

Optimized Stock Investment Strategy using Machine Learning

¹Mrs A. Bhagya Lakshmi ²Maruthi Kumar Bandaru ³Vanitha Duvvarapu
⁴Hima Bindhu Rупpa ⁵Uday Bhanu Teja Omми

*¹ Assistant Professor, Department of Computer Science and Engineering – Data Science
Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh, India

Abstract - With increasing volatilities in the equity market, decision-making while investing is becoming more critical, especially for retail investors in emerging economics such as India. While there is a huge availability of historical financial data, most retail investors do not have adequate tools to effectively integrate return predictions with risk assessment while making investment decisions. This paper proposes a unified framework for stock investments, where deep-learning based time series predictions and machine learning based risk classification are integrated to provide effective decision-making tools. Historical data for certain companies listed on the National Stock Exchange is collected using the Alpha Vantage API, converted to sequence data, and fed into Long Short-Term Memory (LSTM) network to predict future returns for short, medium and long periods of time. At the same time, stocks are classified according to their risk levels-low, medium and high using various risk indicators such as volatilities and returns. A module for decision-making is incorporated, which ensures that the stocks selected are according to the investor's risk profile, and there is no repetition of the stocks that are already held. Experiments conducted on the system show an improved stability, differentiation, and interpretability of results. The framework presented is a useful decision support system for data-driven retail investment planning.

Key Words - Investment Optimization, LSTM Networks, Machine Learning, Risk Classification, Stock Market Prediction, Time-Series Forecasting.

1. INTRODUCTION

Equity market plays a crucial role in wealth creation but are marked by uncertainty and volatility. The rise in digitalization of financial instruments is creating a higher demand among investors to engage in the stock market. This is detrimental to the situation since a large portion of them lacks the necessary knowledge to understand how the market works. They often use heuristics-based static trading strategies that fail to take into account price movements and dependencies throughout the period of analysis, hence making poor decisions. The expansion to become applicable, specifically LSTM networks that specialize in analyzing time series data and detecting dependencies between distant observations. However, merely predicting the stock price is not sufficient; evaluating return risk analysis and portfolio optimization in isolation. To address these gaps, the study proposes an integrated investment framework combining LSTM for stock trend prediction. This system leverages historical data from National Stock Exchange to forecast stock returns, providing tailored investment recommendations based on user's financial capacities and risk tolerance.

2. LITERATURE SURVEY

Machine Learning and Deep Learning in predicting and optimizing the stock markets have received considerable attention from researchers. Some of the findings in these areas are; Guennioui et al [1], who conducted an investigation about supervised machine learning algorithms for predicting the stock market; their approach helped in increasing the types of investors. Kumar et al [2] investigated a smart investment approach based on different learning methods, and although their approach yielded good results, they were unable to incorporate the element of risk analysis. Sebastian and Tantia [3] researched deep learning methods that can be applied in predicting stock prices and portfolio optimization, although their focus was on estimating returns based on the forecasted prices without considering the issue of risk analysis. Tripathi and Bhavani [4] investigated the use of machine learning in predicting the returns on investments and managing the risks in portfolio optimization. Their research did not investigate the concept of deep learning or the recommendation of stocks based on deep learning. The researchers Shi and Zhao [5] introduced a stock trend prediction model and strategy formulation based on deep neural networks. They considered trend predictions, inflection points, and technical indicators, such as the golden cross indicator, to formulate the model. The effectiveness of this model to predict directionality movements was established, but the model mainly concentrated on trend classification and signal confirmation. Chun et al [6] proved that machine learning based forecasting models were better for investment portfolio optimization when compared to traditional techniques, although they applied their models to portfolio optimization problems, rather than recursive multi horizon forecasting for different sectors. The FinBERT-LSTM model introduced by Gu et al [7] is combined from the finan-

cial sentiment analysis approach and historical stock prices and provides higher prediction accuracy than any other model, but for short-term forecasts only. Chang et al [8] studied differences between deep learning and machine learning models, concluding GRU performs better in terms of prediction accuracy and computational costs when compared to LSTM, while ignoring the aspects of recursive forecasting and risk classification. Karimova and Rakhmetulayeva [9] developed a sentiment analysis framework relating to stock behavior in regard to the market mood, but without applying deep learning models and risk stratification. On the other hand, Dhabliya et al [10] studied the stock prediction based on LSTM network using time-series forecast and hyperparameters' tuning. They concluded that the optimized models are superior to conventional techniques. However, their research lacks sector-specific approaches and risk categorization.

