Algo-Trading Optimization using Supervised Learning based on Decision Tree

Sanabam Bineshwor Singh, bineshworsanabam@gmail.com; Mayank Raj,starkbloodraven@gmail.com ;
Sweta Chaudhary , swetaachaudhary14@gmail.com

Department of Computer Science and Engineering, National Institute of Technology Manipur

Abstract Computer and Machine Learning advancements have offered new prospects for enhancing trading processes. Algorithmic trading combines computer programming and financial markets to execute trades at precise moments. It attempts to strip emotions out of trades, ensures the most efficient execution of a trade, places orders instantaneously, lowers trading fees, increases market liquidity, and standardizes trading by removing human emotions from the equation. During the research, we analyzed various trading strategies according to market nature and their yield. In previous research, the dynamic nature of the market and changes due to it was not considered, but we have rigorously analyzed it and noted the subtle changes that it accounts for. In our research, we collected data from different exchanges through customized APIs, and then for the modeling of algorithms, we used Machine Learning Algorithms where we implemented the Decision Tree using different trading strategies giving them proper weightage. Our statistical analysis suggested we got more favorable outcomes than previous works on the same matter and it gave a new perspective to think about when it comes to Algo trading.

Keywords: Machine Learning, Algorithmic Trading, APIs, Decision Tree, Back Testing, Statistical Analysis

1. Introduction

Algorithmic trading (also called Automated trading, Black-box trading, or Algo- trading) uses a computer program that follows a defined set of instructions (an algorithm) to place a trade. The trade, in theory, can generate profits at a speed and frequency that is impossible for a human trader. The defined sets of instructions are based on timing, price, quantity, or any mathematical model. Apart from profit opportunities for the trader, algo-trading renders markets moreliquid and trading more systematic by ruling out the impact of human emotions on trading activities. The relevance of automated stock picking/trading algorithms in the financial sector has lately increased. Economic agents can by digesting information from various traditional and unconventional sources combine such knowledge with traditional financial indicators to generate significant returns. Big Data has enhanced the roles of new, powerful technologies, artificial intelligence (AI), and data science (DS), not just in finance but also in many other fields, including cybersecurity, marketing, economics, and many more. The most common trading algorithms favor the use of past stock market prices as a tool for buying and selling in the financial markets. Among this research, we highlight scholarly works that incorporate data derived directly from financial time series with supervised/unsupervised machine learning techniques. A genetic algorithm, for instance, was suggested by Allen and Karjalainen (1999) to develop technical trading rules that are directly related to the level of returns and volatility. More specifically, this analysis suggests that an investor should enter the market when there are positive returns and low daily volatility and avoid it when there are negative returns and high volatility. The key components for obtaining a trading signal are the volatility and the returns, even if Allen and Karjalainen (1999) mentioned the possibility of expanding their technique by integrating additional forms of information such as fundamental and macroeconomic data.

1.1 The Benefits of Algorithmic Trading

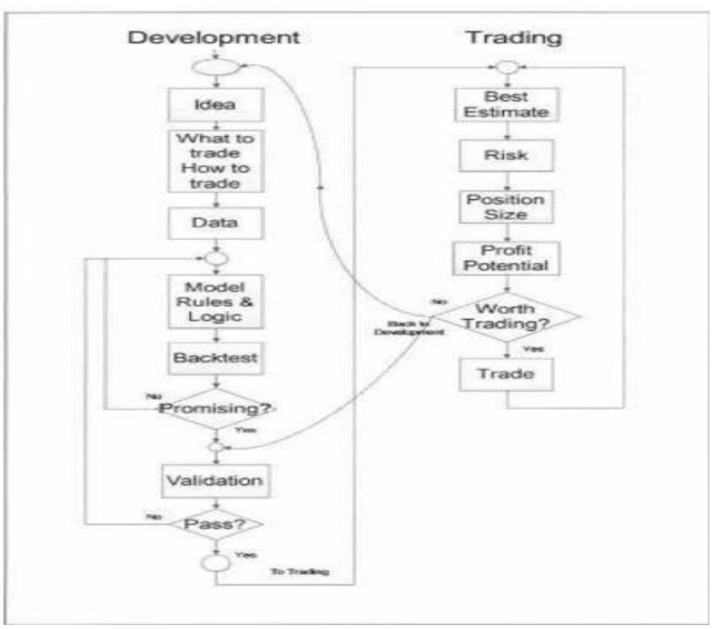
- Trades are executed at the best possible prices.
- Trade order placement is instant and accurate (there isa high chance of execution at the desired levels).
- Reduced transaction costs.
- Simultaneous automated checks on multiple marketconditions.
- Reduced risk of manual errors when placing trades.
- Algo trading can be back-tested using available historical andreal-time data to see if it is a viable trading strategy.
- Reduced the possibility of mistakes by human traders based on emotional and psychological factors. Algo-trading today is high-frequency trading (HFT), which attempts to capitalize on placing many orders at rapid speedsacross multiple markets and multiple decision parameters based on preprogrammed instructions.

