Archives available at journals.mriindia.com

International Journal of Electrical, Electronics and Computer Systems

ISSN: 2347-2820
Volume 14 Issue 01, 2025

Design and Development of Stock Market Prediction for Optimization of Retirement Funds

Jawaharlal Kanoujia[1], Arshia Shaikh[2], Rojid Shaikh[3], Rahul Pandit[4], Dr. Roopali Lolage[5], Mr. Manthan Joshi[6]

[1,2,3,4] UG Student, [5] Professor, [6]Assistant Professor, Department of Computer Engineering, Shree L. R. Tiwari College of Engineering, Mumbai, India.

Peer Review Information	Abstract
Submission: 13 Jan 2025 *Revision: 18 Feb 2025* *Acceptance: 21 March 2025* **Keywords** *Retirement Portfolios* *Stock Market Prediction* *Ensemble Learning* *LSTM* *GRU* *Random Forest* *XGBoost, Market Capitalization* *P/E Ratio* *P/B Ratio*	Even after extensive research, it is difficult to make accurate and trustworthy stock market predictions for retirement portfolios. Automated investment systems today include tax-loss harvesting capabilities but lack advanced prediction models. An ensemble learning framework incorporates LSTM and GRU along with Random Forest and XGBoost systems in order to improve prediction outcomes and decision processes. Market Capitalization, P/E Ratio, P/B Ratio, Net Profit and EPS receive XGBoost and Random Forest analysis to establish trends while minimizing errors. The stock recommendation procedure utilizes time-series inputs from LSTM and GRU that analyze historical input data containing technical indicators such as Closing Price, SMA, EMA, MACD, RSI, Bollinger Bands, and Volume. The unified recommendation system creates final outputs through an aggregation method using all model outputs. For stock identification the evaluation procedure conducts labeling on high-performing stocks that form 30% of the total stocks and trains models to find them before final selections proceed through combined XGBoost and Random Forest scores for optimal investment choices. Backtesting against the NIFTY50 index proved the potential use of the system as a long-term financial security strategy.

INTRODUCTION

Accurate stock market predictions are crucial for managing retirement funds due to market volatility and the need for long-term financial stability. Traditional investment strategies often struggle to provide reliable forecasts, making asset allocation and risk management challenging. Stock market analysis is generally categorized into technical analysis and fundamental analysis. Technical analysis focuses on price fluctuations and historical patterns, while fundamental analysis evaluates financial reports, news, and expert opinions to assess a company's value[1].Existing investment platforms primarily focus on portfolio optimization but lack advanced predictive models. These systems often rely on conventional algorithms that do not fully utilize modern machine learning techniques for enhanced forecasting.

Ensemble learning has become a popular approach to improving prediction accuracy in stock market forecasting. Bagging reduces errors by training multiple models on different data subsets, increasing stability, while stacking refined predictions by combining diverse models in layers. These methods are particularly effective when incorporating market trends and public sentiment. However, achieving higher accuracy requires significant computational power[2,3]. Advances in machine learning have further improved these techniques, making stock price predictions more reliable

Techniques used come from multiple domains, including Deep Learning, Machine Learning, and their combinations. Additionally, Financial Modelling plays a significant role. These domains contribute to stock market prediction for optimizing retirement funds . Machine Learning is a subset of artificial intelligence which enables systems to learn independent of programming and enables automatic performance enhancement through exposure to additional data. XGBoost together with Random Forest serves as the tools we use to make predictions within the project. The prediction system uses XGBoost to learn from mistakes yet achieves better accuracy by executing multiple decision trees through Random Forest. The methods enable both highly effective data processing and forecasting precision[4]. Improves predictions by learning from prediction errors. These analytic methods assist in data assessment to deliver dependable forecasting abilities.

Deep Learning is a subset of Artificial Intelligence that follows an intuition on how the brain works and tries to replicate them. It consists of multiple frameworks among which we will be using the Time Series models like LSTM, GRU [5]. These models keep track of past data to predict future data and they seem to work pretty well on predicting stock prices[6]. It is because of deep learning we can do technical analysis using fundamental insights.