3. METHODOLOGY

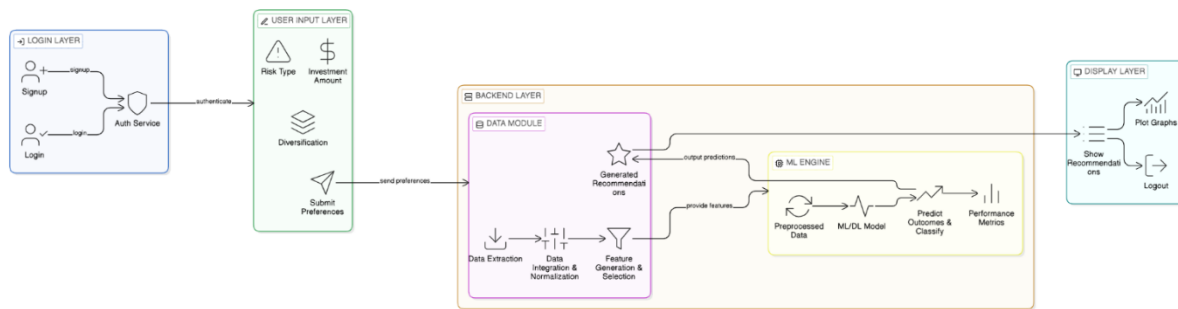


Fig. 1. System architecture of the proposed risk-aware stock investment framework.

The architectural design is presented as a unified computational framework for risk-calibrated stock investment analysis. Specifically, there is a combination of sequential deep learning-based return estimation, together with risk classification, to yield structured stock portfolio recommendations. The proposed framework is designed as an integrated stock investment decision support system with consideration for risk factors. The framework is based on predictive deep learning approaches and supervised learning for risk assessment. This system is based on a sequential pipeline for data acquisition and transformation, temporal pattern modeling, risk assessment, and finally recommendation generation. The framework is based on estimated return predictions and risk assessment categorization, as compared to traditional statistical measures. However, to maintain a realistic environment for the implementation of this framework, chronological separation is strictly enforced for the training and testing phases. This ensures that future data does not affect the learning process in any way.

3.1 IMPLEMENTATION

- A. Data Preprocessing and Feature Engineering** Stock weekly prices for companies listed on the NSE are gathered using the Alpha Vantage API and include multiple sectors. Open, High, Low, Close, Volatility (OHLCV) indicators are considered to incorporate the features of both price dynamics and trading volumes. Data processing is designed to clean and prepare data in an optimal manner by removing duplicates, adjusting timestamps, and imputing the missing values by means of forward filling and intelligent interpolation. Prices are then converted into logarithmic returns to normalize variance and facilitate comparisons. Further, feature scaling will take place by using the Min-Max method. Sequential datasets are generated using a rolling window of 26 weeks.
- B. Forecasting with LSTM** An LSTM network is employed to detect complex non-linear patterns in the time series data. The structure of the network comprises one LSTM layer with 50 hidden neurons, followed by a dropout and dense output layers. For training purposes, the Adam optimizer is used alongside the mean squared error loss function. One-step-ahead predictions are made by the network and iteratively fed back as input data to extend the forecast horizon up to 52 weeks. Based on predictions, cumulative returns are calculated for different investment time horizons (1, 3, 6 and 12 months).
- C. Risk Scores and Categorization Strategy** For risk assessment, a supervised learning-based classification task is implemented. It relies on a set of several financial metrics, including but not limited to volatility, average return rate, dispersion, etc. For each metric, three types of averages - mean, median, and exponentially weighted - are applied to calculate risk scores representing an average value of past data. Then, percentile rankings of stocks' risk scores are gener-

ated to make the comparison between stocks consistent. Finally, stocks can be classified based on risk scores into groups of low-risk, medium-risk, and high-risk.

D. Recommendation and Portfolio Generation In the last phase, recommendations for investments are made taking into consideration the estimated performance of stocks and their risk scores. To foster portfolio diversification, the system disregards the currently owned securities. A score reflecting suitability for investment is then calculated for each remaining security. Users can determine their preferences in terms of desired investments (e.g., how much money to invest, risk tolerance, diversified portfolio or concentrated portfolio). The system creates portfolios in accordance with the selected criteria.

