International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

A Study on Artificial Intelligence & Machine Learning in Stock Index Prediction for Investors

Chavan Prashanth, T. Venkata Ramana

How to Cite this Article:

Prashanth, C. & Ramana, T. V. (2026). A Study on Artificial Intelligence & Machine Learning in Stock Index Prediction for Investors. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.708

https://doi.org/10.55041/ijcope.v2i4.708

ABSTRACT

The study by A Study on AI and ML in Stock Index Prediction to Investors provides the researcher with the opportunity to explore how AI and ML can be used to predict the trends of major stock indexes in India including NIFTY 50, SENSEX, NIFTY BANK, FIN Nifty, and NIFTY AUTO. The study is pertinent because it addresses the growing complexity, volatility and non-linearity of financial markets in which traditional analytical instruments tend to fail in making accurate predictions. The main goals of the study include determining the level of effectiveness of AI/ML models in stock index movement prediction, determining the significant technical and financial aspects that affect the accuracy of predictions and investment decisions, and suggesting a more sophisticated hybrid AI and ML model that would yield higher forecasts and investment decisions. The secondary sources of data used to base the study will include the historical data of the stock index of the National Stock Exchange, Bombay Stock Exchange, and the financial websites like Yahoo Finance. The variables in the dataset are the opening, high, low and closing prices and also the trading volume. It also displays technicals like Moving Averages, Relative Strength Index, Moving Average Convergence Divergence and volatility. Various tools and methods are used in the research, such as Python-based ML models, such as Support Vector Machines, Random Forest, k-Nearest Neighbours , and Deep Learning models, including Long Short-Term Memory and Gated Recurrent Unit .The performance of the models is analysed by such statistical parameters as Accuracy, Precision, Recall, Mean Absolute Error , Root Mean Squared Error and R 2 score. The findings have shown that most of the regression models tend to be ineffective, have high error rates, and negative R 2, implying that they are weak predictors due to the complex and dynamic nature of stock markets. Models such as kNN and Random Forest are more efficient in some indices but the overall accuracy is not so high to be able to make valid predictions. Lastly, the article points out that despite the fact that AI and ML techniques can be successfully implemented to stock market, the effectiveness of the methodology lies in proper feature engineering, model optimization, and data of different types. It suggests that AI and ML hybrid models can be rather helpful in improving the accuracy of predictions and helping to make better investment decisions, risk management, and portfolio optimization in the actual financial markets. The paper discusses the use of AI&ML in stock market prediction, and their use on stock indices and the NIFTY 50, SENSEX, Deep Learning, LSTM, GRU, random forest, k-nearest neighbors, technical indicators, and time series analysis of financial forecasting and hybrid models. Artificial Intelligence, Machine Learning, Stock Market Forecasting, Stock index, NIFTY 50, SENSEX, Deep Learning, LSTM, GRU, Random Forest, k-Nearest Neighbours, Technical indicators, Time Series analysis, financial forecasting, Hybrid models.

KEY WORDS:Artificial Intelligence (AI), Machine Learning (ML), Stock Market Prediction, Stock Index Forecasting, NIFTY 50, SENSEX, FIN Nifty, NIFTY Bank, NIFTY Auto, Deep Learning, LSTM, GRU, Random Forest, k-Nearest Neighbours (kNN), Support Vector Machine (SVM), Technical Indicators, Time Series Analysis, Financial Forecasting, Hybrid Models

