Optimizing Trade Profit Via Rule Based Algorithms After Stock Price Trend Analysis Through Integration of Technical Analysis and Logistic Regression

Amlan Baruah
Department of Business Administration, Gauhati University, Guwahati, India
baruahamlan0@gmail.com

Banajit Changkakati
Department of Business Administration, Gauhati University, Guwahati, India
banajitc@gauhati.ac.in

[Abstract] This study investigates the efficacy of logistic regression in conjunction with rule-based algorithms for optimizing stock price prediction by leveraging technical indicators. The methodology entails the application of logistic regression to predict price movement directions and the implementation of a rule-based algorithm to generate buying and selling signals for the 'ITC' share of National Stock Exchange of India. Analysis of both in-sample and out-sample data reveals a considerable accuracy of 71.4% and 71.2%, respectively, in forecasting price direction. Notably, the rule-based algorithm outperforms the conventional Buy and Hold Model, yielding a superior return. In summary, our research demonstrates a strategy return of 26.65% over a decade between 2013 and 2022 for 'ITC' shares compared to the buy-and-hold strategy. The robustness of the model is rigorously validated, indicating its congruence with observed data and underscoring the potential of predictive analytics in financial markets. This research contributes to the present literature on predictive modeling in finance and offers practical insights for investors looking to refine their trading strategies.

[Keywords] logistic regression, rule-based algorithms, stock price prediction, technical indicators, financial markets, predictive analytics, trading strategies

Introduction

The stock market operates as a complex, non-linear, and nonparametric system, which poses substantial challenges for accurate modeling. Investors endeavor to forecast stock prices and determine optimal trading times. Various studies have utilized fundamental analysis techniques, developing trading rules based on macroeconomic data, industry trends, and company-specific information. According to Reedy (2012) and Kothari (2001), fundamental analysis posits that a stock's price is contingent on its intrinsic value and expected return on investment. This methodology involves a comprehensive analysis of a company's operations and market environment, enabling relatively accurate long-term stock price predictions. However, it is generally accepted that fundamental analysis is less effective for short- and medium-term predictions.

On the other hand, technical analysis has been a broadly utilized method in the financial markets that involves examining historical prices along with the volume data in order to forecast movements of the future price. Technical analysts use indicators which include moving averages, the Relative Strength Index (RSI), and resistance and support levels to help them make trading decisions. They scrutinize historical price patterns as well as trends to pinpoint potential buying and selling opportunities and gauge market sentiment. Despite its popularity, technical analysis has been critiqued for its inconsistency in accurately forecasting future prices.

This study investigates the application of binary logistic regression for predicting the direction of price movements. Alongside price direction forecasting, a rule-based algorithm is implemented to generate 'Buy' and 'Sell' signals. This novel approach aims to assist traders in achieving superior market returns.

Literature Review

Logistic regression is applied in various contexts within stock market prediction, often utilized for estimating the likelihood of certain events occurring, such as price changes or market trends. Its ability to deal with binary outcomes makes it particularly suitable for predicting whether stock prices will go up or down on the basis of several financial indicators. For instance, it can be used effectively to classify stocks into categories based on performance metrics, potentially guiding investment decisions.

Logistic regression is frequently employed for predicting financial outcomes, as evidenced by Ohlson (Ohlson, 1980) & Hassani and Parsadmehr (2012). Ohlson (1980) study applied logistic regression with nine variables for bankruptcy prediction without providing explicit theoretical justification, whereas Hassani and Parsadmehr (2012) focused on the significant differences in financial ratios between solvent and insolvent groups. This comparison underscores the diverse applications of logistic regression in financial forecasting. When employing logistic regression for stock market forecasting, several variables can serve as predictors. These may include financial ratios, historical price patterns, volume of trade, and macroeconomic indicators. By analyzing these factors, logistic regression models can help identify significant predictors that influence stock performance. Studies often use diverse datasets that incorporate various sectors and indexes to enhance the predictive capability of these models (Dutta et al., 2012; Salem et al., 2024).