2. Literature Survey

Algorithmic trading was first thought about sometime between 1980 and the present. Nearly all the literature on algorithmic trading describes what it means. One of the most widely cited authors in this field, Hendershott (2011), defines AT as the use of computer programmers to carry out market orders. Additionally, he claims that AT "automatically makes trading judgments" and is more than simply about execution (Hendershott, et all2011) Peter Gomber relates to the same definition, but he gives it a fresh perspective. Algo trading, according to him, operates in "real-time." Another benefit is that a trader can access numerous markets through AT that others have not yet explored (Gomber,2014) Nearly all the authors emphasize the idea of little human involvement in AT. The SEBI's suggested definition in its discussion paper emphasizes the same connotation. High-Frequency Trading, or HFT, is a notion that has perhaps gained as much popularity as AT. HFT is viewed by some authors as a component of AT and as one of the trading strategies that AT includes. The International Organization of Securities Commissions (IOSCO) provides the most thorough explanation of HFT in its Consultation Report titled "Technological Challenges to Effective Market Surveillance Issues and Regulatory Tools" (August 2012) Numerous pieces of research show that one salient characteristic of HFT is that it involves frequent order updates and large numbers of transactions. (Gomber,2014). The first authors to investigate whether HFT has an impact on asset prices were Cvitanic and Kirilenko (2010).

In the 2021 survey of long-only funds, the average score of respondents is 5.81 – an increase from both the 2020 score (5.71) and the 2019 score (5.74). In 2021, the most impactful features of algorithms are ease of use, customer support and services, dark pool access, execution consistency, and increased trader productivity. Following high scores of 5.96 and 5.92respectively in the 2020 survey, support services and ease of use both scored 6.01 in the 2021's survey. It is interesting to see the ease of use increase its score year-on-year over the past four years, underlining the importance of usable and streamlined technologyin the modern trading environment. Two categories in this year's survey recorded the joint highest year-on-year increase in their score, anonymity, and Algo monitoring. Both categories received an increase of 0.17, putting anonymity at 5.89 and algo monitoring at 5.72. An increase in trader productivity marks the second-highestjump in the score, having increased by 0.16 from 5.80 to 5.97. Thisjump shows the growing role that algos play in boosting the performance of traders. While all scores in 2021 were up from 2020. We will combine many popular strategies like arbitrage, index fund rebalancing, mean reversion, and market timing. Other strategies are scalping, transaction costreduction, and pairs trading in our Algo trading model. So, if any one of the strategies failed to work according to the given conditions/criteria the other strategy will take its place for further process.

3. Proposed Methodology



The proposed method consists of the following steps as given below:
1) Data Collection
2) Visualization of Historical Data
3) Pre-Processing of Data
4) Divide the data into Training and Testing
5) Developing a Hypothesis for the Strategy
6) Back Testing
7) Deploying

Firstly data from different sources are collected with the help of APIs and find the pattern in it and preprocessed it with data cleaning and refinement. The data is then processed with different ML algorithms which are later back-tested for validation. Different rules and strategy being tested for better accuracy selection.

Fig 1. Flow Chart of the work

The implementation of the work is carried out using a framework *Jupyter Notebooks* . An open-source package and environment management system named *Conda* is also being used.

3.1 Data Collection

Datas are collected from different APIs. Few of them are listed below.

i. **API Sandbox.** API sandbox is a feature for simulating and testing the Application Programming Interface (API). For developers, those activities are pivotal.

ii. **IEX Cloud API.** IEX Cloud is a platform that makes financial data and services accessible to everyone. The IEX Cloud API is based on REST, has resource-oriented URLs, returns JSON-encoded responses, and returns standard HTTP response codes.

iii. **YFinance.** Finance came as a support to those who became helpless after the closure of Yahoo Finance's historical data API, as many programs that relied on it stopped working.

