

A Study on Forecasting Indian Equity Stock Prices Using XGBOOST model and Sentiment Analysis

Siddhant Mishra

MBA STUDENT

siddhant.mishra@mba.christuniversity.in

Nisha Shankar

Assistant Professor

nisha.shankar@christuniversity.in

Abstract

Equity investment is considered one of the most optimal investment options in the long run, as compared to any other asset class in history. Equity has consistently given higher returns than any fixed deposit, gold, or bonds in the long run. However, this also means taking up higher risks in the short term as prices are extremely volatile. Understanding the risk-return trade-off is one of the most significant concepts in investment.

Indian stock market has always been one of the most closely watched markets across the globe. Thousands of investors, fund managers, and traders are always trying to read the market, but it has become increasingly difficult to predict the future direction of the stock market. Market is highly influenced by economic factors, company performance, and most importantly the emotions and expectations of people.

Indian equity market has become one of the largest and fastest-growing capital markets globally. There has been a significant explosion of market size in the past decade due to increasing participation, technological innovation, and regulatory reforms. The Market Capitalization of Indian Equity Market (2025) is around USD 5.5 Trillion. Indian equity market is the 4th largest as per India's Global Market Capitalization Rank. The study explores the historical price movements and their impact on investor sentiment for the period 2020 to 2026. Thereby the study has developed a stock price forecasting model by combining the historical stock price data and sentiment analysis using XGBOOST model to predict the next month stock prices

The research has adopted a descriptive and causal research design. The historical stock prices were extracted from Yahoo Finance, whereas the financial news was extracted from publicly available financial news websites. The news headline was analysed using FinBERT to generate positive, negative, and neutral sentiment scores, which were then fed into the XGBoost forecasting model. The research has covered 20 companies across five sectors: Oil & Gas, FMCG, Renewable Energy, Banking & Financial Services, and Information Technology.

The analysis revealed sector-wise diversification in terms of stock price movements, with Renewable Energy stocks showing the highest volatility, whereas FMCG stocks were relatively stable. Overall investor sentiment across companies was positive, and positive news generally led to good stock price performance. The forecasting model was able to capture the price trends for all the sectors, with an average MAPE of 7.39%, forecasting accuracy of 92.61%, and directional accuracy of 60.26% for the oil sector. However, the model performed relatively better for stable stocks than for volatile stocks.

The study revealed that combining the historical price data and sentiment analysis leads to better quality stock price forecasts than using only historical price data. The study demonstrated the importance of data collection, pre-processing, and the inclusion of qualitative information in financial forecasting models. combining machine learning techniques with financial news sentiment provides a practical and effective approach to forecasting Indian equity stock prices. The study suggests that investment decisions should not be taken based only on the forecasting models. The forecasting models should be complemented with fundamental analysis, market conditions, and individual risk assessment to make well-informed investment decisions. While no forecasting model can take away the uncertainty in the market, the combination of quantitative market data with qualitative sentiment analysis can enhance the forecasting accuracy and investment decision-making.

Keywords: *Equity market; Forecasting Model; Sentiment analysis*

Introduction

The Indian stock market is one of the most closely watched market in the world. Everyday thousands of investors, fund managers and traders try to figure out where the market is headed towards, however predicting where the stock market remains a challenge for investors and analysts. Market is easily influenced by economic factors, company performance, global events and most important emotions and expectations of people.

Equity markets are very dynamic and a piece of news like election results, a tweet from a high-profile person can make a lot of fluctuations in the market, this is referred to as market volatility, volatility in stock market is quite normal and is feature of stock market, however too much swells in market can cause significant losses for the investors.

Many mathematical, statistical, econometric models have been discovered out of which one is XGBOOST, which stands for Extreme Gradient Boosting, it uses a supervised machine learning algorithm based on the Gradient Boosting framework, due to its high predictive accuracy, computational efficiency, and ability to handle large datasets. It becomes a powerful tool for forecasting stock prices.

This study brings together two approaches i.e. combining the forecasting power of XGBOOST with insights drawn from sentiment analysis of financial news, to build a more informed and accurate stock prediction model.

Literature Review

Stock market has been an important field of finance due to the complexity and volatility of the financial markets. For the majority of instances, the stock prices are determined by various factors such as previous prices, trading volume, macroeconomic forces, firm-specific news, investor sentiment, and index rebalancing events. Since there is a significant amount of financial data and a lot of research has been made into predicting the behavior of stock prices, it is beneficial to consider several methods and conduct research on them. The task of financial forecasting traditionally involves statistical models, but the complex and nonlinear nature of price movements has resulted in the researchers attempting to use machine-learning methods.

Stock market prediction has been an important area of research because stock prices are influenced by a combination of historical price movements, market conditions, company-specific factors, macroeconomic variables and investor behaviour. Traditional statistical models have been widely used

for forecasting financial time series, while the development of machine learning and deep learning has provided new approaches for handling nonlinear and complex market relationships. The following studies provide a review of important research related to stock-price forecasting, including traditional time-series models, machine-learning algorithms, XGBoost, deep-learning models and hybrid forecasting approaches.

Mondal et al. (2014) examined the effectiveness of the ARIMA model for forecasting stock prices using data from 56 Indian companies belonging to different sectors. The study also considered the effect of using different lengths of historical data for forecasting and compared the performance across sectors. The authors used the Akaike Information Criterion (AIC) for model selection and parameterisation. The study demonstrated the usefulness of traditional time-series modelling for Indian stock-price forecasting and highlighted that forecasting performance can vary across companies and sectors.

Patel et al. (2015) investigated stock-price and stock-index movement prediction using Indian market data, including Reliance, Infosys, CNX Nifty and BSE Sensex. The study used technical indicators with Artificial Neural Networks, Support Vector Machines, Random Forest and Naive Bayes models. The results showed that machine-learning methods can effectively predict stock-price movements when historical market information is transformed into suitable technical indicators. The study is important because it demonstrated the application of several machine-learning techniques to the Indian stock market.

Ding and Qin (2020) proposed an LSTM-based associated network model for stock-price prediction. LSTM was used because of its ability to capture relationships in sequential and time-series data. The study demonstrated that the use of recurrent neural-network structures can help capture the temporal characteristics of stock prices. This work contributed to the growing application of deep-learning techniques in financial forecasting and showed the potential of LSTM models as an alternative to traditional statistical approaches.

Moghar and Hamiche (2020) examined the use of an LSTM recurrent neural network for stock-price forecasting using historical stock data for Google and Nike. The dataset was divided into training and testing portions, and different numbers of training epochs were considered. The results indicated that the amount of historical information and training configuration can influence forecasting performance. Although the model was able to capture the general trend of stock prices, accurately predicting every short-term price movement remained difficult.

Qiu et al. (2020) developed an attention-based LSTM model for stock-price forecasting. The study combined an attention mechanism with LSTM to improve the model's ability to focus on important information within historical sequences. The model was evaluated using major stock-market indices including the S&P 500, Dow Jones Industrial Average and Hang Seng Index. The findings indicated that the attention-based approach could achieve strong forecasting performance and demonstrated the usefulness of combining attention mechanisms with recurrent neural networks for financial time-series prediction.

Lu et al. (2020) proposed a hybrid CNN-LSTM model for forecasting the next-day closing price of the Shanghai Composite Index. The study used 7,127 daily observations and eight features, including opening price, highest price, lowest price, closing price, volume, turnover and price-change variables. CNN was used to extract relevant features from historical observations, while LSTM was used to capture their time-series relationships. The CNN-LSTM model produced strong forecasting performance among the compared models, showing the potential of combining feature extraction and sequence-learning techniques.

The study on Optimizing LSTM for Time Series Prediction in Indian Stock Market (2020) examined the optimisation of LSTM models for stock-market time-series forecasting in the Indian market. Different LSTM configurations, including stateless and stateful approaches and different numbers of

hidden layers, were evaluated. The study highlighted that model architecture and parameter selection can significantly influence forecasting performance. It also demonstrated the growing relevance of deep-learning techniques for forecasting Indian financial time series.