While logistic regression is a powerful tool for stock market prediction, it is often compared with more complex ML algorithms, like SVM or ANN. These advanced methods might handle nonlinear relationships in data more effectively than logistic regression. Nevertheless, logistic regression provides a more interpretable model and can serve as a reliable baseline for predictions (He & Nawata, 2019; Talpová, 2014). Further extending the scope of logistic regression, Han et al. (2008) integrated it with factor analysis to predict financial distress, using a dataset of 72 firms for model development and 28 firms for validation. Their study achieved an impressive accuracy rate of 92.86%, underscoring the effectiveness of their comprehensive methodology and suggesting its potential superiority over traditional methods. Newer methodologies have combined logistic regression with machine learning techniques to enhance predictive accuracy in stock movement analysis (Badaoui et al., 2023; Yang et al., 2022; Zhang, 2022; Zhou et al., 2018).

Similarly, Dutta et al. (2012) utilized a binary logistic regression model to identify determinants of a company's stock market performance. Their research identified eight financial ratios, including percentage change in net sales and price/earnings per share, as key indicators capable of classifying companies with 74.6% accuracy as per their rate of return. The research highlights the practical utility of these ratios for investors, despite challenges such as multi-collinearity, and suggests that these financial ratios can provide clear insights for predicting favorable stock performance. The authors acknowledge, however, that other economic and managerial factors could also influence share prices beyond the scope of their analysis.

Another notable theme is the use of technical indicators in stock price forecasting. Oriani and Coelho (2016) examined the effectiveness of lagging indicators like the Exponential Moving Average, while Chiu and Chen (2009) suggested a dynamic fuzzy model incorporating a Support Vector Machine (SVM) to analyze stock market dynamics. These studies emphasize the importance of combining various technical indicators with closing price data, illustrating the advancement of methodologies in stock market prediction.

Recent studies increasingly integrate machine learning with technical indicators, representing a cutting-edge trend in financial prediction models. Ayala et al. (2021) evaluated multiple machine learning algorithms (Linear Model, ANN, Random Forests, SVR) using daily trading data to enhance trading signals and competitiveness. Similarly, Wagdi et al. (2023) employed Artificial Neural Networks to predict currency movements against the US dollar, demonstrating the application of advanced ML approaches in financial forecasting. Several other studies show that advanced machine learning algorithms frequently yield better predictive performance compared to logistic regression. For example, ML models have demonstrated superior accuracy in predicting stock price movements and returns, with some studies specifically highlighting improvements over traditional linear methods like logistic regression (Hina et al., 2024; Kaur & Sharma, 2022).

It is possible to modify or enhance logistic regression to make it more accurate in intricate market situations. Feature engineering, which entails modifying and producing new features from existing data to better capture complex relationships, is one popular technique. This approach can enhance the model's predictive power by including interactions and nonlinearities that traditional logistic regression may miss. Applying regularization approaches like L1 or L2 regularization can also enhance logistic regression models. These methods help prevent overfitting by penalizing large coefficients, which is particularly beneficial in cases where the number of predictors is high relative to the amount of data available. Regularization keeps the model simple and robust, thus maintaining generalization to unseen data (Montenegro & Navarrete, 2023; Wijma, 2009).

Collectively, these themes reflect the dynamic nature of research in financial prediction models. Researchers continue to adopt diverse methodologies and advanced technologies to improve predictive accuracy, illustrating the evolving landscape of financial forecasting.

Methodology

The study follows mainly a correlational research design with predictive dimensions. There is no primary data involved in the analysis. The secondary data is not manipulated in any way before the analysis.

Data Collection

The analysis which is fully on secondary data. The dataset includes daily price data for ITC shares from 1996 to 2023, collected from the website ‘www.yahoofinance.com’ obtained using Python's 'YAHOO_FIN' module. The data comprises 'High,' 'Low,' 'Close,' and 'Adjusted Close' prices.