The proposed system combines ensemble learning via LSTM, GRU, Random Forest, and XGBoost to break through existing constraints. While LSTM and GRU is used to predict prices from historical stock prices, Random Forest and XGBoost improve accuracy by recommending stocks based on risk classification and financial indicators. Hit deficit in effective decision-making for optimizing retirement fund.

LITERATURE REVIEW

Ananda Chatterjee et al. [7] used time series models together with econometric models and machine learning techniques for creating a predictable stock price forecasting model. This predictive system has shown ability to analyse both historical stock information and market trends in order to create estimates about future stock values. Training of models with historical stock price data combined with trading volume data and financial indicator information took place to enhance accuracy performance according to the authors. The established research confirmed that correctly implemented predictive models successfully forecast stock prices to support improved financial investment choices. The study does not analyse important external elements that influence stock market fluctuations such as worldwide economic developments coupled with news sentiment alongside major financial events. The work does not integrate Long Short-Term Memory (LSTM) network models which exhibit superior performance processing time-series data. The accuracy of predicted models depends significantly on both high-quality inputs and the selection of relevant influencing characteristics. The accuracy of predictions depends heavily on the data quality because biased or incomplete dataset leads to unreliable results. The analysis fails to account for unexpected financial market events since these abrupt shifts and economic instabilities cause stock price effects that existing models might struggle to anticipate. The paper demonstrates that stock market forecasting becomes better through machine learning but it also clarifies that no model provides complete accuracy because of the unpredictable nature of stock market behaviour.

Luckyson Khaidem et al. [8] researched stock price direction forecasting by applying the Random Forest algorithm which has proven successful for diverse classification problems. Researchers applied historical stock prices and technical indicators combined with macroeconomic variables when producing their market prediction results. The research implementation used stock price forecasting through which Random Forest demonstrated its merits against traditional statistical evaluation methods. The collected data demonstrated that Random Forest outperformed basic forecasting models because it achieved better accuracy in predicting market movements specifically during periods of high market uncertainty. A stock market prediction benefits from using Random Forest because this methodology addresses complex financial data in ways that minimize errors in forecasting results. The analysis did not consider external macroeconomic factors which could affect the accuracy of market prediction results. Professionals have identified lack of deep learning-based LSTM network evaluation with the proposed model as a key limitation because LSTM remains highly effective at analyzing sequential financial data patterns. Studied stock indices restrict the model's application scope to broader markets because they were used to draw research findings. The model lacks any representation of intense market volatility conditions and high-speed trading behavior thus creating potential issues with its predictive capabilities. The research demonstrates that machine learning provides predictive capability to stock markets while future studies would benefit from using deep learning and multiple market factors to enhance prediction accuracy.

Ce Guo et al. [9] applied machine learning methods OLS, Random Forest, XGBoost to forecast stock prices within the technology industry. The prediction model for stock prices combined historical price and volume data with financial indicators that included the Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) and KDJ indicator which analyses stock patterns. According to the research analysis various forecasting models were evaluated for determining the optimal forecasting results. Results from the OLS model showed high performance through its lowest MSE value reaching 31.42 combined with 0.82 R^2 rate and best MAPE value of 2.92. The OLS model demonstrated the highest reliability in stock price expectation forecasting based on calculation results. Scientific studies confirm that AI models surpass standard approaches because AI detects vital data correlations in stock market information. The research faced a major drawback through its failure to use deep learning models especially Long Short-Term Memory (LSTM) even though this approach brings enhanced effectiveness in time-series data evaluation. The analysis focused exclusively on Apple Microsoft and Amazon stock behaviour because its restricted scope reduced the ability of results to extend beyond these three companies. The analysis disregarded essential outside variables that include market reactions and financial statement releases and global economic conditions because these factors affect stock market price changes. Limited historical data use in modeling results in performance issues because such models cannot provide effective results in markets with uncertain or unstable conditions.