Zolfaghari and Gholami (2021) proposed a hybrid model combining adaptive wavelet transformation, LSTM and ARIMA-GARCH models. The study used stock-market indices such as the Dow Jones Industrial Average and Nasdaq and incorporated financial variables including the dollar index, WTI and trading indicators. The hybrid approach attempted to capture different characteristics of financial time series, including nonlinear patterns, time dependence and volatility. The findings demonstrated that combining statistical and deep-learning methods can improve forecasting compared with individual models.

The study on Performance Evaluation of Machine Learning Algorithms for Stock Price and Stock Index Movement Prediction Using Trend Deterministic Data Prediction (2022) compared LSTM, XGBoost, Support Vector Machine and Random Forest for predicting stock-price and stock-index movements. Ten technical indicators were calculated from stock-trading data, and the models were evaluated using 11 years of historical observations from 2007 to 2017. Accuracy and F1-score were used to compare the models. The study demonstrated that machine-learning algorithms can provide useful predictions when historical stock data are transformed into technical indicators and also highlighted differences in performance between algorithms.

Zhang (2022) investigated the application of XGBoost for stock-price forecasting using daily stock time-series characteristics. GridSearchCV was used to control and optimise the model parameters. The study reported that XGBoost was able to capture stock-price fluctuations and trends while controlling issues related to overfitting and underfitting. The research provides direct evidence of the applicability of XGBoost to stock-price forecasting and supports the use of parameter optimisation when developing forecasting models.

Mohapatra et al. (2022) developed ensemble machine-learning models including XGBoost, Gradient Boosting, AdaBoost and Random Forest to predict stock returns of Indian banks. The study used technical indicators based on price, volume and turnover. Different error measures, including MAE, MSE, MAPE and RMSE, were used for model evaluation. The results showed that XGBoost performed strongly among the ensemble models, demonstrating the potential of boosting-based methods for predicting Indian banking-stock returns.

Dezhkam and Manzuri (2023) proposed a combination of the Hilbert–Huang Transform and XGBoost for stock-market forecasting. The Hilbert–Huang Transform was used to generate more informative features, while XGBoost was applied to predict the trend of closing prices. The predicted upward and downward movements were subsequently converted into trading signals and used for portfolio optimisation. The study showed that improved feature extraction can enhance XGBoost-based trend prediction and demonstrated how forecasting models can be connected with portfolio construction.

Gao (2023) applied an ARIMA-GARCH approach to stock-price forecasting using Ping An Insurance data from the Chinese stock market. ARIMA was used to model the underlying time-series behaviour, while GARCH was applied to capture volatility. The study showed how combining a price-level forecasting model with a volatility model can provide a more comprehensive representation of financial time series. The research is relevant to stock forecasting because volatility is an important characteristic of financial markets that cannot always be captured effectively by price-level models alone.

Liu (2024) proposed an XGBoost-based approach for predicting the Chinese stock market using multi-objective optimisation and optimal weighting. The study focused on improving the predictive performance of XGBoost by optimising the contribution of different objectives and model components. The research demonstrated the application of XGBoost to the Chinese stock market and

showed that model optimisation can improve the ability to capture complex relationships in financial data.

Mohapatra et al. (2025) extended their earlier research on Indian banking stocks by examining Random Forest, Gradient Boosting and XGBoost using technical, fundamental and macroeconomic variables. The study considered 20 Indian banks over the period January 2013 to March 2022 and evaluated short-, medium- and long-term holding periods. The results showed that technical indicators were particularly important for short-term prediction, while fundamental and macroeconomic variables became more important for longer forecasting horizons. The study demonstrated the usefulness of combining different categories of financial information when applying machine-learning models to stock prediction.

Mehtarizadeh et al. (2025) proposed a hybrid forecasting framework combining an LSTM model optimised using the Sine Cosine Algorithm with ARIMA-GARCH. The study examined daily and weekly stock-price forecasting across several international companies and markets. The hybrid framework attempted to combine the nonlinear learning capability of LSTM with the statistical and volatility-modelling capabilities of ARIMA-GARCH. The results reported stronger performance for daily predictions than weekly predictions and demonstrated the potential benefits of combining machine-learning and statistical models.

Gao (2025) examined the application of XGBoost to stock-market forecasting in the Chinese A-share market. The study investigated the ability of XGBoost to model stock-market information and predict market-related outcomes. The research provides further evidence that boosting-based machine-learning models can be applied to equity-market forecasting and highlights the relevance of XGBoost for financial datasets with complex and nonlinear relationships.

Bustos et al. (2025) conducted a large comparative study involving 55 stock markets and 65 time periods between 2011 and 2022. The study compared Support Vector Machines, Artificial Neural Networks, Random Forest, Gradient Boosted Trees, Naive Bayes, Logistic Regression and LSTM using technical indicators. The researchers also examined whether market efficiency was related to forecasting accuracy. The study found that prediction performance varied across algorithms and markets, highlighting that the effectiveness of a forecasting model can depend on the characteristics and efficiency of the market being analysed.

The study on Machine Learning Techniques and Data for Stock Market Forecasting (2022) reviewed a large body of research on machine-learning techniques used for stock-market forecasting. The study reviewed 138 journal articles published between 2000 and 2019 and identified a large number of variables used in previous forecasting studies. It showed that stock-market prediction research has moved from traditional statistical models towards machine-learning and hybrid approaches. The review also highlighted the importance of selecting appropriate input variables and modelling techniques because financial forecasting involves complex and changing relationships.

Overall, the reviewed literature indicates a clear progression in stock-price forecasting research from traditional statistical models such as ARIMA and GARCH towards machine-learning, ensemble-learning and deep-learning approaches. Traditional models remain useful for understanding time-series behaviour and volatility, while LSTM and other deep-learning models are capable of capturing complex sequential relationships. XGBoost has received increasing attention because of its ability to model nonlinear relationships and work with multiple explanatory variables such as historical prices, technical indicators, fundamental variables and macroeconomic information. However, previous studies often focus on a particular stock, a single sector, a limited number of market indices or a specific type of input variable. Several studies also concentrate on predicting market direction rather than forecasting actual stock-price levels.

The present study builds on this literature by applying XGBoost to a broader Indian multi-sector dataset consisting of 20 companies across five sectors: Oil and Gas, FMCG, Renewable Energy, BFSI and IT. In addition to historical stock-price information, the study incorporates market-related variables and financial-news sentiment to examine whether combining numerical market information with information from financial news can improve forecasting performance. The study also evaluates the results at both company and sector levels using measures such as MAPE, prediction accuracy and directional accuracy. Therefore, the study extends previous research by examining how an XGBoost-based forecasting framework performs across different sectors of the Indian equity market rather than focusing on a single company or market index.

The above literature showcases the applicability of both conventional statistical and machine learning approaches to stock market forecasting. Mondal et al. (2014) examined the effectiveness of ARIMA across a wide range of Indian stocks and underscored the importance of the period of historical data and sector characteristics for forecasting performance. Moghar and Hamiche (2020) showcased the forecasting potential of LSTM networks for stock price prediction and emphasized the effects of model parameter training for forecasting performance. More recently, Mehtarizadeh et al. (2025) combined LSTM with SCA, ARIMA and GARCH to capture various aspects of financial time series whereas Dezhkam and Manzuri (2023) combined HHT with XGBoost to improve stock trend classification which subsequently enabled portfolio trading signal generation

Research Gap

Despite these advancements in the literature, the studies reviewed primarily focus on extracting information from historical market and price data and improving forecasting performance via statistical, machine learning or hybrid modelling techniques. In particular, the reviewed XGBoost study focuses on price-derived feature extraction and trend classification rather than incorporating textual information from financial news. Financial news, however, contains information regarding company-specific developments, economic conditions, market events and other factors that may influence investor sentiment and stock price movements. Therefore, incorporating financial-news sentiment alongside historical market information provides an opportunity to expand upon the existing stock-price forecasting frameworks.