Technical Indicators

The study employs ten technical indicators as independent variables:

1. Simple Moving Average over the last 30 trading periods (SMA 30)
2. Relative Strength Index over 14 periods (RSI 14)
3. Relative Strength Index over 21 periods (RSI 21)
4. Kaufman's Adaptive Moving Average (KAMA)
5. Money Flow Index (MFI)
6. Moving Average Convergence Divergence (MACD)
7. Stochastic Oscillator (SO)
8. Average True Range (ATR)

9. Rate of Change (ROC)
10. Price Volume Trend (PVT)

The values of each technical indicator are normalized using max-min normalization to reduce the impact of scale on the analysis.

Dependent Variable

The dependent variable, 'price signal,' is binary, defined as:

$$\text{Price Signal} = \begin{cases} 1 \; if \; p_t > p_{t-1} \\ 0 \; if \; p_t \leq p_{t-1} \end{cases}$$

Where p_t as current 'Adjusted Close' price and p_{t-1} as previous 'Adjusted Close' price.

Logistic Regression Model

Logistic regression is an essential technique for predicting binary outcomes, such as success versus failure, using a set of predictor variables. Within our study, logistic regression is crucial for evaluating the probability of specific events based on our technical indicators and dataset analysis. The fact that logistic regression does not require assumptions regarding the error terms distribution is one of its advantages.

The core of logistic regression is the binary logistic function, or sigmoid function, which models the probability of event occurrence as a predictor variables function. This function converts real-valued inputs into outputs ranging between 0 and 1, effectively representing probabilities.

The sigmoid function, frequently used in binary classification, underpins logistic regression by transforming real-valued inputs into probabilities. The function is expressed mathematically as:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}}$$

Here, $P\,(Y = 1)$ denotes the probability of event Y occurring, e represents the base of the natural logarithm, β_0 as intercept, and $\beta_1, \beta_2, \ldots, \beta_k$ are coefficients for the predictor variables $X_1, X_2 \ldots, X_k$.

Taking the natural logarithm of both sides, we derive the log-odds or logit function:

$$l\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

The logit function facilitates the interpretation and analysis within logistic regression. The sigmoid transformation ensures the logistic function's output remains between 0 and 1, making it apt for binary classification scenarios where the outcome is dichotomous.

For our study, the logistic model, or logit model, is represented as:

$$logit(p) = l\left(\frac{p}{1-p}\right) = b_0 + b_1SMA30 + b_2RSI14 + b_3RSI21 + b_4KAMA + b_5MFI + b_6MACD + b_7SO + b_8ATR + b_{09}ROC + b_{10}PVT$$

In this model, p signifies the probability of the price direction either rising or remaining stable, while (p/1-p) represents the odds of the price direction falling or remaining stable. The coefficients b_0, b_1, b_2, . . ., b_{10} are determined during the logistic regression modeling process. This model facilitates the interpretation and analysis of the likelihood of changes in price direction, thus connecting technical indicators with predictive modeling.

Assumptions

1. The previous trading day closing price is the same as the current trading day opening price.
2. Opening prices are used for buyer settlements.
3. Only one trade is placed per day.
4. Participants do not allocate more than 10% of their capital to a single transaction.
5. Transaction costs and taxes are ignored.

Trend Determination from Output Trading Signal

Upon the completion of the training process, a novel set of test data has been applied to the trained neural network, resulting in the generation of a corresponding set of output values. The output of the network, referred to as the trading signal (θT_r), is a continuous value ranging from 0 to 1. To form an informed trading decisions, it is crucial to identify trends and determine optimal trading moments. The classification of trends into uptrends and downtrends is achieved by analyzing the output trading signals $\theta T_{ri,}$ according to the following criteria:

$$If\ \theta_{Tr_i} > \frac{1}{n}\sum_{j=1}^{n} \theta_{Tr_j}, the\ predicted\ trend\ is\ up\ (1)$$

Otherwise, the predicted trend is down (0)

Trading Decision Derived from Predicted Trends

Once the direction of stock movement is ascertained, trading points are determined through straightforward trading guidelines as follows:

- If the projected trend for the upcoming day indicates an uptrend, the decision is to buy.
- If a buy decision has been made, then holding is advised.
- If the projected trend for the upcoming day indicates a downtrend, the decision is to sell.
- If a sell decision has been made, then holding is recommended.

