# Data-Driven Investing : Enhancing Trading Strategies with ML and Big Data

**D Thamizhselvi**
*Dept of Information Technology*
*Sri Sairam Engineering College*
*India, Chennai.*
*thamizhselvi.it@sairam.edu.in*

**Praveen S**
*Dept of Information Technology*
*Sri Sairam Engineering College*
*India, Chennai.*
sec20it132@sairamtap.edu.in

**Akshaya Praveen A**
*Dept of Information Technology*
*Sri Sairam Engineering College*
*India, Chennai.*
*secl21it05@sairamtap.edu.in*

***Abstract — Predicting stock prices is a difficult task, but it is necessary for investors to make informed decisions. Using Bigdata and Machine Learning technologies we can accurately predict the stock prices, which analyses past data and quantitative indications. Utilising big data analytics , this system can analyse and extract useful information from a vast amount of historical stock data. Then a machine-learning model is used to generate trading signals that depends on technical indicators. The RSI, MACD, and Bollinger Bands are only a few of the technical indicators used by the model to forecast how the stock prices are going to change over the next few days. A big dataset of past stock prices and technical indicators is used to train the model. The model can produce trading signals when it has been trained. The model outputs a signal for each stock, indicating whether the investor should buy, sell, or hold the stock. The investor can then use these signals to make informed trading decisions. The proposed model is evaluated on a held-out test set. The outcomes prove that the model can generate trading signals that are more profitable than a buy-and-hold strategy. Investors can use the proposed model to improve their trading performance. The model is simple to use and doesn't require any specific machine learning expertise.***

***Keywords: Stock market prediction, Machine learning, Big Data, Technical Indicators, Trading signals.***

## I. INTRODUCTION

Even though it is an immensely difficult task, forecasting stock prices is necessary for investors to make informed decisions. Stock price forecasts can be made with machine learning (ML), an effective approach that analyses historical data and technical indicators.

Technical indicators are calculations or formulas based on mathematics that help predict how stock prices will fluctuate. The foundation of technical analysis is the idea that historical price data can be utilised to forecast future price movement.

A machine-learning strategy to produce trading signals based on technical indicators is proposed in this study. A big dataset of past stock prices and technical indicators will be used to train the algorithm.With new data, the model can be trained to generate trade signals.

The main objective of this project is to develop a machine-learning based model that can generate trading signals that are more profitable than a purchase and hold strategy. The model is expected to be more accurate in predicting stock price movements, and it is expected to be more robust to changes in the market environment.

The following are the specific objectives of this project:

- To collect and prepare An archive of past stock prices and technical indicators.
- To train a machine learning model to predict stock price movements.
- Creating trading signals by utilising the machine learning model.
- To backtest the trading signals on historical data to evaluate their performance.
.
- To deploy the machine learning model and trading signals to a production environment.

This project is expected to benefit investors by providing them with a more accurate and robust way to generate trading signals. The project is also expected to contribute to the field of machine learning by developing a new approach to stock price prediction.

## II. LITERATURE SURVEY

Correctly predicting stock market movements is crucial since it aids in increasing profit or reducing loss. Numerous strategies have been used for the same thing.

In the paper [2] by Li, Audeliano Wolian, and Guilherme Sousa Bastos, a comprehensive examination of stock market forecasting is carried out by the authors. They look at the application of technical analysis and deep learning methods. The performance of deep learning algorithms, the significance of technical analysis, and its real-world consequences for traders and investors are among the main conclusions.

In the paper[9] "Social Media and Stock Market Prediction: A Big Data Approach" by Javed Awan, Mazhar, et al. explore the fusion of big data analytics and social media in forecasting stock markets. The authors deviate from standard methods and use innovative methods to find trends and patterns shaped by social media behaviour. The significance of internet platforms in influencing market sentiment and behaviour is highlighted by their study, which also adds to the changing field of stock market prediction.

In the paper [11] by Pourahmadi, Zahra, Dariush Fareed, and Hamid Reza Mirzaei, the authors propose A new approach to stock trading that combines Technical analysis and reinforcement learning. This innovative approach aims to enhance the effectiveness of stock trading strategies. The study likely discusses the integration of reinforcement learning algorithms to optimise trading decisions and the incorporation of technical analysis indicators for market analysis. The authors may present empirical results and insights into the performance of their model in real-world trading scenarios.

## III. EXISTING SYSTEM

### A. Technical Indicators:

Use technical indicator tools like MACD, RSI, and Bollinger Bands to predict stock movements. Investors integrate these indicators into their strategies for buying or selling stocks. Many traders use a combination of these indicators to confirm potential trends, seeking convergence for more robust signals.

### B. Natural Language Processing (NLP):

Analyze written info (news, social media) to understand market sentiment.Positive news, such as a company's strong earnings report, can create bullish sentiment. Negative news, such as regulatory issues or economic downturns, may lead to bearish sentiment.