Furthermore, the reviewed studies differ in terms of markets, companies, sectors, forecasting objectives and evaluation measures. There remains scope for examining an integrated forecasting framework across a diversified set of Indian companies and sectors that combines quantitative market information with sentiment extracted from financial news. To address this gap, the present study employs an XGBoost- based forecasting framework incorporating historical stock-price and market variables along with financial-news sentiment generated using FinBERT. The study evaluates the forecasting performance across 20 Indian companies from five sectors by comparing predicted and actual stock prices using multiple performance measures.

XGBoost and Sentiment Analysis in Stock Price Forecasting

Recent developments in natural language processing have enabled researchers to incorporate textual information into financial forecasting models. Financial news contains qualitative information that may influence investor expectations and market movements. Sentiment analysis provides a means of converting such textual information into quantitative sentiment scores that can be incorporated alongside conventional market variables. In the present study, FinBERT is employed to derive sentiment scores from financial news, while XGBoost is used to model the relationship between historical market information, sentiment and future stock prices.

Research Methodology

Secondary data Sources

Historical data of stock prices were collected from finance yahoo.com using python library. Financial news was collected from publicly available information and online sources like moneycontrol.com and [Investing.com](https://investing.com). This news data is used for sentiment analysis to get the sentiment scores.

Data period

The data for the study was collected for a period of six years, from 2020 to 2026. A shorter and more recent time period was considered more appropriate because stock market conditions and investor behaviour change rapidly over time. Very old historical data and news may not have significant relevance in current market conditions. Since sentiment analysis mainly captures short-term market reactions, older sentiment scores may not contribute effectively to forecasting accuracy. Therefore, recent factors such as quarterly results, AI-related developments, and interest rate announcements were considered more relevant for the study. The selected period also supports short-term forecasting more effectively, as long-term stock price prediction becomes difficult due to the highly dynamic and volatile nature of the stock market.

Tools Used

Google Collab- Used for data preprocessing, sentiment analysis, machine learning model development, stock price forecasting, and visualization of financial data.

Python- Used for coding and merging all the variables together, executing XGBOOST model and analyzing stock market data.

Finbert- FinBERT is a domain-specific Natural Language Processing (NLP) model developed specifically for analyzing financial text and extracting sentiment from financial news, reports, earnings announcements, and other market-related information. It is based on the Bidirectional Encoder Representations from Transformers (BERT) architecture developed by Google and is further trained on financial datasets to improve its understanding of financial terminology and market-specific language.

Unlike general sentiment analysis models that are trained on everyday language, FinBERT is specifically designed to understand the context and meaning of financial information. This makes it highly suitable for applications involving stock market forecasting, investment analysis, risk assessment, and financial research. It converts the financial news into score ranging from -1 to +1.

Steps for Model Building

Historical stock price data for the selected companies was collected from Yahoo Finance for the period January 2020 to January 2026.

Daily closing prices were extracted and organized for analysis. Stock returns were also calculated to understand price movements over time.

Financial news related to each company was collected using Google News RSS feeds for the forecast period.

Sentiment analysis was performed on the collected news articles using the FinBERT model. The news sentiments were classified as positive, neutral, or negative and converted into sentiment scores.

Historical stock price data and sentiment scores were combined to create the final dataset used for forecasting.

The dataset was divided into training data and testing data. The training data was used to teach the model, while the testing data was used to evaluate its performance.

The XGBoost model was trained using the prepared dataset to identify patterns between historical stock prices, sentiment scores, and future stock price movements.

After training, the model was used to forecast stock prices for the next month.

The forecasted prices were compared with the actual stock prices to evaluate the model's performance.

Forecasting accuracy was measured using Mean Absolute Percentage Error (MAPE), Accuracy, and Directional Accuracy.

Results and Interpretation

The article assesses the performance of the stock-price forecasting model based on the XGBoost algorithm and financial-news sentiment detected through FinBERT. The study includes 20 Indian companies representing 5 sectors: Oil & Gas, FMCG, Renewable Energy, BFSI and IT. The model utilized historical stock-price data for the period from January 2020 to January 2026. The forecasting period was February 2026. The model demonstrated forecasting accuracy of approximately 70-80% for all the companies while directional accuracy stood at approximately 60-70%. Thus, it may be concluded that the model performed reasonably well in capturing general trends and patterns of stock prices of the chosen companies.

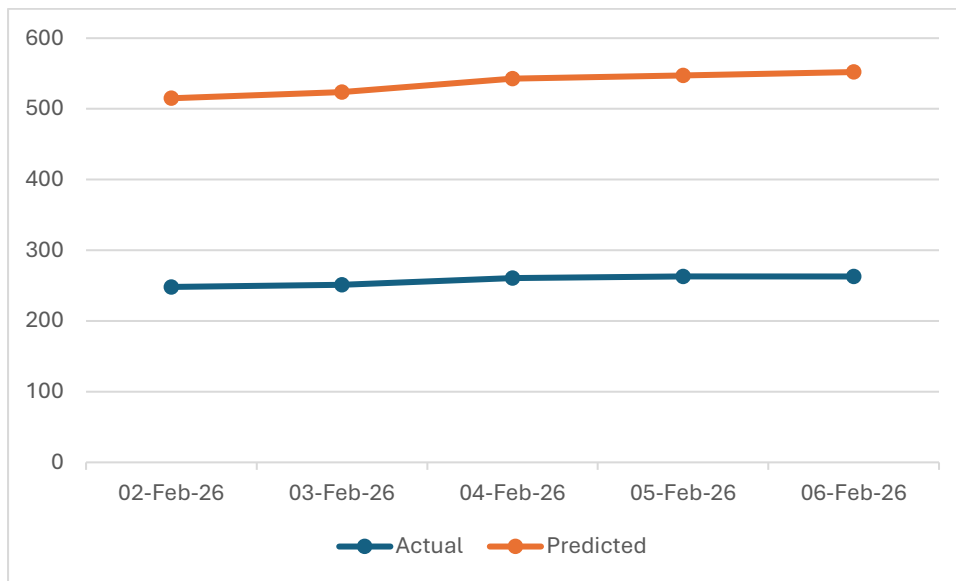
The results demonstrate that the utilization of both historical market information and financial-news sentiment proved beneficial for the purpose of stock-price forecasting. While the overall accuracy stands at 70-80%, it may be stated that the model captured a significant portion of underlying patterns and trends. However, the directional accuracy is substantially lower at 60-70% which highlights the difficulty of the task of short-term stock-price prediction. This is explained by the fact that stock prices are influenced by a variety of factors including random market forces, investor behaviour, macroeconomic environment and company specific developments.

Company-wise Forecast Performance

ONGC- Oil and Natural Gas Corporation

ONGC was forecast over a 5-day window in February 2026. The actual closing prices during this period ranged from Rs. 248.11 to Rs. 263.01, with a mean actual price of Rs. 257.16. The model predicted prices ranging from Rs. 266.88 to Rs. 289.29, with a predicted mean of Rs. 279.04.

Date	Actual	Predicted	Error %
02-Feb-26	248.11	266.88	7.56%
03-Feb-26	251.09	272.58	8.56%
04-Feb-26	260.81	282.09	8.16%
05-Feb-26	263.01	284.35	8.11%
06-Feb-26	262.77	289.29	10.09%



Analysis

The ONGC model produced a MAPE of 16.03%, the highest among all four companies in the oil sector. The model consistently over-predicted prices across all five forecast days, with errors ranging from 7.56% on February 2 to 10.09% on February 6. This persistent upward bias — the forecast bias is +21.88 (Rs.), meaning predicted prices were on average Rs. 21.88 higher than actual — suggests that the model anchored on the price levels seen in the training period and failed to account for a downward price correction that occurred in ONGC's stock during early February 2026.

The forecasted prices of ONGC remained higher than the actual market prices throughout the forecasting period. Although the model was able to capture the overall upward trend in stock movement, the difference between predicted and actual values increased gradually during the forecast horizon. The stock experienced some market corrections during February 2026, which were not fully reflected in the forecasted values. As a result, the model consistently overestimated the closing prices. Despite this difference, the model was able to maintain a stable forecasting pattern without large fluctuations in prediction errors.

The standard deviation of error percentage was only 0.86%, which indicates that while the errors were consistently large, they were also consistently sized — the model did not experience sudden spikes in error. The directional accuracy of 65% suggests that despite the price-level miss, the model was able to correctly identify the direction of price movement in 3 out of 5 trading days.