INTRODUCTION

The stock market is an important part of the financial system, helping to create capital and investment opportunities. However, predicting stock price movements is a complex task due to the dynamic, volatile, and uncertain nature of financial markets. The behavior of stock market indices such as the S&P 500 and the NASDAQ 100 is difficult to predict due to the interconnectedness of many economic, sentiment-based and global factors [1][11]. In the past, stock market prediction techniques such as technical analysis and fundamental analysis were based solely on historical data and financial indicators. While these techniques produce valuable insights into the stock market, they are not able to provide any insight into the complexity and non-linear relationships present in financial time series data [10][12]. Due to this, decision-makers and researchers have begun to turn to artificial intelligence (AI) and machine learning (ML) technologies that allow for the analysis of large amounts of data to identify hidden patterns and trends [8][15]. A number of recent studies have indicated that financial datasets are frequently noisy, non-linear, and affected by unforeseen external factors, thus making accurate forecasting difficult [2][14]. ML models, such as support vector machines, random forests, neural networks, and deep learning architectures, have been widely accepted for use in stock market prediction due to their ability to capture the relationships found in financial assets and increase prediction accuracy as compared to other methods [9][16]. Moreover, systematic literature reviews have suggested that the use of ensemble and hybrid stock market prediction techniques—involving a combination of multiple stock price prediction techniques—offers an increase in prediction stability and performance, while no single technique consistently outperforms all others [17]. Therefore, stock market prediction requires a comparative analysis of multiple stock prediction methodsThe predictive accuracy of various financial forecasting techniques can be significantly affected by the quality of the input data used, as well as the feature selection and the tuning of the forecasting models used [9]. AI and Big Data combined are changing the way financial markets operate (i.e., enhanced speed of data-driven decision making); these advancements in technology improve the efficiency of the market and trading strategies, but they create new difficulties such as increased volatility, overfitting, reduced interpretability, and regulatory uncertainty [4][13]. Furthermore, prior research indicates that reliance solely on historical price data is insufficient to generate accurate price forecasts. In order to produce higher levels of predictive accuracy (e.g., using news sentiment, macroeconomic measures, and market dynamics) beyond just historical prices [5][7]. Although advancements are being made rapidly, reliable and accurate forecast performance is still an ongoing challenge for many AI-based forecasting models. (i.e., certain forecasting models are capable of producing high-quality forecasts but are not able to generalize across different datasets, timeframes, etc.) This creates an immediate need for improvements in the development and application of forecasting models, feature engineering, and evaluating forecasting models [3][6]. Given the lack of research associated with the application of machine learning forecasting models to stocks, the objective of this study is to evaluate and compare many of the different machine learning forecasting models that are available for forecasting stock prices/returns through the examination of the model performance relative to the indices/metrics utilized in the examination of the model performance. By reviewing each machine learning forecasting model's advantages and limitations, the article provides insight into the practical application of these AI/ML models in forecasting financials in real-time.

MATERIALS AND METHODS

It is a quantitative and empirical research design because it is aimed at analyzing numerical values and determining the trends in Stock market indices through the application of the AI&ML methodology. The study relies on the second-hand data, gathered through the trustworthy sources, including NSE, BSE, and Yahoo Finance. These variables are open, high, low, close prices and trading volume; technical indicators comprise of Moving Averages, RSI, MACD and volatility.

Cleaning, normalization, and feature engineering are performed on the collected data to make the data accurate and consistent. The research employs the Machine Learning models (SVM, Random Forest and kNN) and Deep Learning models (LSTM and GRU) to study and predict the stock index moves. The data is separated into the training and testing sets in order to test the model performance and prevent overfitting.

Classification measures used to evaluate the performance of the models include Accuracy, Precision, Recall, and F1 Score and regression measures include MAE, RMSE and R2. There is a comparative analysis to find out the best model in terms

of prediction and error rates. In general, the methodology guarantees a systematic and trustworthy way of researching the stock market prediction with the help of AI techniques.

