

# Impact of News Sentiments on Stock Price Prediction

Archit A. Vohra<sup>1, a)</sup>, Dr. Paresh Tanna<sup>2, b)</sup>

<sup>1,2</sup>*School of Engineering, RK University, Rajkot, India.*

<sup>a)</sup> *Corresponding author: avohra147@rku.ac.in*

<sup>b)</sup> *paresh.tanna@rku.ac.in*

**Abstract.** A significant amount of research is carried out on integrating historical data and sentiment analysis to predict stock prices. This study proposes DTR-SA approach to predict the next day's high price for the randomly selected stocks from the NIFTY 50 Indian stock market index. The approach first calculates sentiment scores based on the news data collected from various news portals. The Decision Tree Regression (DTR) algorithm next combines historical stock prices with a sentiment score to predict the high price for the next day. Results indicate that sentiment analysis of news data along with historical stock prices significantly improves the prediction accuracy and hence the Mean Absolute Percentage Error (MAPE) value. The MAPE value of all stocks considered by this study is better compared to the previous historical results.

**Keywords:** Sentiment Analysis, Decision Tree Regression, Machine Learning, Prediction Accuracy.

## INTRODUCTION

Stock price prediction started by making use of machine learning algorithms like Autoregressive Integrated Moving Average (ARIMA), DTR, Support Vector Regression (SVR), Long Short-Term Memory (LSTM), Neural Network Multi-layer Perceptron (NNMLP) and many more techniques. Since the stock data is non-linear, new algorithms and new approaches improved the prediction accuracy of stock prices. However, even at present none of the approaches is 100% accurate. A large amount of text data became available and the use of digital devices increased, its impact on the human mind also increased. Stock and finance-related text data are mostly available through Twitter, blogs, news portals and other social media platforms.

Certain news headlines on stocks, finance or politics; blogs written by influential financial analysts caused the stock index or certain stocks to move upwards or downwards drastically. This led to the correlation between social media data and the stock market. Moreover, this led to a new area of study known as Sentiment Analysis. It is a Natural Language Processing technique that is used to first analyze whether the text data is positive, neutral or negative. This was initially a classification technique used to mine text data. This technique was further improved to convert the positive, neutral or negative tag to a sentiment score between -1 and 1. Thus this quantitative numeric data can be used with many regression-based algorithms to predict some future values.

Many studies have been carried out on Twitter and other news data to predict the stock price. These studies make use of only social media text data and not historical stock data. As time went on to improve the accuracy of predicted values, historical stock data was combined with social media data.

This research work first calculates the sentiment score based on news data related to stocks, finance and politics as these subjects influence the stock market the most on the next day. The 'nltk' python library is used to calculate

sentiment scores from the collected news data. This study uses the DTR machine learning algorithm to predict the next day's high price based on sentiment score and historical stock data.

### RELATED WORK

Investors earning from the stock market are always keen to know future stock prices in advance. Hence, significant research is already carried out for predicting stock prices. Past research has proved that social media sentiments are correlated to future stock price movement. The study [1] uses LSTM that considers Empirical Mode Decomposition (EMD), sentiment index and attention layer to predict stock prices. The study considers only one stock AAPL. Along with historical data of 5 years, the sentiment index is calculated based on stocktwits.com – only one portal is considered. The proposed model results in an impressive MAPE value of 0.0165.

HiSA-SMFM model again uses LSTM to combine historical and sentiment data as proposed in [2]. Only one stock of Tata Motors is considered and historical data of 1 year is collected. Sentiment data is collected from Twitter. Accuracy of 97.18% resulting in a MAPE value of 0.0282 is achieved by the study. The research work presented in [3] first uses Convolutional Neural Network (CNN) for sentiment analysis and then LSTM to combine historical data and sentiment analysis. One day ahead closing price is predicted for 5 stocks selected from different industries. The proposed hybrid model of CNN-LSTM results in the best MAPE value of 0.0405.

The study [4] compares three models GRU, CNN and LSTM to predict the next-day close price of the Nepal stock exchange index. Here sentiment score is based on news data, fundamental as well as technical indicators are included. LSTM returns the best MAPE value of 0.5826 which is significantly high. Sentiment score based on news articles and Twitter is directly collected from Bloomberg in [5]. The research indicates that the stock price prediction can be significantly improved by making use of sentiment score. The research [6] uses Artificial Neural Network (ANN) to combine sentiment scores and historical prices. The study suggests that explicit sentiments improve the performance of ANN while predicting the closing price direction.