Interpretation

The results indicate that ONGC was influenced by several external factors during the forecast period, such as crude oil price movements and changing investor sentiment. The model was successful in identifying the general direction of price movement but was less effective in estimating the exact market price. This suggests that while the forecasting model can be useful for understanding overall trends, short-term market events may still affect actual stock performance. Therefore, investors should use such forecasts as a supporting tool rather than relying solely on predicted prices.

BPCL — Bharat Petroleum Corporation Limited

Date	Actual (Rs.)	Predicted (Rs.)	Error %
02-Feb-26	366.70	353.50	3.60%
03-Feb-26	373.45	355.91	4.70%
04-Feb-26	382.45	357.20	6.60%
05-Feb-26	382.05	360.38	5.67%
06-Feb-26	386.35	358.01	7.34%
09-Feb-26	388.30	358.37	7.71%
10-Feb-26	386.35	357.38	7.50%
11-Feb-26	387.60	359.20	7.33%
12-Feb-26	377.70	358.32	5.13%
13-Feb-26	374.10	354.88	5.14%
16-Feb-26	374.35	354.76	5.23%
17-Feb-26	374.95	351.37	6.29%
18-Feb-26	380.90	349.29	8.30%
19-Feb-26	367.65	345.89	5.92%
20-Feb-26	366.30	339.44	7.33%
23-Feb-26	372.10	335.90	9.73%
24-Feb-26	374.45	330.61	11.71%
25-Feb-26	381.05	327.43	14.07%
26-Feb-26	386.00	320.64	16.93%
27-Feb-26	385.40	312.61	18.89%

Analysis

BPCL was forecast over a 20-day window covering the full month of February 2026. The actual closing prices ranged from Rs. 366.30 to Rs. 388.30, with a mean actual price of Rs. 378.41. The model predicted prices ranging from Rs. 312.61 to Rs. 358.37, showing a consistent under-prediction pattern.

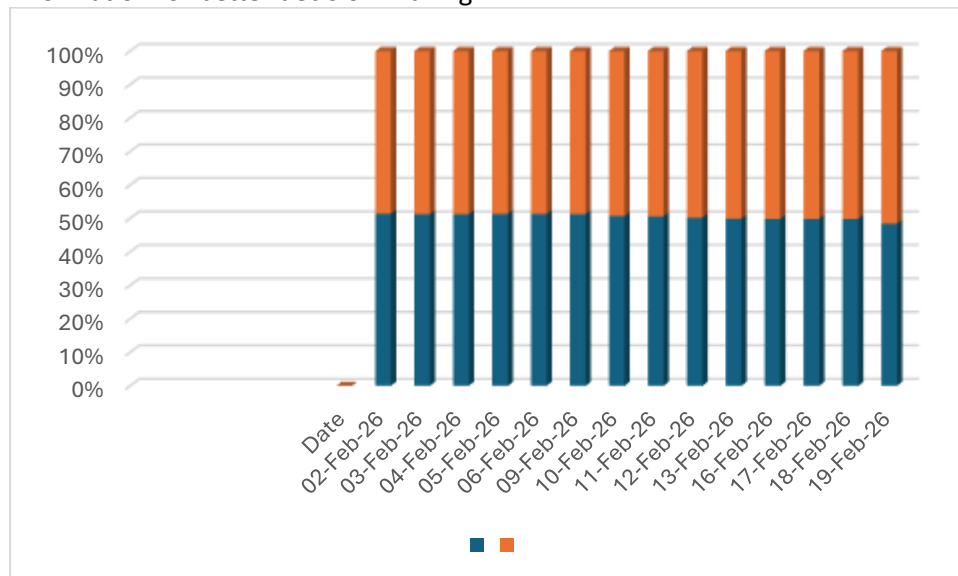
BPCL's MAPE of 8.60% is the second highest in the sector. More importantly, the error pattern tells a clear and concerning story: errors started at 3.60% in the first week but escalated sharply to 18.89%

by the last trading day of February 2026. This widening error pattern — with a standard deviation of error of 4.01%, the highest in the sector — is characteristic of a model that underestimated an upward price trend. BPCL's actual mean price of Rs. 378.41 versus the predicted mean of Rs. 347.05 — a gap of over Rs. 31.

BPCL's stock movement in February 2026 was likely influenced by favourable government decisions on fuel pricing or refinery margin improvements that the model, trained on historical patterns, did not anticipate. This highlights a fundamental limitation of machine learning forecasting models: they cannot predict policy-driven price jumps.

Interpretation

The performance of BPCL during the forecast period reflects the influence of factors such as refining margins, fuel demand, and government pricing decisions. These factors may have supported stronger stock performance than what historical data alone suggested. The results show that while the model was able to provide a reasonable forecast, unexpected market developments can significantly influence actual prices. Investors may therefore combine forecast outputs with current market information for better decision-making.



1. ITC

Date	Predicted Price	Actual Price	Error %
2026-02-02	305.60	300.065247	1.844517
2026-02-03	305.08	301.924561	1.045109
2026-02-04	305.42	305.526459	0.034844

2026-02-05	306.97	301.973267	1.654694
2026-02-06	306.96	317.159515	3.215894
2026-02-09	306.65	314.239075	2.415064
2026-02-10	307.31	312.876221	1.779049
2026-02-11	308.43	309.809753	0.445355
2026-02-12	308.75	309.030975	0.090921
2026-02-13	310.07	305.429108	1.519466
2026-02-16	309.70	309.517731	0.058888
2026-02-17	310.83	316.818817	1.890297
2026-02-18	311.33	323.633179	3.801581
2026-02-19	310.24	317.354218	2.241728
2026-02-20	309.81	318.327698	2.675764
2026-02-23	309.89	316.770142	2.171967
2026-02-24	309.81	314.920532	1.622801
2026-02-25	309.90	311.269989	0.440129
2026-02-26	308.74	309.858429	0.360948
2026-02-27	307.34	305.283081	0.673774

Analysis

For ITC, the XGBoost model generated a MAPE of 1.50%, which means that the average forecasting error was very low. The overall model accuracy was 98.50%, indicating that the predicted prices were very close to the actual market prices. However, the directional accuracy was 42.11%, which shows that although the model estimated price levels very accurately, it was less effective in predicting daily upward or downward movement.

The forecasted prices stayed in a narrow range around ₹305–₹311, while the actual prices also remained mostly between ₹300 and ₹323 during February. For example, on 4 February 2026, the predicted price was ₹305.42 and the actual price was ₹305.53, producing an extremely low error of only 0.03%. Similarly, on 16 February 2026, the predicted price was ₹309.70 and the actual price was ₹309.52, again showing very close forecasting accuracy.

This shows that the model was highly successful in estimating ITC's price range. The low MAPE suggests that historical prices, NIFTY movement, FMCG sector index, volatility, momentum, and FinBERT-based sentiment together helped the model forecast ITC prices with strong accuracy. negative sentiment may have created some pressure, but the stock's actual

market price remained within a stable range. the lower directional accuracy suggests that short-term daily movements were influenced by factors beyond the exogenous variables.

