K³, Amith K R⁴, Abhishek Biradar⁵

¹Dean Professor, Dept Of CSE, ACS College Of Engineering, Bengaluru, India

²Chethan V, Dept of CSE Student, ACS College Of Engineering, Bengaluru, India ³Abhishek K, Dept of CSE Student, ACS College Of Engineering, Bengaluru, India

⁴Amith K R, Dept of CSE Student, ACS College of Engineering, Bengaluru, India

⁵Abhishek Biradr, Dept of CSE Student, ACS College of Engineering, Bengaluru, India

Abstract- Stock markets are highly dynamic and nonlinear systems influenced by multiple factors such as macroeconomic indicators, investor sentiment, and global financial events. Traditional statistical methods are often insufficient to model such complexity. This project proposes a machine learning-based approach for predicting stock prices using historical financial data. The system performs preprocessing techniques such as handling missing values, normalization, and noise reduction. Feature engineering is carried out using technical indicators like Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), moving averages, and volatility. Multiple machine learning models such as Random Forest, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) are used.

Keywords: Stock Price Prediction, Machine Learning, Deep Learning, Long Short-Term Memory (LSTM), Random Forest, Support Vector Machine (SVM), Time Series Forecasting, Financial Data Analysis

1. INTRODUCTION

The stock market is one of the most dynamic and complex financial systems in the world, playing a crucial role in the global economy. It provides a platform for companies to raise capital and for investors to earn returns on their investments. However, predicting stock prices has always been a challenging task due to the highly volatile and nonlinear nature of the market. Stock prices are influenced by a wide range of factors, including macroeconomic indicators, company performance, geopolitical events, investor sentiment, and unexpected global occurrences.

Traditional methods of stock price prediction primarily rely on statistical techniques and fundamental analysis. These approaches often assume linear relationships between variables and fail to capture the intricate patterns present in financial time-series data. As a result, they are not sufficiently effective in handling the complexity and uncertainty associated with stock market behavior.

In recent years, advancements in machine learning have provided new opportunities to address this problem. Machine learning algorithms have the capability to learn from large volumes of historical data, identify hidden patterns, and make accurate predictions without being explicitly programmed. These models can adapt to changing market conditions and improve their performance over time, making them highly suitable for stock price prediction.

This project focuses on developing a machine learning-based framework for predicting stock prices using historical financial data. The system incorporates various preprocessing techniques such as handling missing values, normalization, and noise reduction to improve data quality. Additionally, feature engineering is performed by calculating important technical indicators such as Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), moving averages, and volatility, which play a significant role in understanding market trends.

Multiple machine learning models, including Random Forest, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks, are implemented and evaluated. Furthermore, a hybrid approach that combines regression-based price prediction with classification-based trend prediction is proposed to enhance the accuracy and robustness of the system.

The main objective of this research is to build an intelligent and reliable prediction system that can assist investors in making informed decisions. By leveraging machine learning techniques, the proposed system aims to reduce uncertainty, improve prediction accuracy, and provide valuable insights into stock market behavior.

1.1 Problem Statement

The stock market is highly unpredictable due to its dependency on multiple dynamic factors. Investors often struggle to make accurate decisions due to lack of proper analytical tools.

Traditional prediction methods are not capable of handling large-scale financial data and complex relationships effectively. This creates a need for an intelligent system that can analyze historical data, identify hidden patterns, and provide accurate predictions for better investment decisions.

This project will have a number of phases, including acquiring MRI scans, testing the validity of automated detection, and developing a computer model to identify tumors from MRI scans.

2. RELATED WORK

Stock price prediction has been a widely explored research area due to its importance in financial markets. Various traditional statistical methods and modern machine learning techniques have been proposed to improve prediction accuracy.

Early approaches relied on statistical models such as the **AutoRegressive Integrated Moving Average (ARIMA)** model, which is effective for linear time-series forecasting. However, ARIMA struggles to capture non-linear patterns and sudden market fluctuations, limiting its performance in real-world stock prediction scenarios.

To overcome these limitations, machine learning models such as **Support Vector Machines (SVM)** and **Random Forests** were introduced. SVM has been widely used for classification and regression tasks due to its ability to handle high-dimensional data. Random Forest, an ensemble learning method, improves prediction accuracy by combining multiple decision trees and reducing overfitting. Despite their advantages, these models require extensive feature engineering and may not effectively capture temporal dependencies.