Evaluation of Strategy

Profit and loss are calculated to assess the strategy's effectiveness. Metrics include trade profit, total profit, total loss, average profit, average loss, hit ratio, ending capital, and strategy return. These metrics are compared with actual market returns to evaluate model performance. The formulas for each metric are described in Note A.

Hypothesis Testing

This research addressed the following central research question: How effective are technical indicators in predicting ITC share price movements using a logistic regression model, and how profitable is the trading strategy based on these predictions? The hypothesis and the null hypothesis for the research question were as follows:

H_a-1. Technical indicators significantly predict the direction of ITC share price movements when used through logistic regression.

H_0-1. Technical indicators do not significantly predict the direction of ITC share price movements when used through logistic regression.

H_a-2. The proposed rule-based trading strategy yields a return higher than the actual 'Buy and Hold' return of ITC shares.

H_0-2. The proposed rule-based trading strategy does not yield a return higher than the actual 'Buy and Hold' return of ITC shares.

Results

a. Goodness of Fit of Logit Model

Using the PSPP software program, the Hosmer-Lemeshow Test was utilized to find out the model's goodness of fit. The outcomes demonstrate that the logistic regression model fits the observed data satisfactorily, as Table 1 illustrates. The test statistic yields a p-value of 0.439 with a chi-square value of 7.946 and 8 degrees of freedom. This p-value depicts that the model fits the data well because it is greater than the traditional significance threshold of 0.05, which means that there is inadequate evidence to reject the null hypothesis. This suggests that the predicted probabilities from the logistic regression model are reasonably consistent with the actual outcomes.

Table 1

Hosmer and Lemeshow Test Results

Step	Chi-Square	df	Sig.
1	7.96	8	0.439"

b. Logit Model Formulation

Table 2 presents the coefficients, standard errors, and additional details for the model. The p-value (Sig.), derived from the Wald test, indicates the statistical significance of each predictor variable. Typically, a p-value less than 0.05 suggests that the corresponding predictor is statistically significant

Table 2

Variables in the Equation

Step 1a	B	S.E.	Wald	df	Sig.	Exp(B)
SMA30	6.868	1.040	43.579	1	.000	961.262
RSI14	7.465	.236	999.481	1	.000	1746.171
RSI21	-4.873	.195	626.093	1	.000	.008
KAMA	-6.961	1.042	44.645	1	.000	.001
MFI	-.141	.043	10.987	1	.001	.868

MACD	-.508	.048	113.369	1	.000	.602
SO	-1.731	.066	697.388	1	.000	.177
ATR	.155	.060	6.764	1	.009	1.168
ROC	-.108	.054	4.046	1	.044	.898
PVT	.232	.066	12.209	1	.000	1.262
Constant	.001	.028	.003	1	.959	1.001

In-sample Validation of the Model

The Model had been validated by splitting the sample data into 2 parts. Training and Test data. The training and test data are split in the ratio 80:20 by the method of random sampling. Here the performance of the model is shown through the confusion matrix. Confusion matrix is a table used to calculate classification model performance. It shows how accurate and inaccurate the model's predictions were compared to the ground truth. The confusion matrix is explained here:

Table 3

Confusion Matrix

Actual / Predicted	Downtrend (0)	Uptrend (1)
Downtrend (0)	496	193
Uptrend (1)	214	519

Accuracy: The overall percentage of correct predictions is 71.4%.

It has a higher recall for Uptrend (72.0%) than for Downtrend (70.8%), indicating it is better at correctly identifying Uptrend instances. The precision is fairly balanced between the two classes, with 69.9% for Downtrend and 72.9% for Uptrend. These metrics help in understanding the performance of the classification model and identifying areas for potential improvement.