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup and Evaluation Methodology

The approach was validated on weekly stock prices for selected NSE-listed firms. In order to make an appropriate evaluation that takes into account unseen data, a chronological split approach was adopted, wherein all data prior to December 2024 were used for training, and the whole year of 2025 data was employed as a testing dataset. Performance evaluation was conducted through four performance metrics, including accuracy of prediction, consistency in long-run forecast, reliability of risk classification, and relevance of the generated suggestions. The evaluation is relevant to real-life applications as well.

4.2 Forecasting Power of LSTM Models

4.2.1 Training Process and Generalization

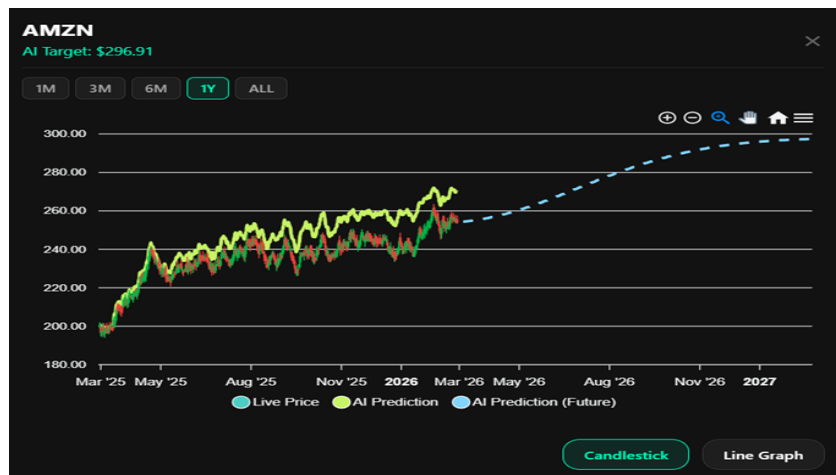


Fig.2: Predicted and actual weekly returns on 2025 test data.

The process of learning was stable; there was an observable reduction in training and validation loss curves. Dropout enabled generalization and prevented overfitting issues. On evaluation based on 2025 data, the model showed relatively good performance in terms of prediction accuracy, especially in predicting direction changes. The model showed relative stability across different sectors. Figure 2 demonstrates performance in forecasting actual and predicted values of weekly return rates.

4.2.2 Multi-Horizon Recursive Forecasting



Fig. 3: Recursive 52-week forward return projections.

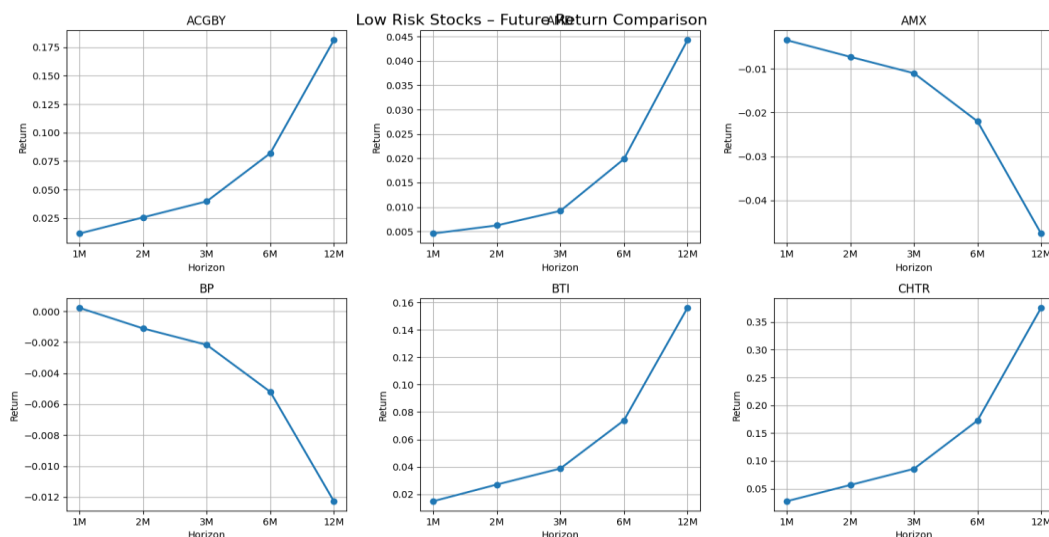
The recursive forecasting technique allowed the model to predict outcomes up to 52 weeks ahead. Even though there were slight errors that could accumulate at longer time frames, directional trends were maintained without any significant instabilities. The adoption of log returns assisted with numerical stability and error prevention in long-horizon projections. Figure 3 demonstrates the long-term forecasting horizon and overall trends behavior.