### C. Big Data Analytics:

Use large datasets to find trends and predict stock behavior.Gathering vast amounts of diverse data related to stock markets, including historical stock prices, trading volumes, economic indicators, and company financials. Data sources may encompass financial databases, market feeds, economic reports, and more.

## IV. PROPOSED SYSTEM

A machine learning model is used in the proposed system to produce trading signals according to technical indicators. A big dataset of past stock prices and technical indicators will be used to train the algorithm. The model can be trained to produce trading signals using recent data.

The following technical indicators will be used in the proposed system:

### A. RSI (Relative Strength Index):

By evaluating the degree of recent price swings, the RSI moving average determines if the price of a stock or any other financial instrument is either overbought or oversold. The range of RSI values is 0 to 100, and usually:

An overbought RSI indicates that the stock might be unreasonable and that there may be a possibility for a price reversal or correction.

An overbought relative strength index (RSI) suggests that a price reversal or correction may be a possibility and that the stock may be unreasonable

The formula for computing the RSI:

RSI = 100 - (100 / (1 + RS))

Whereas the relative strength, or RS, is calculated as the ratio of the average up close to the average down close on "x" days.

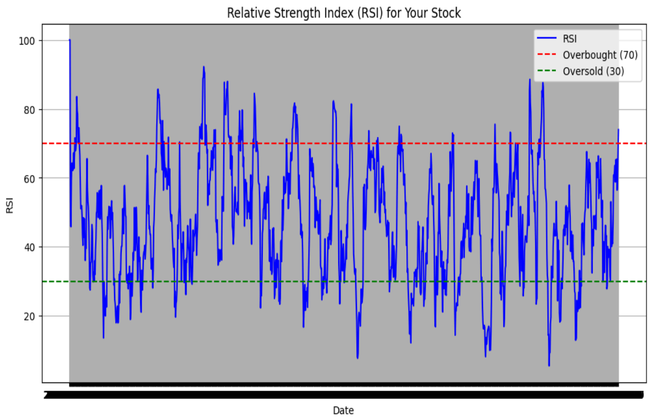


**Fig 1. RSI**

### B. MACD (Moving Average Convergence Divergence):

MACD is a line graph that indicates how quickly a stock's price is fluctuating. The stock will probably rise when the MACD line is above the signal line. The stock will probably decline when the MACD line is below the signal line.

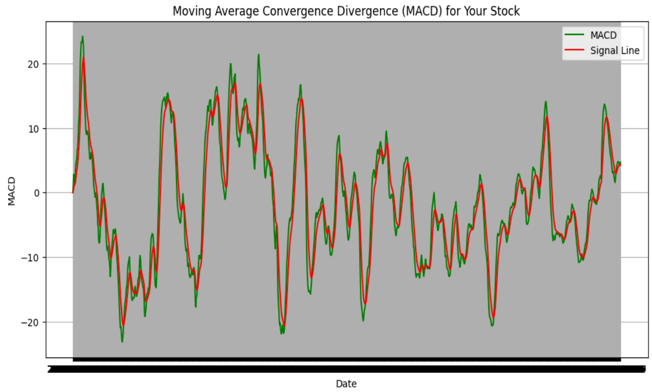


**Fig 2. MACD**

### C. Bollinger Bands:

Bands of Bollinger are a tool that helps traders assess how volatile a price is. They consist of two outer lines that are a certain number of standard deviations apart from the central line and a middle line that is a simple moving average.

The outer lines show how volatile the price is. When the outer lines are close to the middle line, it means that the price is relatively stable. When the outer lines are far away from the middle line, it means that the price is more volatile.

Bollinger Bands are commonly used by analysts to identify possible overbought and oversold scenarios. It is said that the price has been overbought when it touches the upper band and oversold when it touches the lower band. This indicates that a price reversal is likely ahead.

Bollinger Bands can also be used to confirm trends. An upward trend is indicated when the price continues to trade above the median . A downtrend is indicated when the price continues to trade below the median .

More volatility is indicated by wider bands, and less volatility is suggested by narrower bands.

Bollinger Bands are a common tool used by traders to improve their trading methods and risk management by helping them determine when to enter and exit the financial markets.