Out-of-Sample Validation of the Model

The model was evaluated utilizing out-of-sample data encompassing the utilization of technical indicators derived from the price action of ’HDFC Bank Limited [Symbol=HDB], a company listed on the NSE, India. The sample data spanned from January 1, 2023, to December 31, 2023, during which the model generated a total of 64 signals. The Classification Report, presented in Table 4, reveals an accuracy rate of 71% for the predicted symbols.

Table 4

Classification Report on Out-Of-Sample Data for HDFC.NS

“Class	Precision	Recall	F1-Score	Support
0	0.66	0.87	0.75	159
1	0.81	0.55	0.66	157
Accuracy	0.71			
Macro Avg	0.73	0.71	0.70	316
Weighted Avg	0.73	0.71	0.70	316”

Trade Simulation and Analysis

Once the model has been adequately trained and tested on both in-sample and out-of-sample data, the trading rules detailed above are applied to the predicted signals (1 and 0). The returns generated by this strategy are then compared to the annual returns achieved by a buy-and-hold strategy for the same asset. According to data from the website *Senthilstocktrader*(n.d.), the annual returns for the stock 'ITC' from 2013 to 2022 were as follows: 12.1%, 14.5%, -11.1%, 10.9%, 9.1%, 7%, -15.7%,-12.1%, 4.3%, and 52.1%. An initial investment of Rs. 10,000 at the beginning of 2013, if held until the end of 2023, would result in a final value of Rs. 16,688.94, representing a cumulative appreciation of 166.89% over ten years.

A trader implementing the rule-based strategy outlined in this study can anticipate specific results over a ten-year period, assuming adherence to the model's assumptions. The outcomes are detailed in Table 5. According to our strategy, the final capital amounts to Rs. 21,137.24, reflecting an appreciation of 211.37% over ten years. This performance demonstrates an advantage of 26.65% over the buy-and-hold strategy during the same timeframe.

Table 5

Trading Summary

*Initial Capital =Rs. 10000.

Year	Trades	Positive Trades	Negative Trades	Avg. Profit	Avg. Loss	Hit Ratio	Ending Capital	Strategy Return
2013	61	44	17	23.59	13.69	0.72	10805.15	8.05
2014	60	42	17	22.04	5.35	0.70	11639.99	7.73
2015	63	48	15	22.60	14.64	0.76	12505.07	7.43
2016	60	42	18	27.55	9.39	0.70	13493.44	7.90
2017	69	47	22	27.94	10.40	0.68	14577.61	8.03
2018	63	41	23	31.71	9.27	0.65	15664.36	7.45
2019	61	42	19	23.33	10.69	0.69	16441.09	4.96
2020	65	45	20	46.91	33.63	0.69	17879.53	8.75
2021	59	36	22	50.67	13.15	0.61	19414.42	8.58
2022	62	50	13	41.93	28.74	0.81	21137.24	8.87

Analyses of Hypotheses

This study sought to explore the relationship (if any) between the technical indicators in predicting ITC share price movements using a logistic regression model, and how profitable is the trading strategy based on these predictions. The analysis of the given technical indicators as independent variables (table 2) found each of them significant at p=0.05 level in the analysis of the price direction of ITC shares. So, we reject the null hypothesis H_0*-1* and accept the alternative hypothesis H_a*-1* that technical indicators significantly predict the direction of ITC share price movements when used through logistic regression. Also as seen from the results of the trading summary in the table 5, we see that the final profit of following our strategy is Rs.21,137.24 which is greater than the return of Rs. 16,688.94 in the Buy and Hold strategy. So, we reject the null hypothesis H_0*-2* and accept the alternative hypothesis H_a-2 that the proposed rule-based trading strategy yields a return higher than the actual 'Buy and Hold' return of ITC shares.

Discussion

This study's objective was to form a robust logistic regression model for predicting ITC share price movements and to compare the performance of a rule-based trading strategy derived from this model against a traditional buy-and-hold strategy. The dataset, spanning from 1996 to 2023, provided a comprehensive basis for analysis, leveraging various technical indicators known to influence market dynamics.