Yifan Zhang et al. [10] conducted stock price prediction using the XGBoost algorithm because of its effective structured data processing capabilities and high accuracy rate. Stock prices together with trading volume data and technical indicators Moving Averages and Relative Strength Index and Bollinger Bands formed the basis for future stock price predictions. Research conducted an evaluation that involved XGBoost versus support vector machines and random forest as traditional machine learning algorithms to discover the best prediction method. Throughout the analysis XGBoost delivered better performance than competing models because it produced superior prediction accuracy together with decreased error rates. The authors determined that XGBoost represents an outstanding solution for stock price prediction since it excels at processing big datasets combined with discovering intricate patterns in financial information. The research analysis failed to include Long Short-Term Memory (LSTM) networks although these deep learning models are renowned for their capability to handle sequential financial data processing successfully. The prediction model would have achieved better accuracy if it included external variables such as market sentiment and financial news together with macroeconomic indicators. XGBoost demands thorough parameter adjustment to stop overfitting which results in complicated deployment for actual trading systems. The research failed to investigate how sudden market fluctuations or unpredictable events known as black swan affect stock prices because they significantly undermine prediction reliability. Ironically this research proves successful stock market prediction through machine learning but researchers should enhance accuracy by using external market data and deep learning methodologies.

PROPOSED SYSTEM

The proposed system uses historical stock market data from Yahoo Finance which consists of three years of daily prices alongside quarterly financial indicators. The system architecture has three sequential phases: The first stage employs Random Forest and XGBoost classifiers to find strong stocks through fundamental-based classification methods then the second stage uses LSTM and GRU to perform time series forecasting for stock price projections three weeks into the future before executing return-based classification in the third stage to calculate expected returns from stock forecasts which are consolidated via a soft voting ensemble to identify the top 50 ranked stocks. Out of all stocks analyzed there will be fifty recommended stocks which resulted from this scoring process.

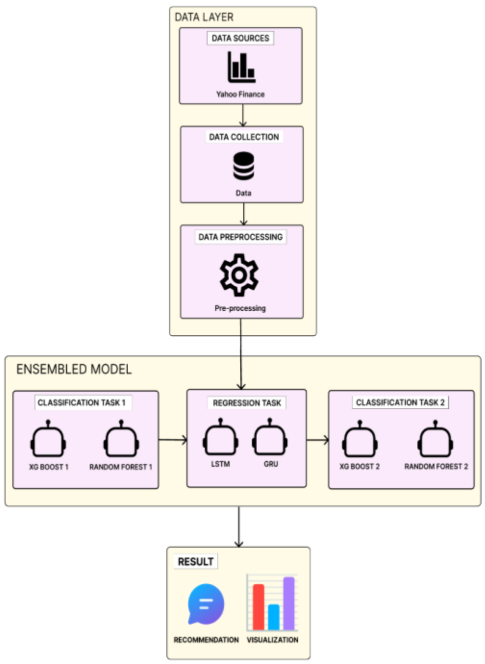

Fig 1. The Proposed System Architecture

1. **Data Acquisition and Preprocessing**
   The proposed system uses Yahoo Finance API through yfinance library to acquire data as its initial process. The system gathers fundamental data about NSE-listed companies containing P/E ratio, Net Profits, P/B ratio, Debt-to-Equity and Market Capitalization alongside EPS and inclusive of historical stock prices for three years. Machine learning functions of SMA RSI and MACD are computed using the pandas-ta library from price data collection. A normalization process through z-score transforms all feature values to standardize their scales before training deep learning models and machine learning systems.

2. **Fundamental-Based Classification**
   The company's classification process was through financial assessment standards that require Net Profit reaching ₹500 crore and P/B ratio under 5 while P/E ratio should rest between 15 to 30 and Debt-to-Equity requirements stay at 2 or below. The trained classifiers include Random Forest and XGBoost which utilize these labels as their source material. Random Forest employs class_weight='balanced' to handle class imbalance and XGBoost uses scale_pos_weight to achieve the same effect. The models produce probabilities for assessing stock fundamental strength in their final outputs. Multiple probability scores are connected to form an ensemble confidence score that strengthens classification reliability.

3. **Time Series Price Forecasting**
   The deep learning ensemble model uniting LSTM and GRU networks enables daily stock prediction for the upcoming three weeks. The machine learning system receives historical Close price, Volume, SMA_10 and SMA_30 values from 30 sequential weeks for training purposes. The two recurrent layers along with dropout regularization and dense prediction layers make up the complete structure of both LSTM and GRU architectures. Adam optimizer together with mean squared error (MSE) functions as the optimization technique for the models. Model predictions are combined through averaging after training in order to generate a strong forecast regarding future stock value.