Results and Discussion

Table 1: Sensex Model Scores

Model	MAE	RMSE	MAPE (%)	R2 Score
SVR	23747.11	24325.32	31.04	-20.2907
RandomForest	25149.76	25996.98	32.9	-23.3174
kNN	24355.83	25540.92	31.81	-22.4717
VotingEnsemble	24435.92	25175.26	31.94	-21.8045
StackedModel	30143.73	30600.15	39.53	-32.6914

Regression findings of the SENSEX index are a clear indication that all the above machine learning models used are ineffective in forecasting stock prices. This is indicated by the negative values of the R 2 in all the models which means that the models cannot explain the variability in the dependent variable. Practically, this implies that the models cannot describe the dependence between historical returns and future prices and are even poorer than a simple model which forecasts the mean value.Support Vector Regression (SVR) demonstrates the lowest possible MAE and RMSE values which means that the model predicts the results slightly better than the other models. Nevertheless, the change is not significant and much-needed to be deemed as trustworthy. Random Forest and kNN models also demonstrate similar performance, which implies that both tree and distance-based algorithms can only detect limited patterns but still cannot be accurate predictors.This is because the Voting Ensemble model yields more reliable conclusions but does not substantially increase the performance, which means that simple averaging models to each other does not boost the predictive power. Conversely, the Stacked Model has the lowest performance, and the largest MAE, RMSE, and MAPE values and the lowest R 2 value. This implies that stacking brings about overfitting whereby the model turns out to be overly complex and learns noise rather than significant patterns.The MAPE values are between 31 and 39, which show that the model has high percentage errors, therefore it is not reliable to use in practical forecasting. The findings, in general, indicate that lag-based features alone cannot be used to capture the dynamic nature and complexity of the SENSEX index

Table 2: Nifty 50 Model Scores

	MAE	RMSE	MAPE (%)	R2 Score
SVR	7543.43	7738.94	32.46	-19.0605
Random Forest	7999.92	8270.45	34.44	-21.9107
kNN	7850.54	8203.47	33.77	-21.5411
Voting Ensemble	7772.81	8012.87	33.45	-20.5058
Stacked Model	9558.29	9712.63	41.29	-30.5976

The NIFTY 50 regression analysis has a trend to the SENSEX. All the models have R^2 values. This means the models do not work well. They cannot explain the changes in stock prices. They do not find the patterns in the data.

The SVR model works a little better than the models. It has MAE and RMSE values.. The differences are small and do not help improve the accuracy of the predictions. The Random Forest and kNN models are fair; they help to identify relationships, but do not qualify to generate universal solutions. The results from Voting Ensemble model are not better than from Random Forest model; therefore, having an ensemble of the predictions does not solve the issue. On the other hand, the Stack Model performed least well, indicating that increased complexity in a model does not inherently improve its performance. In some cases it obstructs the model's generalisation ability due to over-fitting. MAPE values for NIFTY 50 regression and SENSEX are between 32%-41% which indicate the performance of the model is well below acceptable limits for real-world applications and does not effectively uncover current market trends due to a lack of adequate input data. Overall both NIFTY 50 and SENSEX pose barriers with sub-par predictive capabilities.

Table 3: Nifty Bank Model Scores

Model	MAE	RMSE	MAPE (%)	R2 Score
SVR	14535.01	14823.61	29.35	-24.929
RandomForest	12939.37	13566.86	26.07	-20.7189
kNN	12587.52	13359.49	25.37	-20.06
VotingEnsemble	13333.85	13795.29	26.89	-21.4564
StackedModel	18162.6	18415.62	36.76	-39.0176

The performance of all the models predicting performance of the NIFTY BANK is poor. All four models produced high error rates meaning they perform poorly in predicting market movements. MAPE values vary from 25% - 36% showing a clear indication of large prediction errors. kNN performed marginally better than the other three models in terms of MAE (12587.52), RMSE (13359.49), and MAPE (25.37%). This improvement is low enough that it cannot be considered meaningful. Random Forest and SVR both perform poorly as well and show a significant degree of error indicating their inability to effectively ascertain patterns. Voting Ensemble produced consistent results, but not enough to provide any additional performance enhancements. Stacked Model performed the poorest in terms of MAE (18162.6), RMSE (18415.62), and MAPE (36.76%) and also produced the lowest R2 value (-39.0176), indicating overfitting and generalization error. Collectively, all four models produced significantly negative R2 values, suggesting an inability to explain stock price movements within this highly volatile and changing environment of the financial sector. Therefore, the results of these four predictive-modeling techniques cannot be relied upon to make accurate forecasts for the performance of the FIN NIFTY index.