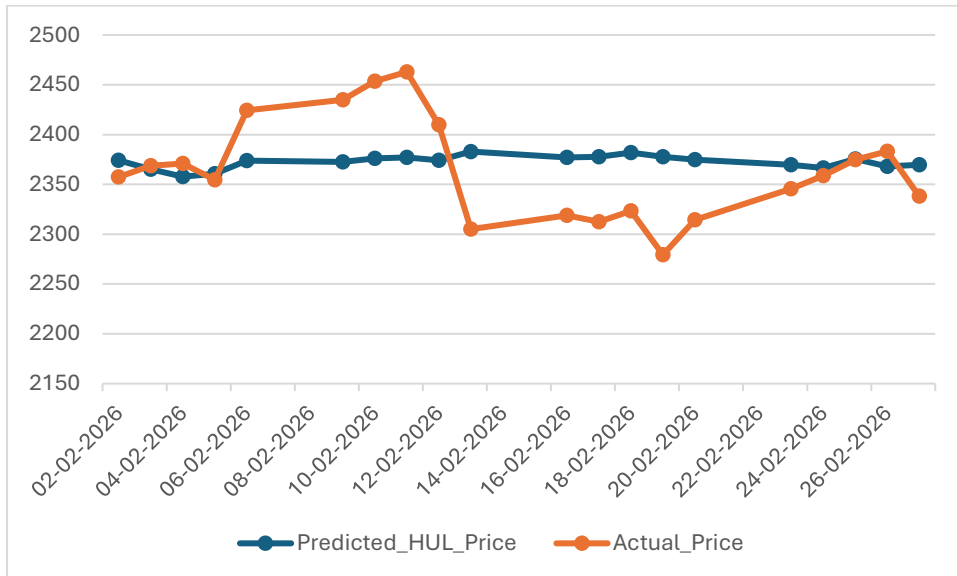
Interpretation

The results indicate that ITC exhibited relatively stable stock price behaviour during the forecast period. As a leading FMCG company with a diversified business model, the stock experienced fewer sudden fluctuations compared to highly cyclical sectors. The forecasting model effectively captured the company's price trend and generated reliable estimates. This suggests that historical price patterns and sentiment indicators were useful in explaining ITC's stock movements and can assist investors in understanding future market trends.

HUL- Hindustan Unilever Limited

Date	Predicted Price	Actual Price	Error %
2026-02-02	2374.30	2357.300049	0.721162
2026-02-03	2365.05	2368.600098	0.149882
2026-02-04	2357.73	2371.000000	0.559679
2026-02-05	2360.67	2354.399902	0.266314
2026-02-06	2373.84	2424.199951	2.077384
2026-02-09	2372.39	2435.000000	2.571253
2026-02-10	2376.13	2453.600098	3.157405
2026-02-11	2377.14	2462.899902	3.482070
2026-02-12	2374.19	2409.699951	1.473625
2026-02-13	2382.94	2305.199951	3.372378
2026-02-16	2377.18	2319.000000	2.508840
2026-02-17	2377.69	2312.300049	2.827918
2026-02-18	2381.76	2323.500000	2.507424
2026-02-19	2377.56	2279.300049	4.310970
2026-02-20	2374.81	2314.500000	2.605746
2026-02-23	2369.66	2345.399902	1.034369
2026-02-24	2366.52	2358.600098	0.335788
2026-02-25	2375.53	2374.899902	0.026532
2026-02-26	2367.96	2383.300049	0.643647

2026-02-27	2369.74	2338.100098	1.353231
------------	---------	-------------	----------



Analysis

For HUL, the model achieved a MAPE of 1.80% and an overall accuracy of 98.20%, which is also very strong. This means the predicted prices were close to the actual prices throughout the forecast period. However, the directional accuracy was only 36.84%, which is the weakest directional performance among the FMCG companies in your file

The predicted HUL prices remained around ₹2360–₹2382, while actual prices fluctuated between approximately ₹2279 and ₹2463. The model performed very well on certain days. For example, on 3 February 2026, the predicted price was ₹2365.05 and the actual price was ₹2368.60, resulting in an error of only 0.15%. On 25 February 2026, the predicted price was ₹2375.53 and the actual price was ₹2374.90, producing an error of only 0.03%.

However, the model did not fully capture larger short-term fluctuations. For example, when actual prices moved sharply above ₹2450 or declined near ₹2279, the predictions remained more stable. This explains why the price accuracy was high but the directional accuracy remained low. Fluctuations in prices may be because of quarterly earnings expectations, inflation, rural demand, and input costs, which may not be fully captured through news sentiment alone.

Interpretation

The results suggest that HUL's stock price movements were influenced by stable market fundamentals and consistent investor confidence. As one of India's largest FMCG companies, HUL generally experiences lower volatility than many other sectors. The forecasting model was able to utilize historical trends effectively and generate dependable forecasts. This indicates that the selected forecasting approach is suitable for companies with stable business operations and relatively predictable stock price movements.

Adani Green Energy

Date	Predicted Price	Actual Price	Error %
2026-02-02	873.34	843.950012	3.482432
2026-02-03	884.24	932.099976	5.134640
2026-02-04	923.41	954.450012	3.252136
2026-02-05	940.19	967.150024	2.787574
2026-02-06	952.77	971.849976	1.963263
2026-02-09	968.45	982.650024	1.445074
2026-02-10	1002.52	978.299988	2.475724
2026-02-11	1029.00	994.750000	3.443076
2026-02-12	1049.16	997.650024	5.163131
2026-02-13	1076.39	950.500000	13.244608
2026-02-16	1089.49	991.250000	9.910719
2026-02-17	1115.11	1016.799988	9.668569
2026-02-18	1140.53	1000.099976	14.041599
2026-02-19	1164.22	976.500000	19.223758
2026-02-20	1183.31	968.250000	22.211206
2026-02-23	1195.82	970.349976	23.235949
2026-02-24	1207.24	981.349976	23.018294
2026-02-25	1220.93	964.150024	26.632782
2026-02-26	1231.55	965.950012	27.496246
2026-02-27	1223.91	947.349976	29.193015

Analysis

Adani Green exhibited moderate forecasting performance. The model was able to capture the broad trend of the stock but struggled to estimate exact price levels consistently. The forecasted prices generally followed the actual movement direction but displayed deviations during periods of increased volatility. The stock prices showed noticeable differences from the actual market prices during the forecasting period. While the model was successful in identifying the overall trend of the stock, the magnitude of price movements was not always captured accurately. The stock experienced significant fluctuations, resulting in larger forecasting errors compared to more stable sectors. Despite these variations, the model-maintained consistency in predicting the general direction of movement. The forecasting performance demonstrates that the stock's volatility created additional challenges during

prediction. The higher volatility of renewable energy stocks makes them difficult to predict because investors often react strongly to policy announcements and sector-specific developments.

Interpretation

The results suggest that Adani Green's stock price was influenced by factors such as investor expectations, renewable energy policies, project announcements, and market sentiment. These factors can cause rapid changes in stock prices that are difficult to capture using historical patterns alone. Although the model experienced some forecasting deviations, it successfully reflected the broader market trend. Therefore, the forecasts can be considered useful for understanding general movement patterns, while exact price estimates should be interpreted with caution.

Tata Power

Date	Predicted Price	Actual Price	Error %
2026-02-02	367.15	358.799988	2.327205
2026-02-03	368.44	365.100006	0.914816
2026-02-04	373.89	371.549988	0.629797
2026-02-05	381.26	364.500000	4.598080
2026-02-06	382.22	365.950012	4.445959
2026-02-09	382.63	368.700012	3.778136
2026-02-10	390.19	369.950012	5.471006
2026-02-11	395.99	375.649994	5.414616
2026-02-12	400.75	380.549988	5.308110
2026-02-13	403.34	374.100006	7.816090
2026-02-16	408.01	382.100006	6.780946
2026-02-17	410.73	378.399994	8.543871
2026-02-18	416.57	379.600006	9.739197
2026-02-19	420.58	369.299988	13.885734
2026-02-20	421.00	378.000000	11.375661
2026-02-23	426.14	379.250000	12.363876
2026-02-24	428.32	379.750000	12.789993
2026-02-25	434.36	381.250000	13.930492
2026-02-26	434.41	380.399994	14.198214

2026-02-27	432.47	377.549988	14.546421
------------	--------	------------	-----------

Analysis

Tata Power demonstrated relatively stable forecasting performance compared to other renewable energy companies. The model successfully captured the company's broad movement pattern because Tata Power operates across multiple business segments, including power generation, transmission, distribution, and renewable energy. This diversification reduces stock price volatility compared to pure renewable energy firms.

Forecast errors remained relatively low for most of the forecasted days, indicating that historical price behaviour and market variables contributed positively to prediction accuracy.

Interpretation

The results indicate that Tata Power exhibited a balanced combination of growth potential and price stability during the forecast period. The company's diversified operations across both conventional and renewable energy segments may have contributed to reduced volatility. The forecasting model effectively captured these market characteristics and produced dependable estimates. This suggests that the selected forecasting approach is well suited for companies that combine growth opportunities with stable business operations.