4. **Return Calculation and Binarization**
   The calculation of expected return depends on deep learning ensemble predictions and current price values which result in percentage-based return measurements. An expected return serves as a classification benchmark which labels stocks into two categories where those achieving high returns get marked as Class 1 while others remain in Class 0. The quantile-based stock selection method identifies the most promising assets before they advance to the subsequent phase.

5. **Return-Based Recommendation Classification**
   A second classification process takes place while employing return-based labels as inputs. XGBoost and Random Forest training occurs with features including current price and predicted price. Similar to the previous phase both prediction models create probability scores for stock suggestion. The final confidence score results from the soft voting method that combines different scores through averaging. The methodology strengthens stock rankings through model-based label refinement and regulates the relationship between model variables and prediction errors.

6. **Portfolio Construction and Performance Evaluation**

   Selection of top 50 stocks according to final ensemble score leads to the formation of the investment portfolio. The selected stocks are the system's most recommended options. A performance evaluation involves calculating the total returns of the established top-50 fund while comparing it to the Nifty 50 index data over three years. A Matplotlib visualization displays the performance results which enable viewers to see the recommended fund's relationship with the market benchmark through its computed daily return product.

7. **System Integration**
   All system components integrate to form a single unified system while dependent on the information presented in the system architecture diagram (Fig 1). The system performs sequential processing from data acquisition and preprocessing activities before enabling parallel training of machine learning and deep learning pipelines which leads to stock rank decision-making. The system structure combines modules with cohesive elements to extract maximum value from temporal and financial factors when making stock predictions.

RESULTS AND DISCUSSIONS

The models were trained on financial information obtained from the Yahoo Finance API, such as stock tickers and major fundamentals like Market Capitalization, P/E ratio, and EPS. From a classification problem, a Top 50 stock fund was chosen using an ensemble of Random Forest and XGBoost models to forecast stock outperformance. A regression problem was also done on Nifty 50 daily closing prices to study more general market trends. Classification output was used to inform portfolio construction, while regression output assisted with market benchmarking. The fund's performance was tested for a three-year time frame with daily price data, with important metrics cumulative return, annualized return, volatility, Sharpe ratio, and maximum drawdown compared to the Nifty 50. This analysis provided valuable insights into stock behavior and long-term, risk-adjusted performance for retirement planning.

The evaluation metrics which were applied to a deep learning ensemble model (LSTM + GRU) for three-week stock price prediction. The comparison section includes stock price comparisons for current market values versus predicted time-series points combined with fundamental regression error measurements. The assessment of relative prediction accuracy uses MAPE (Mean Absolute Percentage Error) while MSE (Mean Squared Error) and MAE (Mean Absolute Error) determine prediction deviation magnitude and RMSE (Root Mean Squared Error) indicates standard deviation of errors and R^2 (R-squared) measures the extent of price variance explanation.

Table 1: Performance Metrics for Stock Predictions

Ticker	Current Price	Predicted Price (3 Weeks)	MAPE	MSE	MAE	RMSE	R^2
ZYDUSLIFE.NS	₹831.55	₹886.17	3.92%	2087.98	37.77	45.69	0.08
WELCORP.NS	₹779.15	₹789.65	4.31%	1884.05	34.63	43.41	-0.21
IGL.NS	₹176.95	₹179.68	4.87%	167.88	9.86	12.96	-1.39

The deep learning ensemble model exhibits remarkable capacity for short-term stock forecasting because it tracks rising trends in multiple stock selections.Table 1 shows that model forecasts increasing stock prices for ZYDUSLIFE.NS and WELCORP.NS through future price ranges that expand from ₹831.55 to ₹886.17 and ₹779.15 to ₹789.65 which indicates minimal bullish tendencies. Performance indicators from the model demonstrate strong results through Mean Absolute Error (MAE) values that remain below 37.77 for ZYDUSLIFE.NS and under 34.63 for WELCORP.NS which indicates that predicted and actual prices stay near one another. The prediction model shows strong performance through limited Mean Squared Error (MSE) scores which stayed near 2087.98 and 1884.05 thus demonstrating its ability to minimize substantial errors in predicted data. The model demonstrates reliable short-term forecasting capability as well as the ability to detect stock price momentum patterns during analysis.