Table 4: Nifty Financial Services Model Scores

Model	MAE	RMSE	MAPE (%)	R2 Score
SVR	6046.25	6253.12	26.87	-14.3905
RandomForest	5967.88	6318.58	26.47	-14.7144
kNN	5662.26	6120.37	25.03	-13.744
VotingEnsemble	5905.62	6201.66	26.18	-14.1382

StackedModel	8052.71	8224.0	35.94	-25.621

If we consider the degree of discrepancy between the NIFTY Financial Services forecasts and actual results, it appears to have performed fairly well. The magnitude of the errors is relatively small (with MAPE values between 25 and 26 percent), which indicates a reasonably close forecast to actual outcomes.

In terms of comparative performance, the kNN model provides superior forecasts than the Random Forest model, suggesting that both types of models are effective (tree-like and similarity-based) for identifying patterns in the NIFTY Financial Services data. However, the R^2 values are not particularly high.. They are better than what we see with other indices. This means the models are a little better at predicting what will happen with the NIFTY Financial Services.

The SVR model is okay. The Voting Ensemble gives us steady results but it does not do much better than the other models. The Stacked Model does not do well which shows us that using models does not always make things better.

Overall the NIFTY Financial Services results are better, than what we see with indices but the models are still not good enough to use in real life. This means we need to find ways to look at the data and add more information to make the models work better.

Table 5: Nifty Auto Model Scores

Model	MAE	RMSE	MAPE (%)	R2 Score
SVR	12051.23	12459.99	53.02	-14.5455
RandomForest	11652.57	12161.97	51.05	-13.8107
kNN	11470.12	11978.24	50.31	-13.3666
VotingEnsemble	11710.16	12170.5	51.39	-13.8315
StackedModel	13280.96	13654.45	58.65	-17.6688

The NIFTY AUTO results are really bad. The models are not doing a job they have a lot of mistakes. The error is than 50%, which means they are not good at predicting what will happen.

Among the models kNN does a little better. It is not a big difference. The Random Forest and SVR models are also not doing well. The Voting Ensemble does not help much either.

The Stacked Model is the worst it has the mistakes and a very bad score. This means it is not working well and is trying to be too perfect. All the models have scores, which means they do not understand what is happening with the stock prices.

The reason for this is that the auto industry is very unpredictable and has a lot of things that're specific to that industry. The models are not good at dealing with this. Overall the models are not good, at predicting the NIFTY AUTO index.

Table 6: Classifications of SENSEX

Model	Accuracy	Precision	Recall	F1 Score	AUC
SVM	0.5556	0.5583	0.9576	0.7054	0.478
RandomForest	0.5051	0.55	0.6	0.5739	0.4856
kNN	0.5051	0.5523	0.5758	0.5638	0.4931
VotingEnsemble	0.5084	0.5531	0.6	0.5756	0.4964
StackedModel	0.5556	0.5556	1.0	0.7143	0.5033

The SENSEX classification results are not very good with accuracy of 50 to 55 percent, which is almost like guessing.

The SVM and the Stacked Model are really good at finding the trends but they are not very good at being precise.

This means they are biased towards saying the trends will go up.

The F1 scores for the SENSEX classification results of these models are okay. The AUC values are very low around 0.5, which means they are not good at classifying.

The Random Forest and kNN and Voting models, for SENSEX classification results are a bit more balanced. They are still not very good so we cannot really use any of these models for real purposes.

Table 7: Classifications of NIFTY 50

Model	Accuracy	Precision	Recall	F1 Score	AUC
SVM	0.5507	0.559	0.9641	0.7077	0.4767
RandomForest	0.4932	0.544	0.6287	0.5833	0.4964
kNN	0.4899	0.546	0.5689	0.5572	0.4768
VotingEnsemble	0.5101	0.5539	0.6766	0.6092	0.4826
StackedModel	0.5608	0.5636	0.982	0.7162	0.5211

The NIFTY 50 classification performance is similar to the SENSEX performance. The accuracy of the NIFTY 50 is between 49% and 56%. This means the NIFTY 50 models do not have a lot of power to predict what will happen.