The prediction model proposed in [7] uses Naïve Bayes for sentiment analysis and K-NN to predict the future trend of rise and fall of stock prices. K-NN integrates sentiment scores based on financial news and historical stock prices. The best accuracy of 89.8% is obtained when the sentiment score is used by K-NN. The future trend of stock buy or sell was predicted by [8]. News and Twitter data were considered to calculate sentiment scores. The various deep learning algorithms use both basic stock indicators and sentiment scores. The study reveals that sentiment score is related to the next day's share value. The work in [9] uses LSTM to prove that collective emotional states are correlated to the stock transaction data over time. Basic stock indicators of six stocks from the Shanghai stock exchange were used in the experiment. The study indicates that financial data predictions can be improved by using public opinion along with stock data.

Six different models are implemented in [10] to predict stock prices. Five years of stock data from the Hong Kong stock exchange and news articles from FINET were considered in the study. It gives an important conclusion that sentiment analysis improves prediction accuracy. The research [11] uses StockTwits for sentiment analysis. The study uses 9 months of data for five stocks. It concludes that sentiment analysis improves the stock price direction prediction accuracy. Sentiment analysis and neural network techniques are combined in [12] to predict the stock faith recommendation. Stock faith enables the investor to buy or sell stocks. The study [13] uses Nave Bayes and Random Forest to classify tweets and calculate sentiment scores. Next linear regression is used to build a prediction model. Experiments indicate that sentiment score helps in improving prediction accuracy.

Deep learning models Recurrent Neural Network (RNN) and LSTM are used to improve stock price prediction accuracy by making use of financial news articles in [14]. Five years of stock data for Apple was considered in this study. RNN-LSTM that included sentiment score performed best with a MAPE value of 0.0203. A rare result of a study [15] states that there is a negligible effect of sentiment analysis on the stock price prediction. An ensemble model was proposed that combines tweets about Infosys along with basic and technical stock indicators.

## RESEARCH METHODOLOGY

### Research Data

The historical stock data was extracted from www.nseindia.com making use of the nsepy python library. Historical data of certain randomly selected stocks are collected from 1-Jan-2000 for the DTR model. Here, the basic stock price indicators – open, low, close, previous close and Volume-Weighted Average Price (VWAP) are considered.

Web scraping is carried out to extract news data from RSS links. Reputed news portals like moneycontrol, economic times, mint, investing, business-standard etc. were used to scrap news data. News data is scraped from 1-Jun-2022 onwards. Around 200 news items are scraped on each stock trading day. The news data along with historical data is used for DTR-SA model. News data is collected using ‘BeautifulSoup’ library of Python. Next, the sentiment score is calculated using the ‘nltk’ python library. The range of sentiment scores is between -1 and +1, a negative value indicates negative sentiment about the news and a positive value indicates positive sentiment

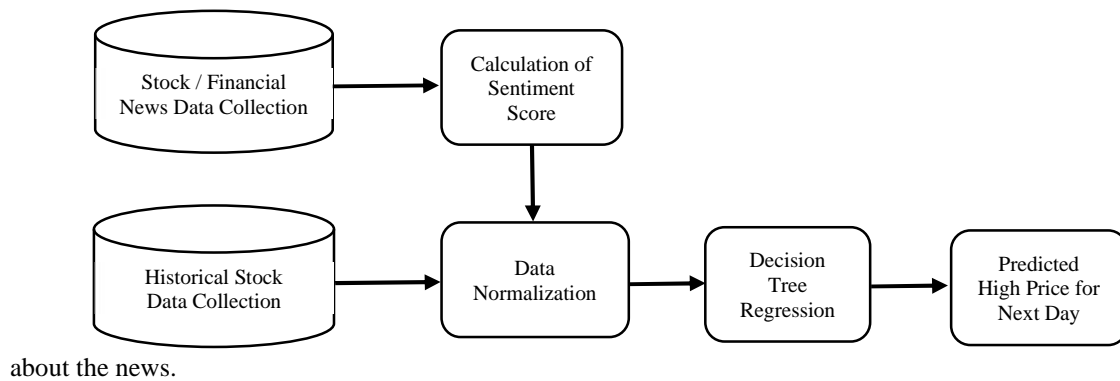


Figure 1. DTR-SA Model.

As part of data pre-processing, normalization of stock prices and sentiment score is carried out using MinMaxScaler from Sklearn python library. Next Decision Tree Algorithm is applied to the normalized data to predict the High stock price for the next day.