The ensemble model uses Random Forest together with XGBoost for stock ranking according to predicted three-week returns. Soft-voting ensemble scoring produces average confidence scores through its computation method. The highest returning stocks along with the highest level of confidence will identify top rankings.

Table 2: Stocks Based on Ensemble Confidence Score and Expected Return

Ticker	Current Price (₹)	Predicted Price (₹)	Expected Return (%)
CYIENT.NS	1189.5	1618.5	36.07%
APARINDS.NS	4902.5	8494.84	73.28%
BASF.NS	4469.8	6917.12	54.75%
GRWRHITECH.NS	3173.2	5516.84	73.86%
CDSL.NS	1241.9	1793.16	44.39%

Short-term stock growth identification can be achieved through the ensemble classification model that combines Random Forest and XGBoost predictive methods. The system generates valuable performance predictions about future stock performance through stock evaluations based on expected returns and confidence scores.Table 2 demonstrates that GRWRHITECH.NS will increase its value by 73.86% while APARINDS.NS will experience a 73.28% growth resulting in price changes from ₹3173.2 to ₹5516.84 and ₹4902.5 to ₹8494.84. The stock performance projections indicate CDSL.NS along with BASF.NS will deliver 44.39% and 54.75% returns respectively. The projection indicates CYIENT.NS will deliver a 36.07% return as stock values should rise from their current ₹1189.5 price point to reach ₹1618.5. The ensemble model shows exceptional abilities in identifying stocks with substantial upside potential and maintains strong abilities to detect investment opportunities which makes the model very applicable for short-term investment approaches.

Fig 2. Cumulative Returns Top 50 vs Nifty50 (3 years)

The model pension fund was built using classification-based selection of top 50 performers among stocks, and regression-based short-run return prediction. To evaluate long-term performance, a cumulative return over three years was determined using historical daily closing prices. The Top 50 Fund registered a cumulative return of 1.89, while the Nifty 50 index returned 1.42, thereby registering better overall performance. From Fig 2, the fund consistently performed better than the benchmark index, particularly post the first year. The upward direction trend supports the predictive ability of the ensemble model used for stock selection. The findings reinforce the model's potential in generating optimized retirement portfolios trading off high potential returns against market-consistent risk exposure.

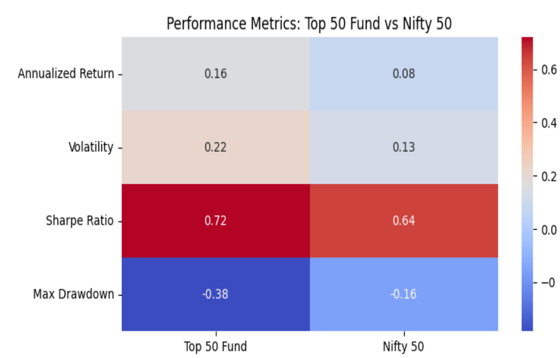

Fig 3. Performance metrics of top50 vs nifty50

Over a 3-year time horizon, the Top 50 Fund, constructed on selected stocks, was compared to the benchmark Nifty 50 on the basis of key financial performance measures. As can be observed from Fig 3, the Top 50 Fund possessed a much greater annualized return of 16% compared with 8% for the Nifty 50, indicating better long-term performance. This greater return was achieved at greater volatility 22% compared with 13% indicating greater sensitivity of the fund to the market. With greater risk, the Top 50 Fund demonstrated a better Sharpe Ratio of 0.72 compared with 0.64 for the Nifty 50, indicating better risk-adjusted returns. Nonetheless, the fund also demonstrated greater losses in bear markets, a maximum drawdown of -38% compared with -16% for the Nifty, indicating greater downside exposure. Overall, the Top 50 Fund bettered the benchmark in absolute as well as risk-adjusted returns but at an increased level of risk, indicating appropriateness for investors with higher risk tolerance and long-term focus.

To predict future stock prices, the models were trained using the past 30 days of historical data and relevant technical indicators. GRU and LSTM models were both optimized with the Adam optimizer and trained for 200 epochs. But there are issues to be addressed in handling high market volatility and accurate forecasting in the long term. Various methods have been employed to improve the performance of time series models, such as reducing the forecasting horizon, expanding the historical data, and increasing the number of training epochs with caution not to overfit. Although a decreasing trend in prediction accuracy over time is expected, the ensemble approach averaging the outputs of LSTM and GRU produced more stable and credible results compared to using a single model.