The models for NIFTY 50 have a time doing better than just guessing. This shows that it is really hard to predict what the stock market will do.

The Stacked Model for NIFTY 50 does the job of remembering when the market goes up. It can recall when the market goes up about 98% of the time. This is very good at finding when the NIFTY 50 goes up.. Sometimes when a model is this good at recalling when the market goes up it can predict that the market will go up too much. This makes it not very good at predicting what will happen with information.

The SVM model for NIFTY 50 is also good at recalling when the market goes up.. It has a problem with saying the market will go up when it will not. This is called a positive. The AUC for all the NIFTY 50 models is low 0.47 to 0.52. This means the models are not very good at telling the difference between when the market will go down.

The Random Forest model for NIFTY 50 the kNN model for NIFTY 50 and the Voting Ensemble model for NIFTY 50 are more balanced.. They do not do a lot better, than the other models. Overall the NIFTY 50 models do not do a job of finding the patterns in the data. This means we need to find ways to look at the data and make the models work better.

Table 8: Classifications of Nifty Bank

Model	Accuracy	Precision	Recall	F1 Score	AUC
SVM	0.5488	0.5464	0.9876	0.7035	0.5014
RandomForest	0.4815	0.5226	0.5031	0.5127	0.4924
kNN	0.5118	0.5541	0.5093	0.5307	0.524
VotingEnsemble	0.5219	0.5605	0.5466	0.5535	0.5271
StackedModel	0.5354	0.539	0.9876	0.6974	0.485

The results for NIFTY BANK classification are okay not great. The accuracy is between 50% to 54%. This means the models are just a little better than guessing.

* The SVM model is really good at finding the cases with a recall of around 0.98.. It makes many mistakes in predicting positive cases. This shows that the model tends to predict many upward movements.

Precision, recall, and F1 metrics are all better with the Voting Ensemble than other models. This tells us that when we combine these models together, our predictions will be more consistent. The kNN and Random Forest models range from average performance to below average performance on predicting.Stacked Model finds good cases but does not really help with classification accuracy, indicating overfitting.AUC values are approximately 0.48-0.52, indicating the models do not differentiate much between different types of data. Overall, while some of the models provide some level of accuracy for finding trends, they are not of practical use in real-world applications. The classification models for NIFTY BANK require improvement.

Table 9: Classifications of Nifty Financial Services

Model	Accuracy	Precision	Recall	F1 Score	AUC
SVM	0.5068	0.5531	0.7353	0.6313	0.5218
RandomForest	0.5236	0.5924	0.5471	0.5688	0.5187
kNN	0.4764	0.5497	0.4882	0.5171	0.5067
VotingEnsemble	0.4899	0.559	0.5294	0.5438	0.5065
StackedModel	0.5743	0.5743	1.0	0.7296	0.4941

The performance of the financial sector on this index is mediocre, with an overall accuracy of about 47% to 57%. This is little better than chance alone. The stacked model has the best performance and is also the only one of the four that has completely accurate results. However, it does have a number of errors, and one model has had a lot of general consistency between its predictions and reality. While this gives no reason to believe the same will happen with this index going

forward, three of the four models have average classification capabilities with scores indicating they cannot be relied upon to classify correctly.The NIFTY Financial Services models are just not reliable.