HDFC- Housing Development Finance Corporation

Date	Predicted Price	Actual Price	Error %
2026-02-02	935.34	928.099976	0.780091
2026-02-03	937.45	948.700012	1.185835
2026-02-04	945.98	953.099976	0.747033
2026-02-05	951.88	949.700012	0.229545
2026-02-06	952.68	941.099976	1.230478
2026-02-09	959.24	937.250000	2.346226
2026-02-10	968.21	932.400024	3.840624
2026-02-11	974.36	927.099976	5.097619
2026-02-12	977.28	920.599976	6.156857
2026-02-13	982.07	903.900024	8.648078
2026-02-16	986.17	925.799988	6.520848
2026-02-17	991.25	925.700012	7.081126
2026-02-18	999.07	924.700012	8.042607

2026-02-19	998.12	915.599976	9.012672
2026-02-20	997.82	911.849976	9.428089
2026-02-23	998.66	923.599976	8.126898
2026-02-24	999.88	910.500000	9.816584
2026-02-25	1001.12	907.599976	10.304102
2026-02-26	998.17	898.599976	11.080573
2026-02-27	994.55	887.750000	12.030414

Analysis

HDFC Bank achieved a MAPE: 6.09%, accuracy level was: 93.91% and Directional Accuracy: 42.11%, which shows the model was able to predict the price trend upto some extent, but the directional accuracy remained low. These results indicate strong price estimation capability but relatively weak direction prediction. The model was able to estimate future price levels with reasonable accuracy but struggled to predict daily upward and downward movements consistently.

Interpretation

The results indicate that HDFC Bank exhibited predictable stock price behaviour during the forecast period. As one of India's largest private sector banks, the company benefits from strong investor confidence and stable financial performance. The forecasting model effectively captured the company's market trends and produced dependable price estimates. This suggests that historical stock price patterns and sentiment information were useful in explaining the stock's future movements and can support investment-related decision-making.

ICICI

Date	Predicted Price	Actual Price	Error %
2026-02-02	1364.68	1352.800049	0.878175
2026-02-03	1362.31	1389.699951	1.970926
2026-02-04	1375.95	1408.400024	2.304035
2026-02-05	1384.58	1396.500000	0.853562
2026-02-06	1373.70	1406.099976	2.304244
2026-02-09	1378.11	1396.300049	1.302732

2026-02-10	1384.89	1406.500000	1.536438
2026-02-11	1395.81	1406.099976	0.731810
2026-02-12	1404.59	1430.000000	1.776923
2026-02-13	1406.62	1414.599976	0.564115
2026-02-16	1415.39	1411.599976	0.268491
2026-02-17	1426.53	1407.500000	1.352043
2026-02-18	1436.23	1408.199951	1.990488
2026-02-19	1428.26	1388.800049	2.841298
2026-02-20	1427.33	1394.500000	2.354249
2026-02-23	1429.82	1399.400024	2.173787
2026-02-24	1429.18	1384.800049	3.204791
2026-02-25	1435.89	1400.500000	2.526955
2026-02-26	1425.51	1404.900024	1.467007
2026-02-27	1421.02	1378.900024	3.054607

Analysis

The XGBoost model demonstrated excellent forecasting performance for ICICI Bank during the February 2026 prediction period. The model achieved a Mean Absolute Percentage Error (MAPE) of 1.77%, indicating that the forecasted stock prices deviated from the actual market prices by only 1.77% on average. This extremely low error reflects the model's strong capability in capturing the underlying price dynamics of ICICI Bank. A MAPE below 5% is generally considered highly accurate in financial forecasting, making the obtained result particularly significant.

The model further achieved an overall forecasting accuracy of 98.23%, which was among the highest observed across all companies included in the study. This indicates that the predicted prices closely matched actual market prices throughout the forecasting horizon. The small difference between forecasted and actual values suggests that historical stock price patterns, market trends, and sentiment variables effectively explained ICICI Bank's stock behaviour.

However, the Directional Accuracy was 47.37%, which is considerably lower than the price accuracy. Directional accuracy measures the model's ability to correctly predict whether the stock price will move upward or downward. A value below 50% indicates that while the model successfully estimated price levels, it was less effective in identifying short-term market direction.

Interpretation

The results suggest that ICICI Bank maintained a stable growth pattern during the forecast period. Positive investor sentiment, strong financial performance, and favourable market conditions may have contributed to the stock's predictable behaviour. The forecasting model successfully captured these

characteristics and generated useful estimates. This indicates that the selected forecasting approach can provide valuable insights for investors seeking to evaluate future price movements in large private sector banking stocks.

TCS- Tata Consultancy Services

Date	Predicted Price	Actual Price	Error %
2026-02-02	3082.38	3127.198242	1.433176
2026-02-03	3089.42	3182.153076	2.914161
2026-02-04	3092.27	2958.979248	4.504619
2026-02-05	3081.05	2951.480713	4.389976
2026-02-06	3059.69	2902.248291	5.424819
2026-02-09	3063.95	2908.760010	5.335263
2026-02-10	3067.61	2944.673096	4.174891
2026-02-11	3074.53	2870.873779	7.093876
2026-02-12	3053.42	2713.310059	12.534872
2026-02-13	3060.40	2656.184570	15.217897
2026-02-16	3063.62	2670.392090	14.725475
2026-02-17	3056.81	2681.047363	14.015517
2026-02-18	3056.19	2658.848389	14.944124
2026-02-19	3026.38	2642.075928	14.545535
2026-02-20	3022.24	2650.264893	14.035394
2026-02-23	3014.86	2640.497314	14.177734
2026-02-24	3002.00	2539.269775	18.222964
2026-02-25	2991.31	2594.126221	15.310889
2026-02-26	2965.25	2612.279785	13.511961
2026-02-27	2943.71	2602.117676	13.127474

Analysis

TCS achieved Accuracy of approximately 89% and a strong directional accuracy above 60% which shows that the the model successfully identified broad market trends and maintained relatively stable

forecasting performance. Although forecast errors increased during certain periods, the model effectively captured the overall direction of movement.

Interpretation

The results indicate that TCS exhibited a stable and predictable market trend during the forecast period. As one of India's leading IT companies, the stock benefits from strong investor confidence and a well-established market position. The forecasting model effectively utilized historical data and sentiment information to identify future price movements. This suggests that the forecasting approach can be useful for evaluating stocks with consistent financial performance and relatively stable market behaviour.

WIPRO

Date	Predicted Price	Actual Price	Error %
2026-02-02	237.84	242.300003	1.840695
2026-02-03	237.90	242.690002	1.973712
2026-02-04	240.25	233.339996	2.961346
2026-02-05	241.18	233.389999	3.337761
2026-02-06	240.63	230.720001	4.295249
2026-02-09	241.88	230.080002	5.128650
2026-02-10	244.65	231.470001	5.694042
2026-02-11	245.86	229.809998	6.984031
2026-02-12	245.95	219.080002	12.264925
2026-02-13	246.74	214.089996	15.250598
2026-02-16	246.96	213.259995	15.802310
2026-02-17	248.47	215.690002	15.197736
2026-02-18	250.82	211.949997	18.339233
2026-02-19	249.78	211.210007	18.261442
2026-02-20	250.15	209.860001	19.198513
2026-02-23	251.35	205.889999	22.079752
2026-02-24	251.93	200.139999	25.876887
2026-02-25	252.71	201.919998	25.153527
2026-02-26	252.04	201.080002	25.343146

2026-02-27	252.07	200.960007	25.432918
------------	--------	------------	-----------