CONCLUSION

All models in this study act in unison, reaching a consensus between providing prediction accuracy and aesthetic risk exposure. The stock market suggestions originate from XGBoost and Random Forest machine learning algorithms with LSTM and GRU performing time-series forecast predictions. The organizational structure uses a combination of time-series technical data together with financial metrics from formal documentation. The system provides accurate stock market forecasts that boost retirement fund management quality. Model integration enables the system to identify successful stocks thereby improving future price estimates. The use of ensemble methods outperforms standalone models when assessing NIFTY50 index backtesting during three years because it leads to better prediction effectiveness combined with improved classification precision. The built Top 50 Fund portfolio outperformed the benchmark index by delivering better returns on cumulative returns and on yearly basis. The portfolio demonstrates superior risk-adjusted returns through its high Sharpe Ratio in uncertain periods thus serving as a suitable long-term investment choice for risk-taking investors. Implementing ensemble methods that combine multiple model predictions represents the optimal strategy to reduce model biases. The improvement in model performance originates from its reaction to market patterns and unpredictable events in the market. Since retirement investing uses the proposed model it represents an acceptable approach yet its performance requires upgrading to handle prolonged market swings and unanticipated financial alterations. Future development requires combining sentiment analysis with attention-based deep learning models combined with macroeconomic factors to increase model performance and extend flexibility. The implemented measures create stronger protection during retirement investment security.

FUTURE SCOPE

One such key progression for this project would be an incorporation of real time market predictions as well as the use of adaptive fund management to make smarter and more efficient investment strategies that optimize retirement portfolios.Automated fund rebalancing gives investors the chance to replace underperforming stocks with market-leading ones which generates better returns with reduced risks.The application of skillful machine learning techniques enables to get a more precise stock identification and to better predict the fish biomass. Additionally, the inclusion of sentiment analysis from social media and financial newswire can identify market sentiment helpful information. A second way to increase the stability of portfolio investments and long term financial gains is using an adaptive portfolio management system that reallocates to changing real time data, risks, and economic indicators.

References

Shah Raj and Tambe Ashutosh and Bhatt Tej and Rote Uday, “Real-Time Stock Market Forecasting using Ensemble Deep Learning and Rainbow DQN” in the 3rd International Conference on Advances in Science & Technology (ICAST), 2020

Ujjwal Pasupulety, Aiman Abdullah Anees, Subham Anmol and Biju R Mohan, “Predicting Stock prices using Ensemble Learning and Sentiment Analysis”, 2019

Asad, M. “Optimized Stock market prediction using ensemble learning”. 2015

Sabyasachi Mohapatra, Rohan Mukherjee, Arindam Roy, Anirban Sengupta, and Amit Puniyani, “Can Ensemble Machine Learning Methods Predict Stock Returns for Indian Banks Using Technical Indicators?” in the Journal of Risk and Financial Management, vol. 15, no. 350, 2022.

M. S. Hegde ; G. Krishna ; R. Srinath “An Ensemble Stock Predictor and Recommender System” IEEE dated 03 December 2018.

Xinxin Jiang, Shirui Pan, Jing Jiang, Guodong Long, “Cross-Domain Deep Learning Approach For Multiple Financial Market Prediction” IEEE dated 15 October 2018.

Ananda Chatterjee, Hrisav Bhowmick, and Jaydip Sen, “Stock Price Prediction Using Time Series, Econometric, Machine Learning, and Deep Learning Models,” Praxis Business School, Kolkata, 2021.

Luckyson Khaidem, Snehanshu Saha, and Sudeepa Dey, “Predicting the Direction of Stock Market Prices Using Random Forest,” IEEE Conference on Computational Intelligence, 2016.

Ce Guo, “Technology Industry Stock Price Prediction Based on OLS, Random Forest, and XGBoost,” University of Wisconsin-Madison, 2023.

Yifan Zhang,"Stock Price Prediction Method Based on XGBoost Algorithm," International Conference on Machine Learning and Financial Markets, 2023