Our methodology utilized a diverse set of technical indicators as independent variables, each selected for its ability to capture different aspects of market behavior. The selection process ensured that only statistically significant indicators were included, enhancing the model's predictive power. The logistic regression model was employed due to its suitability for binary outcome prediction, specifically predicting whether the price would rise or fall.

We assessed the performance of the logistic regression model with both in-sample and out-of-sample data. With a p-value of 0.439, the goodness of fit, as determined by the Hosmer-Lemeshow test, showed that the model adequately explains the data and that there is no compelling evidence to refute its adequacy. This finding gives rise to confidence in the model's strong capability to generalize to new data.

In-sample validation using a confusion matrix revealed an accuracy of 71.4%, with a precision of 69.9% for predicting downtrends and 72.9% for predicting uptrends. The recall rates for downtrends and uptrends were 72.0% and 70.8%, respectively. These metrics highlight that the model performs consistently across both classes, though with a slight edge in predicting uptrends.

The out-of-sample validation further demonstrated the model's reliability. Applied to HDFC Bank Limited's data for 2023, the model achieved an accuracy of 71.20%, with precision and recall rates indicating balanced performance. This consistency between in-sample and out-of-sample validations underscores the model's robustness and generalizability.

The trading strategy based on the logistic regression model's predictions was evaluated by comparing its returns to a buy-and-hold strategy over a ten-year period (2013-2022). The rule-based strategy, guided by the model's trading signals, yielded a final capital of Rs. 21,137.24, reflecting an appreciation of 211.37%. In contrast, the buy-and-hold strategy resulted in a final capital of Rs. 16,688.94, with a cumulative appreciation of 166.89%.

The higher returns from the rule-based strategy highlight its effectiveness in capitalizing on market trends predicted by the logistic regression model. The strategy's success can be attributed to its ability to dynamically adjust to market conditions, as reflected in the higher hit ratio and average profit per trade compared to average losses.

The technical indicators incorporated into the model played a crucial role in its predictive accuracy. Indicators like the SMA, RSI, Kaufman's Adaptive Moving Average (KAMA), and others provided a multi-faceted view of market conditions. Each indicator contributed uniquely to capturing various aspects of price movement, such as trend direction, momentum, volatility, and market pressure.

The significant coefficients from the logistic regression analysis indicated the relative importance of each indicator. For instance, SMA30 and RSI14 had high positive coefficients, suggesting their strong influence on predicting price increases. Conversely, negative coefficients for indicators like RSI21 and KAMA indicated their association with price declines. This nuanced understanding of technical indicator contributions helps us in refining trading strategies and improving predictive models.

Several assumptions underpinned the analysis, such as ignoring transaction costs and taxes, assuming consistent opening and closing prices, and limiting capital allocation per trade. While these assumptions facilitated model development and analysis, they also introduced limitations. Real-world trading involves costs and taxes that can affect profitability. Additionally, market conditions such as liquidity and trading volume variations could impact strategy performance. The exclusion of these factors suggests that the model's performance in real-world scenarios might differ. Future research could enhance the model by incorporating these elements, providing a more comprehensive and realistic evaluation of trading strategies.

Conclusions

The study successfully developed a logistic regression model using a comprehensive set of technical indicators to predict ITC share price movements. The model's predictive accuracy and robustness were validated through in-sample and out-of-sample tests, demonstrating its reliability in different market conditions. The rule-based trading strategy derived from the model outperformed the traditional buy- and-hold strategy, highlighting its potential for achieving higher returns.

The findings underscore the importance of leveraging technical indicators and logistic regression for developing effective trading strategies. The study also acknowledges the limitations and assumptions inherent in the analysis, suggesting avenues for future research to address these aspects and enhance model performance. Overall, the research contributes valuable insights into the application of statistical models in financial markets, offering a promising approach for traders and investors to navigate market complexities and optimize returns.