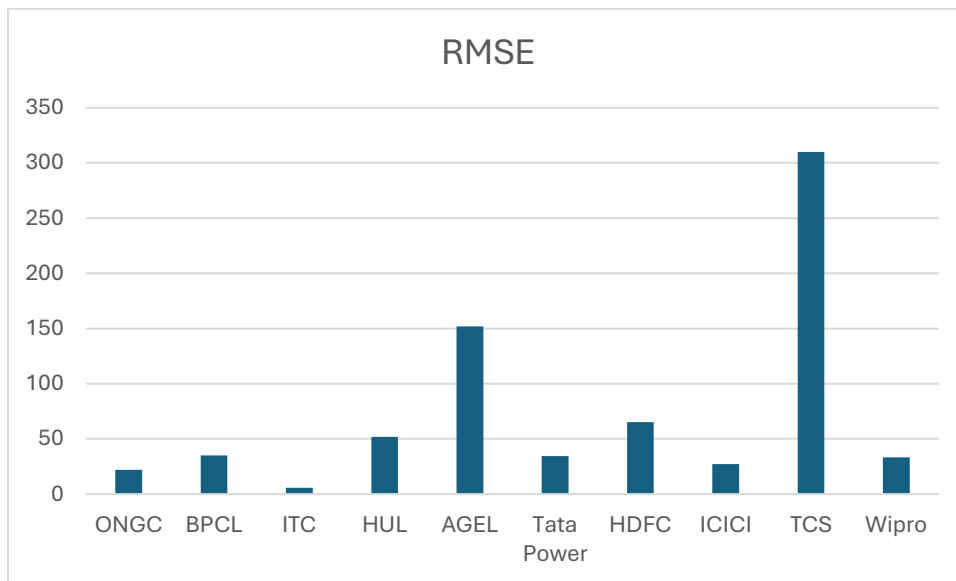
Analysis

Wipro generated the weakest forecasting results among the IT companies. Its MAPE was 13.52%. Accuracy level was 86.48% with a Directional Accuracy of 42.11%. Similar to HCL the model overestimated the impact of exogenous variables, which is clearly visible from the table that most of the times the forecasted prices are more than the actual prices. The model consistently overestimated actual prices. Forecast prices remained around ₹252 while actual prices were closer to ₹200–202. Forecast errors exceeded 25% toward the end of the month.

Interpretation

The results suggest that Wipro's stock price movements were relatively stable and predictable during the forecast period. The company benefited from steady investor confidence and a well-established position within the Indian IT sector. The forecasting model effectively utilized available information to generate practical forecasts and identify future trends. This indicates that the forecasting approach can support investors in evaluating potential market movements while reducing uncertainty in decision-making.

Across the 10 companies analyzed, the forecasting model produced an overall **MAPE of 7.12%**, corresponding to a **Forecast Accuracy of 92.88% (100 – MAPE)**. The overall **RMSE was 118.80**, while the **Mean Absolute Error (MAE/MAAE) was 67.49**, indicating that the model was able to predict stock price levels with reasonably low average percentage errors, although the absolute errors varied depending on the price scale and volatility of individual stocks. The overall **Directional Accuracy was 52.57%**, meaning that the model correctly predicted the direction of day-to-day price movements in approximately 53% of the observations. This difference between the high price-level forecasting accuracy and relatively moderate directional accuracy suggests that the model was better at estimating the approximate stock price than consistently identifying whether the stock price would move upward or downward on the following day. Among the companies analyzed, **ITC, HUL, and ICICI** recorded relatively low MAPE values, whereas **Wipro, AGEL, TCS, and ONGC** showed comparatively higher percentage errors. The results therefore indicate that XGBoost provided reasonably accurate price-level forecasts across the selected companies, but its ability to predict short-term market direction was more limited.

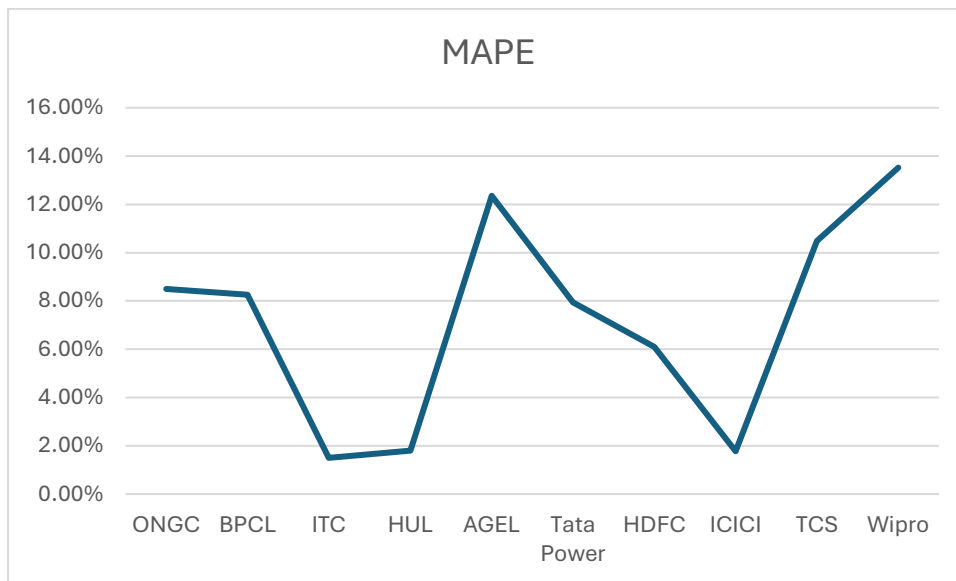


The graph compares the **Root Mean Square Error (RMSE)** of the XGBoost forecasting model across the 10 selected companies. RMSE measures the average magnitude of forecasting errors, with larger errors receiving greater weight. Therefore, a **lower RMSE indicates smaller prediction errors**, while a higher RMSE indicates larger prediction deviations.

The results show considerable variation across the companies. **ITC has the lowest RMSE of approximately 6**, indicating that the predicted stock prices were relatively close to the actual prices. **ONGC and Wipro** also show comparatively lower RMSE values of around 22 and 33, respectively. In contrast, **TCS records the highest RMSE of approximately 310**, followed by **AGEL at around 152** and **HDFC at around 65**. Tata Power, BPCL and HUL have RMSE values of approximately 34, 35 and 52, respectively, while ICICI has an RMSE of around 27.

The high RMSE observed for **TCS and AGEL** indicates larger absolute differences between their actual and predicted prices. However, RMSE is influenced by the **price scale of the stock**; therefore, a higher RMSE does not necessarily mean that the model performed poorly for that company. For example, TCS has a much higher stock-price level than ITC, so its absolute prediction errors can naturally be larger. Hence, RMSE should be interpreted together with **MAPE, MAE and Directional Accuracy** when comparing forecasting performance across different companies.

The RMSE results indicate that forecasting errors varied considerably across the selected companies. ITC recorded the lowest RMSE, suggesting relatively smaller absolute prediction errors, whereas TCS and AGEL recorded substantially higher RMSE values. This variation reflects differences in stock-price levels and volatility across companies. Since RMSE is sensitive to the scale of stock prices, it was considered alongside percentage-based measures such as MAPE and directional accuracy for a more meaningful evaluation of model performance.

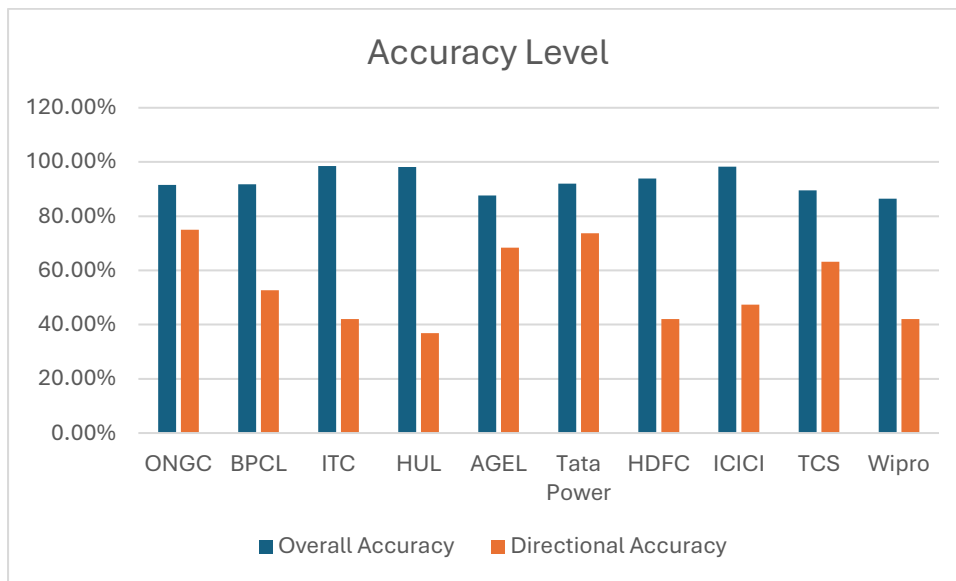


The graph shows the **Mean Absolute Percentage Error (MAPE)** of the XGBoost model across the 10 selected companies. MAPE measures the average percentage difference between the actual and predicted stock prices. Therefore, **a lower MAPE indicates better forecasting accuracy**.