Table 10: Classifications of Nifty Auto

Model	Accuracy	Precision	Recall	F1 Score	AUC
SVM	0.4932	0.5314	0.7697	0.6287	0.5676
RandomForest	0.5135	0.5652	0.5515	0.5583	0.5169
kNN	0.4797	0.5342	0.5212	0.5276	0.4918
VotingEnsemble	0.5068	0.5562	0.5697	0.5629	0.4978
StackedModel	0.5236	0.5465	0.8545	0.6667	0.5138

The NIFTY Auto index results are a little better. Still just okay compared to other indices. The accuracy of the models is around 47% to 53%. This means that the models are doing a bit better than guessing. They are still having trouble predicting what the market will do. The Stacked Model is doing the best with an accuracy of 52.36% and a good F1 score of 0.6667. It is also very good at finding trends with a recall of 0.8545. However the NIFTY Auto index models might be predicting many upward movements, which could be a problem.

The SVM model is good at finding trends with a recall of 0.7697 and a balanced F1 score of 0.6287.. It is not very precise and not very accurate overall. The Random Forest model is consistent and balanced with an accuracy of 51.35% and precision and recall values.. It is not doing much better than the other models, which means it is not very good at predicting.

The Voting Ensemble model is balanced. It is not doing much better than the individual models. This means that combining models is not making the predictions much better. The kNN model is not doing well with lower accuracy and F1 score, which means it is having trouble with the complexity of the sectors data. The AUC values are around 0.49–0.57 which means the models are not very good at telling the difference between things.

Overall the NIFTY Auto index models are a little better, than the indices but they are still not good enough to make confident predictions. This means that the models need to be improved and informative features need to be added. The NIFTY Auto index needs models to make good predictions.

Conclusion

The stock market is really hard to predict because it is influenced by a lot of things and it keeps changing all the time. Most of the models we used in this project did not work well they had a lot of errors and were not accurate. The stock market prediction models like kNN and Random Forest worked a little better. They are still not good enough to make real predictions.

The results also show that it is easier to predict if the stock market will go up or down than the exact price but even that is not very accurate. One important thing we noticed is that the stock market can be very volatile which makes it even harder to predict.

Overall, the study shows that while Artificial Intelligence and Machine Learning can be very useful in predicting the stock market, we need models, more data and new techniques to make them more accurate and useful in real life.

References

[1] Rezaei, A., Abdellatif, I., & Umar, A. (2025). *Artificial intelligence and machine learning techniques in stock market prediction: A comprehensive review.*

[2] Mansilla-Lopez, M., Mauricio, R., & Narváez, L. (2025). *Machine learning techniques for volatility prediction in financial markets.*

[3] Chouhan, M. A. (2025). *AI-based stock market prediction in Indian markets.*

[7] Wawer, M., & Chudziak, J. (2025). *Integrating technical analysis with AI for stock market forecasting.*

[8] Kumar, S. (2025). *AI-based intelligent decision systems for financial forecasting.*

[11] Ekundayo, F. (2024). *Economic implications of AI-driven financial markets and big data integration.*

[12] Oyewole, A. T., Adeoye, O. B., Addy, W. A., et al. (2024). *Predicting stock market movements using neural networks.*

[13] Musale, D. S. (2024). *Enhancing stock market predictions through artificial intelligence.*

[16] Ajiga, D. I., Adeleye, R. A., Asuzu, O. F., et al. (2024). *AI techniques in financial forecasting and stock market analysis.*

[21] Kureljusic, M., & Karger, E. (2024). *Forecasting in financial accounting with artificial intelligence: A systematic review.*

[24] Srividhya, Y., Rajesh, C., Srinivas, K. B., et al. (2023). *Stock market prediction compliance using AI and ML.*

[25] Mintarya, L. N., Halim, J. N. M., Angie, C., et al. (2023). *Machine learning approaches in stock market prediction: A systematic review.*

[26] Sonkavde, G., Dharrao, D. S., Bongale, A. M., et al. (2023). *Forecasting stock market prices using machine learning and deep learning models.*

[34] Milana, C., & Ashta, A. (2021). *Artificial intelligence techniques in finance and financial markets: A survey of the literature.*

[38] Singh, M., Jain, K., Singh, D. K., et al. (2024). *Stock price prediction using AI and ML.*

[39] Patel, A., Patel, D., & Yadav, S. (2021). *Prediction of stock market using artificial intelligence.*