Contribution

This study contributes to financial market prediction by developing a robust logistic regression model to forecast ITC share price movements using a diverse set of statistically significant technical indicators. Demonstrating consistent performance in both in-sample and out-of-sample validations, the model's reliability and generalizability are established. A rule-based trading strategy based on the model outperformed a traditional buy-and-hold strategy over ten years, indicating its effectiveness in dynamic market adaptation and profit maximization. Insights from the technical indicators enhance the understanding of market behaviors, aiding in the refinement of predictive models and trading strategies. This research provides a valuable framework for future studies incorporating real-world trading costs and conditions to improve the practical applicability of predictive models in financial markets.

Future Research

Future work could focus on expanding the range of technical indicators, exploring ML approaches like neural networks and SVM, and incorporating macroeconomic factors to enhance predictive accuracy. Additionally, applying the model to different stocks and market conditions could further validate its generalizability and robustness

References

Ayala, J., García-Torres, M., Noguera, J. L. V., Gómez-Vela, F., & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge-Based Systems*, *225*, 107-119. https://doi.org/10.1016/j.knosys.2021.107119

Badaoui, M. E., Raouyane, B., Moumen, S. E., & Bellafkih, M. (2023). Impact Machine Learning Classification and Technical Indicators, Forecast the Direction of Bitcoin. *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 1–7.

Chiu, D. Y., & Chen, P. J. (2009). Dynamically exploring internal mechanism of stock market by fuzzy-based support vector machines with high dimension input space and genetic algorithm. *Expert Systems with Applications*. https://www.sciencedirect.com/science/article/pii/S0957417407005817

Dutta, A., Bandopadhyay, G., & Sengupta, S. (2012). Prediction of stock performance in the Indian stock market using logistic regression. *International Journal of Business and Information*, *7*(1), 105.

Han, D., Ma, L., & Yu, C. (2008). Financial Prediction: Application of Logistic Regression with Factor Analysis. *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, 1–4. https://doi.org/10.1109/WiCom.2008.2308

Hassani, M., & Parsadmehr, N. (2012). The Presentation of Financial Crisis Forecast Pattern (Evidence from Tehran Stock Exchange). *International Journal of Finance and Accounting*, *1*(6), 142–147.

He, H., & Nawata, K. (2019). *The Application of Machine Learning Algorithms in Predicting the Borrower's Default Risk in Online Peer-to-Peer Lending*. https://api.semanticscholar.org/CorpusID:203647606

Hina, M., Jayani, S., Dipika, M., & Patel, K. (2024, April 4). *PREDICTING STOCK PRICES THROUGH MACHINE LEARNING TECHNIQUES AND SENTIMENT ANALYSIS*. https://api.semanticscholar.org/CorpusID:270261567

Kaur, R., & Sharma, A. (2022). *Machine Learning Algorithms for Prediction of Blue-Chip Stocks*. https://api.semanticscholar.org/CorpusID:251581894

Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, *31*(1), 105–231. https://doi.org/10.1016/S0165-4101(01)00030-1

Montenegro, C., & Navarrete, R. (2023). S&P 500 Index Value Forecasting Using Decision Fusion Regression Model. *2023 10th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 22–26.

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131. https://doi.org/10.2307/2490395

ONGC Share Price History from 1996 to 2024—Senthil Stock Trader—Senthilstocktrader.com. (n.d.). https://senthilstocktrader.com/ongc-share-price-history/#: :text=ONGC%20Shares%3A%20A%20Remarkable%20Growth%20from%201996%20to%202024&text=Impressive%20Returns%3A%20Investors%20saw%20a,have%20increased%20to%20%E2%82%B9407565.22.

Oriani, F. B., & Coelho, G. P. (2016). Evaluating the impact of technical indicators on stock forecasting. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8.

Reedy, V. (2012). Intrinsic Value Estimation Through Fundamental Analysis–A Case Study of Dr. Reddy's Laboratories Ltd., Hyderabad. *SSRN Electronic Journal, March 2012*.