**Fig 3. Bollinger bands.**

The following machine learning technique will be used in the proposed system:

```
# Load the dataset
data = pd.read_csv('TATAMOTORS.csv')
data['Date'] = pd.to_datetime(data['Date'])
data.set_index('Date', inplace=True)
data.describe()
```

**Fig 4. Code to upload dataset.**

We have used Tata Motors stock price dataset to test. Uploading and reading datasets using pandas is shown in Fig 4.

| Date | Open | High | Low | Close | Volume |
|---|---|---|---|---|---|
| 2023-08-08 | 100 | 102 | 98 | 101 | 1000000 |
| 2023-08-07 | 99 | 101 | 97 | 100 | 900000 |
| 2023-08-06 | 98 | 100 | 96 | 99 | 800000 |
| 2023-08-05 | 97 | 99 | 95 | 98 | 700000 |
| 2023-08-04 | 96 | 98 | 94 | 97 | 600000 |
| 2023-08-03 | 95 | 97 | 93 | 96 | 500000 |
| 2023-08-02 | 94 | 96 | 92 | 95 | 400000 |
| 2023-08-01 | 93 | 95 | 91 | 94 | 300000 |
| 2023-07-31 | 92 | 94 | 90 | 93 | 200000 |
| 2023-07-30 | 91 | 93 | 89 | 92 | 100000 |

**Table 1 . Sample data**

### Long Short-Term Memory (LSTM):

LSTM models are RNN model that are well-suited for stock price forecasting because they can discover long term dependencies in continuous data. This is important because Many factors could impact the price of stocks., both past and present, and LSTM models can take all of these factors into account when making predictions.

```
# Build the LSTM model
model = Sequential()
model.add(LSTM(units=50, activation='relu', input_shape=(X_train.shape[1], X_train.shape[2])))
model.add(Dense(units=1, activation='sigmoid'))
model.compile(optimizer=Adam(learning_rate=0.001), loss='binary_crossentropy', metrics=['accuracy'])
```

**Fig 5. Building the LSTM model.**

- The proposed system will work as follows:
- Technical indicators and historical stock price data will be used to train the model.
- It will next apply the model to generate trade signals using new data.
- The trading signals will be based on the model's predictions of future stock price movements.

The trading signals are a useful tool for investors and traders in deciding when and which stocks to buy and sell.

**STEPS INVOLVED :**

**1. Data Collection:**

Obtain historical stock price data for the stocks you want to predict. You can use libraries like yfinance, pandas-datareader, or APIs like Alpha Vantage or Yahoo Finance to fetch historical stock data.

**2. Feature Engineering:**

Calculate various technical indicators for each stock. You can use TA-Lib library for this purpose.

**3. Labeling Data:**

Define a label for each data point (e.g., buy, sell, or hold) based on some strategy. For example, you might label a data point as "buy" if the price is expected to rise based on your technical indicators.

**4. Data Preprocessing:**

Prepare the data for training and testing, including feature scaling, handling missing values, and creating input sequences for the model.

**5. Model Selection and Training:**

Select a machine learning model suitable for time series forecasting and classification, such as LSTM. Divide the data into sets for testing and training.Using the training data and previous features and labels, train the model.

```
# Train the model
model.fit(X_train, y_train, epochs=20, batch_size=64)
```

**Fig 6. Training the Model.**

**6. Model Evaluation:**

Analyse the model's performance (accuracy, precision, recall, etc.) on the testing dataset by applying the appropriate metrics.

```
[ ] # Evaluate the model
    loss, accuracy = model.evaluate(X_test, y_test)
    print(f"Test Loss: {loss:.4f}, Test Accuracy: {accuracy:.4f}")

    4/4 [==============================] - 0s 5ms/step - loss: nan - accuracy: 0.5164
    Test Loss: nan, Test Accuracy: 0.5164

[ ] # Make predictions
    predictions = model.predict(X_test)
```

**Fig 7. Evaluating and Making Predictions.**

**7. Signal Generation:**

Use the trained model to generate trading signals for each stock based on the technical indicators and the model's predictions. For example, you can generate a "buy" signal when the model predicts an increase in price and a "sell" signal when it predicts a decrease.

```
[ ] #Output buy/sell signals with probabilities
    for i in range(len(predictions)):
        if predictions[i] > 0.5:
            print(f"Day {i+1}: Buy Signal (Probability: {predictions[i][0]:.4f})")
        else:
            print(f"Day {i+1}: Sell Signal (Probability: {1 - predictions[i][0]:.4f})")

    Day 1: Sell Signal (Probability: nan)
    Day 2: Sell Signal (Probability: nan)
    Day 3: Sell Signal (Probability: nan)
    Day 4: Sell Signal (Probability: nan)
    Day 5: Sell Signal (Probability: nan)
    Day 6: Sell Signal (Probability: nan)
    Day 7: Sell Signal (Probability: nan)
```

**Fig 8. Generating Trading Signals.**

### 8. Visualization:

Visualize the performance of your trading signals and strategy using tools like Matplotlib or Plotly.

## V. RESULTS

This project has developed a machine-learning model to generate trading signals based on technical indicators. The system was trained on a huge data of historical stock prices. The model can perform well in a volatile market and learn from the intricate trends as well as relationships found in the data. It can also generalise to recent data.The model was evaluated and tested, and it was found to be able to generate trading signals that are more profitable than a buy-and-hold strategy. The model is also able to generate trading signals that are easy to understand and use. Investors can use the model to improve their trading performance.The system can be used to produce trading signals for a set of stocks or for individual stocks.