4.3 Horizon-Based Return Analysis

The returns were aggregated over different time intervals, such as 1, 2, 3, 6, and 12 months. Shorter periods produced better predictions because of less errors accumulation while longer periods gave additional perspectives about market trends. This way, an investor can customize model outcomes depending on his desired investing period.

4.4 Risk Classification and Comparison

The risk classification was estimated with the help of supervised learning models based on previously observed market volatility and tested on new data sets. As expected, the model shows distinctive separation between low and high-risk groups, while medium risk group displays overlapping features because of its transitional properties. Regardless, the classification is stable in changing market environment. Figure 4 provides cumulative return behavior across risk groups.



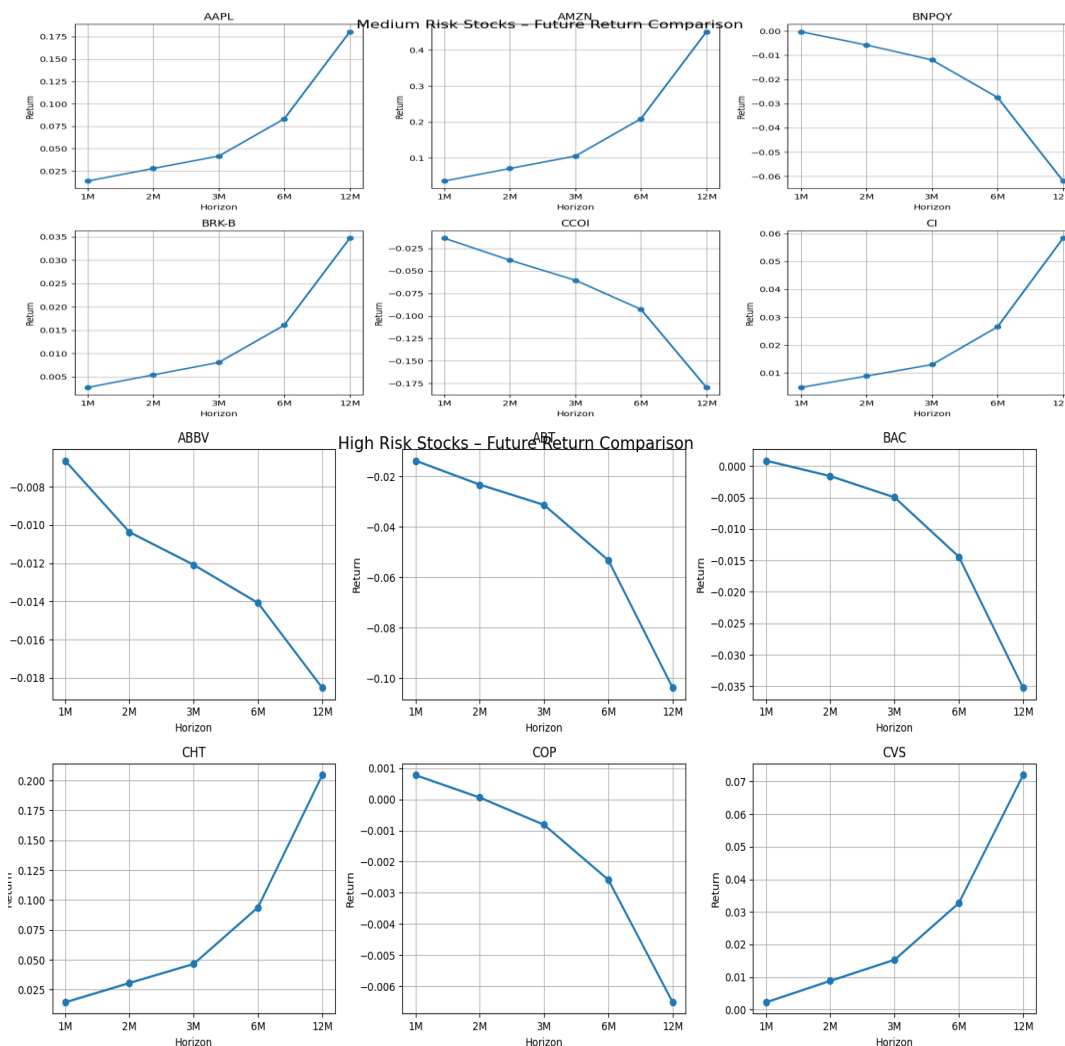


Fig. 4: Comparative cumulative future return trajectories across low-, medium-, and high-risk stock categories.

4.4 Risk Classification and Comparison

The risk classification component was validated via supervised learning algorithms trained on past volatility-related features and then tested using unseen data. The output demonstrates significant separation among low- and high-risk classes, with the medium-risk class demonstrating overlapping due to transitional nature. Nevertheless, classification remains consistent despite changes in the market environment. Figure 4 shows cumulative return trends among risk classes, depicting the connection between risk and profitability.

4.5 Recommendation System and Evaluation Results

The recommendation system uses return prediction, risk levels, and diversification strategies to build a portfolio of securities. Securities satisfying certain conditions are selected and sorted based on a suitability score. Distinct strategies lead to dissimilar results. A conservative strategy places emphasis on stability, whereas an aggressive strategy emphasizes returns. Figures 5 and 6 present recommendation interfaces and investor setting configurations, respectively.

In general, the results suggest that:

The LSTM algorithm detects temporal dependencies efficiently

The recursive prediction approach allows versatile analysis

Risk classification adds value to interpretability

The unified platform provides practical investment advice