### Decision Tree Regression

Regression trees are generally used when the dependent or target variable is continuously valued, i.e., when the problem is prediction type. In a regression tree, a regression model is fit to the dependent variable using each of the independent variables. After this, the data is split at several points for each independent variable considered for study. At each such point, the error between the actual value and predicted values is squared to get ‘A sum of squared errors’ (SSE). The SSE is compared across the variables and the variable or point which has the lowest SSE is chosen as the split point. This process is continued recursively. The biggest advantage of using a regression tree is that it creates a very simple model which predicts the values very quickly. No assumptions are required as the regression trees are non-parametric and non-linear. However, overfitting is one of the biggest problems as far as DTR are concerned.

To validate the performance of DTR-SA and DTR models MAPE (Mean Absolute Percentage Error) is used.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (1)$$

In the above equation  $Y_i$  is the actual high price and  $\hat{Y}_i$  is the predicted high price of the stock under study.

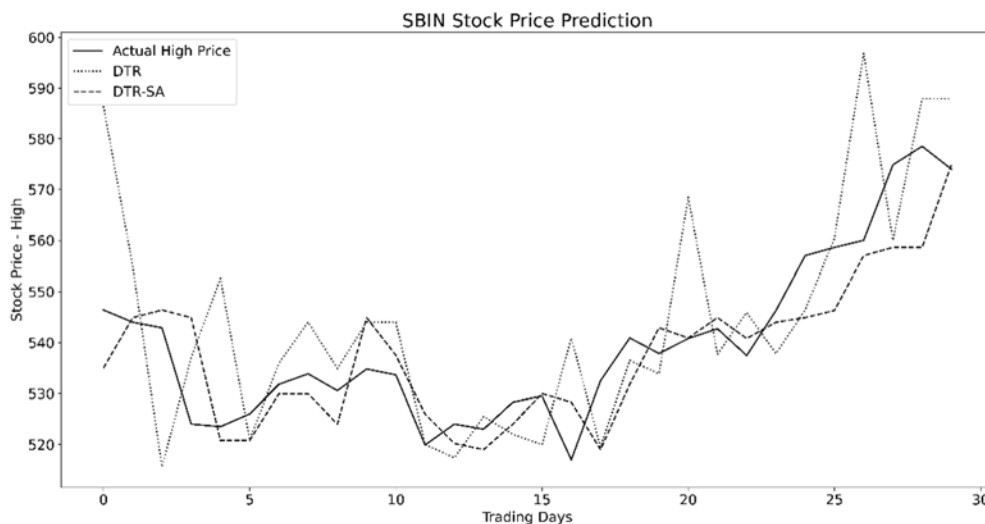
## RESULTS AND ANALYSIS

Some stocks from the NIFTY50 dataset are selected at random. To study the effect of news sentiment on stock price prediction, first the DTR algorithm is applied to only the historical data and results are obtained. Next, the DTR-SA i.e. the same DTR algorithm is applied, but this time to historical and sentiment scores and results are obtained. MAPE is used to compare both results.

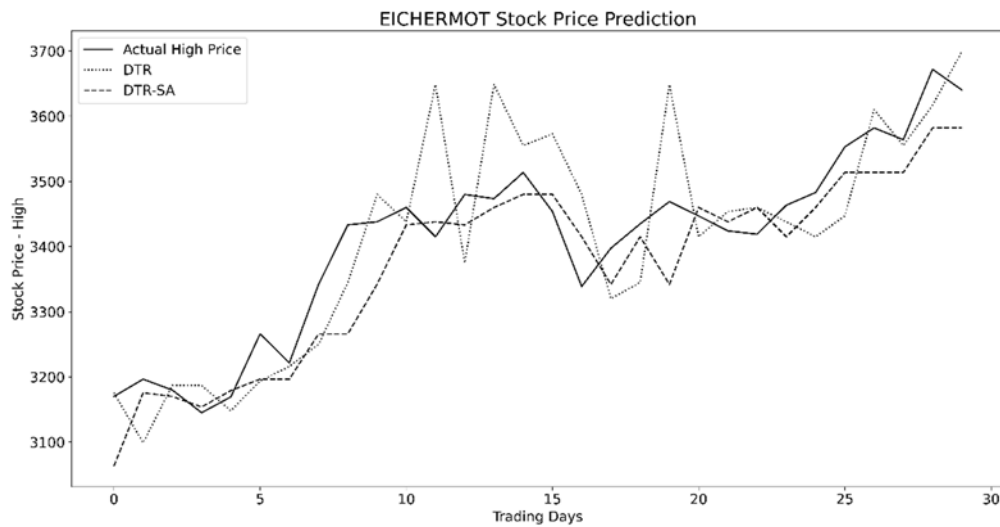
**Table -1:** Comparison of MAPE results.