With advancements in deep learning, **Artificial Neural Networks (ANNs)** gained popularity for stock price prediction. ANNs can model complex non-linear relationships in financial data. However, they do not inherently consider sequential dependencies in time-series data. To address this issue, **Recurrent Neural Networks (RNNs)** and particularly **Long Short-Term Memory (LSTM)** networks were developed. LSTM models are designed to capture long-term dependencies and have shown superior performance in time-series forecasting tasks. Many researchers have demonstrated that LSTM-based models outperform traditional machine learning models in predicting stock trends due to their ability to remember past information. Recent studies have also explored hybrid models combining techniques such as **LSTM with Convolutional Neural Networks (CNN)** or integrating sentiment analysis from news and social media to enhance prediction accuracy. These approaches aim to incorporate both numerical and textual data for better decision-selection methods have been applied to improve model efficiency and accuracy. Despite these advancements, stock price prediction remains a challenging task due to market volatility, external economic factors, and unpredictable investor behavior. learning applications.

3. OVERVIEW OF THE PROPOSED PROJECT

The proposed project focuses on developing an intelligent system for predicting stock prices using advanced machine learning and deep learning techniques. The primary objective is to analyze historical stock market data and generate accurate predictions of future price movements to assist investors in making informed decisions.

The system utilizes historical data such as stock prices (open, close, high, low), trading volume, and other relevant financial indicators. This data is preprocessed through techniques such as normalization, handling missing values, and feature selection to improve model performance. The processed dataset is then used to train predictive models.

The proposed model primarily employs **Long Short-Term Memory (LSTM)** networks, a type of Recurrent Neural Network (RNN), which is highly effective for time-series forecasting. LSTM is capable of capturing long-term dependencies and trends in stock data, making it suitable for predicting sequential patterns in financial markets.

4. SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed stock price prediction system is designed as a structured pipeline that transforms raw financial data into meaningful predictions using machine learning techniques. The architecture consists of multiple interconnected modules that work sequentially to ensure efficient data processing, model training, and prediction generation.

Initially, the system begins with the data collection module, where historical stock market data is gathered from reliable financial sources. This data typically includes attributes such as opening price, closing price, highest price, lowest price, and trading volume. The collected data serves as the foundation for the prediction process.

Following data collection, the preprocessing module is applied to clean and prepare the data. This stage involves handling missing values, removing inconsistencies, and normalizing the data to a standard scale. Normalization is particularly important as it ensures that all features contribute equally to the model's learning process. The data is then transformed into a suitable format for time-series analysis and divided into training and testing datasets. classification/categorisation.

The next component is the feature engineering module, where relevant features are selected or derived to improve model performance. Additional indicators such as moving averages and trends may be generated to enhance predictive

The methodology of the proposed system follows a systematic approach to develop an accurate and efficient stock price prediction model using machine learning techniques.

The process begins with data acquisition, where historical stock price data is collected from financial datasets or APIs. A sufficient amount of data is required to ensure that the model can learn meaningful patterns and trends.

Once the data is collected, preprocessing is performed to clean and normalize the dataset. This step includes removing missing values, scaling the data, and converting it into sequences suitable for time-series forecasting. Typically, a fixed number of past observations are used to predict future values.

The next step involves designing the predictive model using Long Short-Term Memory (LSTM) networks. The LSTM architecture is chosen due to its ability to retain long-term dependencies and effectively model sequential data. The model is configured with appropriate layers, activation functions, and regularization techniques to prevent overfitting.

After designing the model, it is trained using the prepared dataset. The training process involves feeding input sequences into the model and adjusting its parameters to minimize prediction error. Optimization algorithms such as Adam are commonly used to improve convergence and accuracy.

Following training, the model is tested using unseen data to evaluate its generalization capability. The performance is measured using error metrics such as RMSE, which indicates how closely the predicted values match the actual stock prices.

Finally, the trained model is used to predict future stock prices. The results are analyzed to identify trends and patterns, which can assist investors in making informed financial decisions.

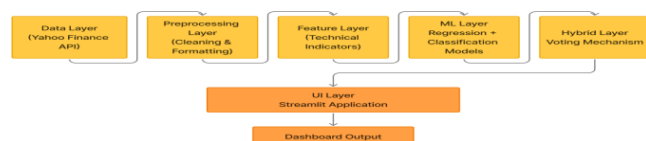


Fig 4.2: System architecture

5. PERFORMANCE MATRIX

The performance of the proposed stock price prediction model is evaluated using several standard metrics that measure the difference between actual and predicted values. These metrics help in assessing the accuracy, reliability, and effectiveness of the model in forecasting stock prices.

Mean Squared Error (MSE) is used as a primary evaluation metric. It calculates the average of the squared differences between the actual stock prices and the predicted values. By squaring the errors, this metric gives more importance to larger deviations, making it useful for identifying significant prediction errors. A lower MSE value indicates that the model's predictions are closer to the actual values.

Root Mean Squared Error (RMSE) is another important metric derived from MSE. It represents the square root of the average squared errors and provides the prediction error in the same unit as the stock prices. This makes RMSE easier to interpret in practical scenarios. A smaller RMSE value indicates better predictive performance and higher accuracy of the model.