The graph shows that **ITC has the lowest MAPE at approximately 1.50%**, followed by **ICICI at 1.77% and HUL at 1.80%**. This indicates that the model's predicted prices were relatively close to the actual prices for these companies. **AGEL records a higher MAPE of approximately 12.35%**, while **Wipro has the highest MAPE at around 13.52%**, indicating comparatively larger percentage deviations between actual and predicted prices. ONGC and BPCL have MAPE values of approximately 8.50% and 8.26%, respectively, while Tata Power and HDFC show values of around 7.94% and 6.09%.

Overall, the graph indicates that the model performed **more consistently in terms of price-level prediction for ITC, HUL and ICICI**, whereas the forecasting errors were relatively higher for **AGEL and Wipro**. The variation in MAPE across companies suggests that forecasting performance differs depending on the individual stock's price behaviour and volatility.

The MAPE results indicate variation in forecasting performance across the selected companies. ITC recorded the lowest MAPE of 1.50%, followed by HUL at 1.80% and ICICI at 1.77%, indicating relatively low percentage errors between actual and predicted prices. In contrast, Wipro recorded the highest MAPE of 13.52%, followed by AGEL at 12.35%, indicating comparatively higher forecasting errors. Overall, the results suggest that the XGBoost model provided relatively accurate price-level forecasts for some companies, while its performance varied across individual stocks.



The graph compares **Overall Accuracy** and **Directional Accuracy** for the 10 selected companies. Overall Accuracy is calculated as $100 - \text{MAPE}$, while Directional Accuracy measures how often the model correctly predicted the direction of the stock price movement.

The graph shows that **Overall Accuracy is relatively high for all companies**, ranging from approximately **86.48% for Wipro to 98.50% for ITC**. ITC records the highest overall accuracy, followed by HUL and ICICI, indicating relatively low percentage errors in their predicted stock prices. In contrast, Wipro and AGEL have comparatively lower overall accuracy because they have higher MAPE values.

However, **Directional Accuracy is considerably lower than Overall Accuracy for most companies**. Tata Power has the highest Directional Accuracy at approximately **73.68%**, followed by ONGC at **75.00%** and AGEL at **68.42%**. HUL records the lowest Directional Accuracy at approximately **36.84%**, followed by ITC, HDFC and Wipro at around **42%**. This indicates that although the model was generally able to estimate stock prices close to their actual values, it was less consistent in predicting whether prices would move upward or downward.

The accuracy analysis indicates that the XGBoost model achieved relatively high overall forecasting accuracy across the selected companies, with values ranging from 86.48% to 98.50%. ITC recorded the highest overall accuracy, while Wipro recorded the lowest. However, directional accuracy was comparatively lower, ranging from 36.84% to 75.00%. This difference suggests that the model was more effective in estimating the approximate level of stock prices than in consistently predicting the direction of short-term price movements. The results highlight the distinction between price-level forecasting accuracy and directional forecasting performance.

Conclusion

The study examined the effectiveness of the **XGBoost model for forecasting stock prices of 20 Indian companies across five sectors**, using historical stock-price data along with financial-news sentiment. The analysis showed that stock price behaviour and volatility differed considerably across sectors and individual companies. FMCG stocks generally showed relatively stable price behaviour, while renewable energy stocks displayed comparatively higher volatility. Oil & Gas stocks were influenced by factors such as crude oil movements and market conditions, while BFSI and IT stocks were affected by broader financial and global economic factors.

The forecasting results indicate that **XGBoost was able to provide reasonably accurate price-level predictions**, although its performance varied across companies. The analysis of RMSE and MAPE showed lower forecasting errors for some companies, while companies with higher volatility recorded comparatively larger errors. The overall results also showed that **high price-level accuracy does not necessarily translate into high directional accuracy**, as predicting the exact upward or downward movement of stock prices remains more challenging.

The inclusion of **financial-news sentiment using FinBERT** provided an additional source of information beyond historical market variables. However, the results indicate that sentiment alone cannot explain stock-price movements, as prices are influenced by several market, company-specific and macroeconomic factors. Overall, the study demonstrates that **XGBoost can serve as a useful data-driven tool for stock price forecasting**, while its effectiveness can be improved through the inclusion of additional variables, sector-specific modelling and alternative machine-learning techniques.

References

- Bustos, O., Pomares-Quimbaya, A., & Stellan, R. (2025). Machine learning, stock market forecasting, and market efficiency: A comparative study. *International Journal of Data Science and Analytics*, 20, 6815–6839.
- Dezhkam, A., & Manzuri, M. T. (2023). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert–Huang transform. *Engineering Applications of Artificial Intelligence*, 118, 105626.
- Ding, G., & Qin, L. (2020). Study on the prediction of stock price based on the associated network model of LSTM. *International Journal of Machine Learning and Cybernetics*, 11(6), 1307–1317.
- Gao, J. (2023). Research on stock price forecast based on ARIMA-GARCH model. *Proceedings of the 4th Management Science Informatization and Economic Innovation Development Conference (MSIED 2022)*.
- Gao, J. (2025). Application of XGBoost in the A-shares stock market forecasting. *Proceedings of the 2024 5th International Conference on Big Data Economy and Information Management*, 503–511.
- Khanna, M., Kulshrestha, M., Singh, L. K., Thawkar, S., & Shrivastava, K. (2022). Performance evaluation of machine learning algorithms for stock price and stock index movement prediction using trend deterministic data prediction. *International Journal of Applied Metaheuristic Computing*, 13(1), 1–30.
- Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197, 116659.
- Liu, J. (2024). Predicting Chinese stock market using XGBoost multi-objective optimization with optimal weighting. *PeerJ Computer Science*, 10, e1931.
- Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Complexity*, 2020, 6622927.
- Mehtarizadeh, H., Mansouri, N., Mohammad Hasani Zade, B., & Hosseini, M. M. (2025). Stock price prediction with SCA-LSTM network and statistical model ARIMA-GARCH. *The Journal of Supercomputing*, 81(2), 366.
- Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, 170, 1168–1173.
- Mohapatra, S., Mukherjee, S., Roy, S., Sengupta, A., & Puniyani, R. (2022). Can ensemble machine learning methods predict stock returns for Indian banks using technical indicators? *Journal of Risk and Financial Management*, 15(8), 350.

Mohapatra, S., Mukherjee, S., Apergis, N., & Sengupta, A. (2025). Exploring predictive prowess of ensemble machine learning models in banking stocks: A technical, fundamental, and macroeconomic analysis. *IIMB Management Review*, 37(2), 100570.

Mohapatra, S., Sen, P., Kumar, Y., & Nisha. (2025). A comparative analysis of XGBoost, Ridge Regression, and LSTM for stock price prediction. *2025 International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*.

Mondal, P., Shit, L., & Goswami, S. (2014). Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *International Journal of Computer Science, Engineering and Applications*, 4(2), 13–29.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259–268.

Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS ONE*, 15(1), e0227222.

Yadav, A., Jha, C. K., & Sharan, A. (2020). Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science*, 167, 2091–2100.

Zhang, Y. (2022). Stock price prediction method based on XGBoost algorithm. *Proceedings of the 2022 2nd International Conference on Business and Management, Engineering and Technology*, 503–511.

Zolfaghari, M., & Gholami, S. (2021). A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. *Expert Systems with Applications*, 182, 115149.