STOCK NAME	TOTAL TEST RECORDS	DTR MAPE	DTR-SA MAPE	No. of Records with improvement in predicted price.
SBIN	30	0.022	0.012	17
HINDALCO	30	0.021	0.016	16
CIPLA	30	0.0116	0.0089	18
EICHERMOT	30	0.0205	0.0144	19
HEROMOTOCO	30	0.0117	0.0086	19
HDFC	30	0.0163	0.0093	18
BPCL	30	0.0208	0.0148	18
BHARTIARTL	30	0.0166	0.0152	16
HCLTECH	30	0.0142	0.01	18
INFY	30	0.0181	0.0142	16
KOTAKBANK	30	0.0194	0.0119	22
RELIANCE	30	0.0135	0.008	16
BAJFINANCE	30	0.0159	0.0114	18

The MAPE value for all the above stocks improves when DTR includes the sentiment score. Thus, clearly indicating that news has a significant impact on the decision-making ability of investors. Moreover, the obtained MAPE value of all the stocks considered in the study is far better than those achieved by [1], [2], [3], [4] and [15].



**Figure-2:** SBIN Stock Price Prediction



**Figure-3:** EICHERMOT Stock Price Prediction

Two specimen stock price prediction performances are presented in figure-2 and figure-3 respectively. In both the above stocks – SBIN & EICHERMOT the graph also validates the fact that the price predicted using DTR-SA approach that includes news sentiment is far better than the one that uses only historical data.

## CONCLUSION

This study proposes a new DTR-SA approach to predict the stock's HIGH PRICE for the next day. It is a simple approach that utilizes sentiment scores and historical stock prices. The sentiment score is calculated based on news data related to stock, finance and politics collected from news portals. Results clearly indicate that news sentiments have an impact on investors. Hence, including a sentiment score while predicting the next day's high price improves the prediction accuracy. This improvement is validated by both MAPE value and graphs. Moreover, the obtained results when compared with previous studies indicate that DTR-SA simple approach gives better results than other approaches and results obtained so far.

## REFERENCES

- [1] Jin Z, Yang Y, Liu Y., "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9713-9729, 2020.
- [2] Gupta I, Madan TK, Singh S, Singh AK., "HiSA-SMFM: Historical and Sentiment Analysis based Stock Market Forecasting Model," *arXiv preprint arXiv:2203.08143*, 2022.
- [3] Jing N, Wu Z, Wang H., "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, vol. 178:115019, 2021.
- [4] Pokhrel NR, Dahal KR, Rimal R, Bhandari HN, Khatri RK, Rimal B, Hahn WE., "Predicting neps index price using deep learning models," *Machine Learning with Applications*, vol. 9:100385, 2022.
- [5] Vanstone BJ, Gepp A, Harris G., "Do news and sentiment play a role in stock price prediction?" *Applied Intelligence*, vol. 49(11), pp. 3815-3820, 2019.
- [6] Rechenthin M, Street WN, Srinivasan P., "Stock chatter: Using stock sentiment to predict price direction," *Algorithmic Finance*, vol. 2(3-4), pp. 169-96, 2013.

- [7] Khedr AE, Yaseen N., "Predicting stock market behavior using data mining technique and news sentiment analysis," *International Journal of Intelligent Systems and Applications*, vol. 9(7), pp. 22, 2017.
- [8] Mehta P, Pandya S, Kotecha K., "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning." *PeerJ Computer Science*, vol. 7, pp. e476, 2021.
- [9] Zhang G, Xu L, Xue Y., "Model and forecast stock market behavior integrating investor sentiment analysis and transaction data," *Cluster Computing*, vol. 20(1), pp. 789-803, 2017.
- [10] Li X, Xie H, Chen L, Wang J, Deng X., "News impact on stock price return via sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 14-23, 2014.
- [11] Gupta R, Chen M., "Sentiment analysis for stock price prediction," *In 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (IEEE 2020)*, pp. 213-218.
- [12] Pathak A, Shetty NP., "Indian stock market prediction using machine learning and sentiment analysis," *In Computational Intelligence in Data Mining (Springer 2019)*, pp. 595-603.
- [13] Cakra YE, Trisedya BD., "Stock price prediction using linear regression based on sentiment analysis," *International conference on advanced computer science and information systems (ICACSIS) (IEEE 2015)* pp. 147-154.
- [14] Mohan S, Mullapudi S, Sammeta S, Vijayvergia P, Anastasiu DC., "Stock price prediction using news sentiment analysis," *IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService) (IEEE 2019)* pp. 205-208.
- [15] Pasupulety U, Anees AA, Anmol S, Mohan BR., "Predicting stock prices using ensemble learning and sentiment analysis," *IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) (IEEE 2019)* pp. 215-222.