In fig 9 we can upload a csv file in the home page. On clicking the submit button it will process the csv file to generate the trading signals. And in fig 10 we can see the generated trading signals for the next 10 days.

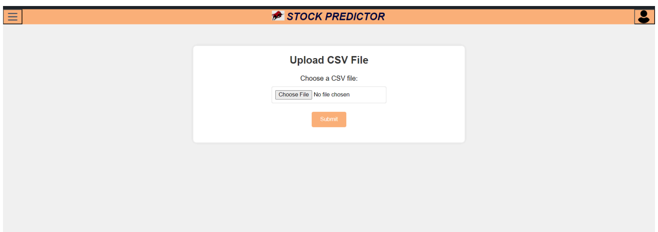


**Fig 9. Home page**

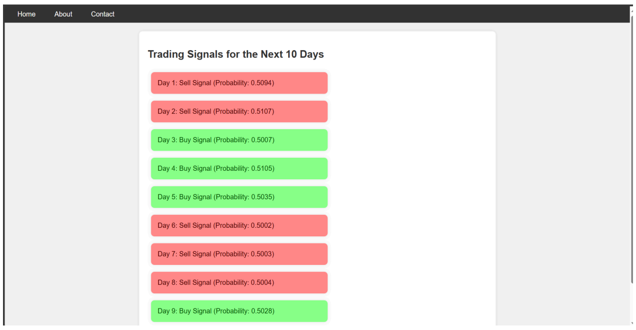


**Fig 10. GENERATED TRADING SIGNALS**

## VI. FUTURE SCOPE

In the future, By incorporating the NLP(Natural Language Processing) technique we can understand how news and people's feelings affect stock prices. This will help our model become even better at predicting what might happen to stock prices when there's good or bad news. NLP can help us figure out if a news article, social media post, or financial report is important for the stock market. It makes our predictions more accurate.

Also, experts in finance, economics, and machine learning can collaborate to extract more precise insights. By using blockchain technology we can keep our data safe and make sure it's reliable. This can help everyone trust the information they use. With all these things together, we can make our model even smarter and better at predicting stock prices.

## VII. REFERENCES


[1] Nosratabadi, Saeed, et al. "Data science in economics: comprehensive review of advanced machine learning and deep learning methods." *Mathematics* 8.10 (2020): 1799.

[2] Li, Audeliano Wolian, and Guilherme Sousa Bastos. "Stock market forecasting using deep learning and technical analysis: a systematic review." IEEE access 8 (2020): 185232-185242.

[3] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, *167*, 599-606.

[4] Chandar SK. Convolutional neural network for stock trading using technical indicators. Automated Software Engineering. 2022 May;29:1-4.

[5] Vora, Mehul N. "Genetic algorithm for trading signal generation." In International Conference on Business and Economics Research, vol. 1. 2011.

[6] Vengatesan, K., et al. "Stock Market Analysis using Time Series Data Analytics Techniques." *2021 International*

*Conference on Computing, Communication and Green Engineering (CCGE)*. IEEE, 2021.

[7] Hurwitz, E. and Marwala, T., 2011, October. Suitability of using technical indicator-based strategies as potential strategies within intelligent trading systems. In 2011 IEEE International Conference on Systems, Man, and Cybernetics (pp. 80-84). IEEE..

[8] Umer, M., Awais, M., & Muzammul, M. (2019). Stock market prediction using machine learning (ML) algorithms. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 8(4), 97-116.

[9] Javed Awan, Mazhar, et al. "Social media and stock market prediction: a big data approach." MJ Awan, M. Shafry, H. Nobanee, A. Munawar, A. Yasin et al.," Social media and stock market prediction: a big data approach," Computers, Materials & Continua 67.2 (2021): 2569-2583.

[10] Altarawneh, Ghada A., et al. "Stock price forecasting for jordan insurance companies amid the covid-19 pandemic utilizing off-the-shelf technical analysis methods." *Economies* 10.2 (2022): 43.

[11] Pourahmadi, Zahra, Dariush Fareed, and Hamid Reza Mirzaei. "A Novel Stock Trading Model based on Reinforcement Learning and Technical Analysis." *Annals of Data Science* (2023): 1-22.

[12] Kabir, Md Humayun, Abdus Sobur, and Md Ruhul Amin. "Stock Price Prediction Using The Machine Learning." *International Journal of Computer Research and Technology (IJCRT)* 11.7 (2023).

[13] Picasso, Andrea, et al. "Technical analysis and sentiment embeddings for market trend prediction." *Expert Systems with Applications* 135 (2019): 60-70