Salem, A.-B. M., Yurynets, R., Yurynets, Z., Konieczny, G., & Kolisnichenko, P. (2024). Forecasting the Dynamics of Cryptocurrency Rates Based on Logistic Regression. *International Conference on Computational Linguistics and Intelligent Systems*. https://api.semanticscholar.org/CorpusID:269188765

Talpová, S. Ž. (2014). *Logistic Regression: An Option for a Management Research?*

https://api.semanticscholar.org/CorpusID:61901372

Wagdi, O., Salman, E., & Albanna, H. (2023). Integration between technical indicators and artificial neural networks for the prediction of the exchange rate: Evidence from emerging economies. *Cogent Economics & Finance*, *11*(2), 2255049.

Wijma, K. E. (2009, February). *Improving logistic interactions between Technomed Europe and Technomed Asia: Shifting responsibilities*. http://essay.utwente.nl/60435/

Yang, J., Wu, J., & Li, X. (2022). Penalized logistic regressions predict stock price trends. *Other Conferences*. https://api.semanticscholar.org/CorpusID:251845884

Zhang, X. (2022). Research on Credit Risk Forecast Model Based on Data Mining Technology. *Proceedings of the 8th International Conference on Computing and Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:250497005

Zhou, F., Qun, Z., Sornette, D., & Jiang, L. (2018). Cascading Logistic Regression Onto Gradient Boosted Decision Trees to Predict Stock Market Changes Using Technical Analysis. *Econometrics: Econometric & Statistical Methods - Special Topics eJournal*. https://api.semanticscholar.org/CorpusID:125294394

Notes

Note A: Profit and Loss Calculation

To evaluate the effectiveness of the proposed trading strategy, it's essential to assess its profitability and risk management. It is calculated through a series of formulas to ultimately arrive at the strategy return. The return is compared with the actual YoY return provided by the asset in the real market to gauge the efficiency of our model.

1. **Trade Profit:** This metric quantifies the profit gained from individual trade. It's computed as the difference between the selling price and the buying price, normalized by the initial investment.

$$\text{trade profit} = \left(\frac{\{\text{sell price}\}}{\{\text{buy price}\}} - 1\right) \times \{\text{initial investment}\}$$

2. **Total Profit:** The total profit represents the cumulative profit generated from positive trades. It's calculated as the sum of trade profits for all positive trades.

total profit = trade profit (for positive trades)

3. **Total Loss:** The total loss accounts for the cumulative losses incurred from negative trades. It's calculated as the sum of absolute trade profits for all negative trades.

total loss = abs(trade profit (for negative trades)

4. **Average Profit:** This metric provides insights into the typical gains per trade. It's calculated by dividing the total profit by the number of positive trades.

$$\text{average profit} = \frac{Total\ Profit}{Positive\ Trades}$$

(if positive trades > 0, else 0)

5. **Average Loss:** Similarly, average loss quantifies the typical losses per trade. It's calculated by dividing the total loss by the number of negative trades.

$$\text{average loss} = \frac{Total\ Loss}{Negative\ Trades}$$

(if negative trades > 0, else 0)

6. **Hit Ratio:** The hit ratio measures the proportion of successful trades relative to the total number of trades. It indicates the model's accuracy in predicting profitable trading opportunities.

$$\text{hit ratio} = \frac{Positive\ Trades}{Total\ Trades}$$

(if trades > 0, else 0)

7. **Ending Capital:** This metric represents the final capital amount after accounting for all profits and losses incurred during the trading period.

$$\text{ending capital} = \text{capital} + \text{total profit} - \text{total loss}$$

8. **Strategy Return:** The strategy return quantifies the percentage change in capital over the trading period. It's calculated relative to the previous year's ending capital.

$$\{\text{strategy return}\} = \left(\frac{\{\text{ending capital}\} - \{\text{previous year ending capital}\}}{\{\text{previous year ending capital}\}}\right) \times 100$$