Mean Absolute Error (MAE) is also used to evaluate the model's performance. It calculates the average of the absolute differences between actual and predicted values. Unlike MSE, MAE does not square the errors, making it less sensitive to outliers. It provides a clear understanding of the average magnitude of prediction errors.

6. EXPERIMENTAL SETUP AND RESULT ANALYSIS

The experimental setup for the proposed stock price prediction system is designed to evaluate the effectiveness of the Long Short-Term Memory (LSTM) model using historical stock market data.

The dataset used in this study consists of historical stock price records, including attributes such as opening price, closing price, highest price, lowest price, and trading volume. The data is collected from reliable financial sources and spans a significant time period to ensure meaningful pattern extraction.

Before training the model, the dataset undergoes preprocessing. This includes handling missing values, removing inconsistencies, and normalizing the data using scaling techniques such as Min-Max normalization. The normalized data is then transformed into sequential format suitable for time-series analysis, where a fixed number of past observations are used to predict future values.

The dataset is divided into training and testing sets, typically in an 80:20 ratio. The training set is used to train the LSTM model, while the testing set is used to evaluate its performance on unseen data.

The LSTM model is implemented using deep learning frameworks such as TensorFlow or Keras. The architecture consists of one or more LSTM layers followed by dense layers for output prediction. Dropout layers may also be included to prevent overfitting. The model is trained using an optimizer such as Adam and a loss function such as Mean Squared Error (MSE).

Training is carried out over multiple epochs with a defined

batch size to ensure convergence. Hyperparameters such as learning rate, number of layers, and number of neurons are tuned to achieve optimal performance.

The performance of the proposed model is analyzed by comparing the predicted stock prices with the actual values from the test dataset.

The results indicate that the LSTM model is capable of capturing the underlying trends and patterns in stock price movements. The predicted values closely follow the actual stock prices, demonstrating the model's effectiveness in time-series forecasting.

Quantitative evaluation is performed using performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). The obtained values show that the model achieves low prediction error and high accuracy, indicating good generalization capability.

Graphical analysis is also used to visualize the performance of the model. Plots comparing actual and predicted stock prices reveal that the model performs well in capturing overall trends, although minor deviations may occur during periods of high market volatility.

The results further demonstrate that the LSTM model outperforms traditional machine learning techniques by effectively handling sequential dependencies in stock data. However, certain limitations are observed, such as reduced accuracy during sudden market fluctuations caused by external factors like economic events or news.

Overall, the experimental results confirm that the proposed system is reliable and efficient for stock price prediction, providing valuable insights for investors and financial analysts.

7. CONCLUSION

In this project, a stock price prediction system was successfully developed using machine learning techniques, with a primary focus on the Long Short-Term Memory (LSTM) model. The system effectively analyzes historical stock market data to identify patterns and trends, enabling the prediction of future stock prices with improved accuracy.

The implementation of LSTM proved to be highly suitable for time-series forecasting, as it can capture long-term dependencies and sequential relationships within the data. Compared to traditional statistical and machine learning methods, the proposed approach demonstrates

better performance in handling complex and dynamic market behavior.

The experimental results indicate that the model achieves satisfactory accuracy, with low error rates as measured by evaluation metrics such as MSE, RMSE, and MAE. The comparison between actual and predicted values shows that the model is capable of closely following stock price trends, making it a useful tool for financial analysis and decision-making.

However, stock price prediction remains a challenging task due to market volatility and the influence of external factors such as economic conditions, political events, and investor sentiment. These factors can introduce uncertainty and affect prediction accuracy.

In conclusion, the proposed system provides an efficient and reliable approach for stock price prediction using deep learning techniques. It can assist investors and analysts in making informed decisions. Future enhancements may include integrating real-time data, incorporating sentiment analysis from news and social media, and exploring hybrid models to further improve prediction accuracy.

This research article describes a highly functional and responsive framework designed to facilitate the automated detection of brain cancer tumours via magnetic resonance imaging (MRI) scans using machine learning algorithms. The framework will help produce consistent results given the variance seen in the different types of medical imaging by creating a pre-processed image of MRI scans that has been standardised prior to being processed; creating an adaptive training system based on the characteristics of the extracted feature set; and providing assistance with the classification of the resulting data based on how successfully the images were analysed and classified

REFERENCES

- [1] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.
- [2] A. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162–2172, 2015.
- [3] Y. Bao, Z. Yue, and C. Rao, "A deep learning framework for financial time series using stacked autoencoders and long short-term memory," *PLoS ONE*, vol. 12, no. 7, 2017.
- [4] K. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1–2, pp. 307–319, 2003.
- [5] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, 2015.
- [6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
- [8] J. Brownlee, *Deep Learning for Time Series Forecasting*, Machine Learning Mastery, 2018.
- [9] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. Menon, and K. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017.
- [10] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, 2012.