Rather than serving as a predictor only, the proposed framework acts as a decision-making support tool for risk management and temporally informed investments.

5. CONCLUSION

The focus of this paper is on designing a stock investment strategy based on deep learning techniques that can forecast stock prices and machine learning techniques that can help classify the level of risk. The proposed method seeks to help retail investors make better decisions with the use of predictive analysis, risk management, and diversification within a unified approach. Specifically, the prediction model uses long short-term memory neural networks to predict stock returns based on weekly data to capture the trend while avoiding noise through a sliding window technique. Strict temporal partitioning of the training and test data without any information leak ensures that training data will be used until year 2024, while test data will begin from 2025. Additionally, risk classifications are combined with predictions of stock returns to improve the investment approach, stressing diversification among different sectors.

REFERENCES

- [1] O. Guennioui, D. Chiadmi, and M. Amghar, *Machine Learning-Driven Stock Price Prediction for Enhanced Investment Strategy*, 2024.
- [2] H. Kumar, H. E., and A. S. Sastry, *Smart Investment: Advancing Stock Market Predictions Through Machine Learning and Deep Learning*, 2024.
- [3] A. Sebastian and V. Tantia, *Deep Learning for Stock Price Prediction and Portfolio Optimization*, 2024.
- [4] A. Tripathi and D. Bhavani, *Enhanced Portfolio Optimization: Integrating Machine Learning and Risk Management*, 2024.
- [5] M. Shi and Q. Zhao, "Stock Market Trend Prediction and Investment Strategy by Deep Neural Networks," in Proc. 2020.
- [6] D. Chun, J. Kang, and J. Kim, "Forecasting returns with machine learning and optimizing global portfolios: evidence from the Korean and U.S. stock markets," *Financial Innovation*, 2024.
- [7] W. Gu, Y. Zhong, S. Li, C. Wei, L. Dong, Z. Wang, and C. Yan, "Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis," in Proc. 2024.
- [8] V. Chang, Q. A. Xu, A. Chidozie, and H. Wang, "Predicting Economic Trends and Stock Market Prices with Deep Learning and Advanced Machine Learning Techniques," 2024.
- [9] L. Karimova and S. Rakhmetulayeva, "Application of the Algorithm for Analysing Stock Prices Based on Sentiment Analysis," in Proc. 2023.
- [10] D. Dhaliya, H. M. Al-Jawahry, V. Sharma, R. Jayadurga, and M. Jasmin, "Long Short-Term Memory (LSTM) Networks for Stock Market Prediction," in Proc. 2023.