3.2 Data Refining

For data refining , different python modules are used namely *Numpy* and *Pandas*. Numpy is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms basic linear algebra, basic statistical operations, random simulation and much more. Pandas, on the other hand, is a fast, powerful, flexible, and easy-to-use open-source data analysis and manipulation tool, built on top of the Python programming language.

3.3 Developing Hypothesis for Strategy

Different Hypothesis for Strategy as given below are used in the work.

I. **S&P 500 Equal Weight Index (EWI).** This is an equal-weight version of the popular S&P 500 Index. Although both indexes are composed of the same stocks, thedifferent weighting schemes result in two indexes with different properties and different benefits for investors.

II. **Quantitative Momentum Investing Strategy.** It is an investment strategy that selects for investment the stocks whose price appreciated the most during a period.

III. **Quantitative Value Investing Strategy.** It is an investing strategy that selects the highest- quality cheapest stocks using state-of-the-art computer algorithms for investment. Implementation of Quantitative Value has generated returns with relatively low volatility and low asset turnover.

IV. **Implementation of various Decision Trees depending upon the Nature of Charts.**

a. **LSTM.** LSTMs are very powerful in sequence prediction problems because they can store past information. This is important in this case because the previous price of a stock is crucial in predicting its future price.

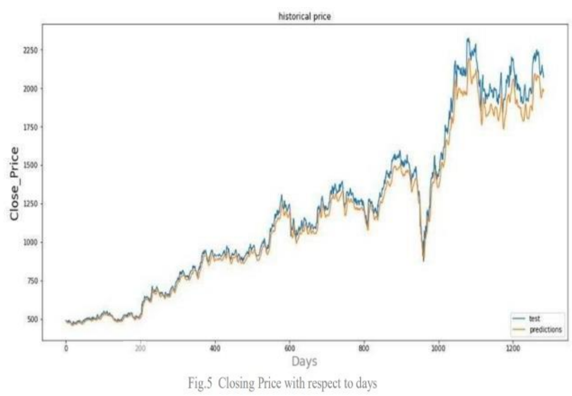


Fig 2. Closing price with respect to days

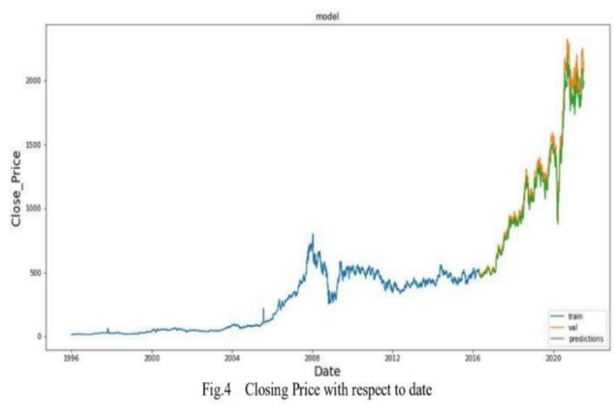


Fig 3. Closing price with respect to the date

b. Exponential Moving Average (EMA). An exponential moving average (EMA) is a type of moving average (MA) that places a greater weight and significance on the most recent data points.

1) Exponential Moving Average (EMA: is widely used in trading practices. It smooths the effects of price changes by giving more weight to the latest data. Our model uses a smoothing factor N=9 to calculate EMA as given below:

$$EMA = Value_{Today}\frac{2}{N+1} + EMA_{yesterday}\left(1-\frac{2}{N+1}\right) \qquad (1)$$

c. Moving Average Convergence-Divergence (MACD). The moving average convergence/divergence (MACD, or MAC-D) line is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA. The signal line is a nine-period EMA of the MACD line. MACD is best used with daily periods, where the traditional setting of 26/12/9 days is the norm.

i. MACD=12-Period EMA − 26-Period EMA

d. The Relative Strength Index. The relative strength index (RSI) is a momentum indicator used in technical analysis. RSI measures the speed and magnitude of a security's recent price changes to evaluate overvalued or undervalued conditions in the price of that security.

RSI = 100 – (100 / (1+ (Average Gain / Average Loss)))

e. Average Directional Index. Average Directional Index (ADX): works together with a Minus Directional Indicator (-DI) and a Plus Directional Indicator (+DI). The directional movement indicators are used to detect stock trends, typically used to determine entry and exit points. The ADX is derived from the smoothed averages between +DI and –DI.

f. Triple Exponential Average (TRIX). The triple exponential average (TRIX) is a momentum indicator used by technical traders that shows the percentage change in a moving average that has been smoothed exponentially three times.

g. Bollinger Bands. A Bollinger Band is a technical analysis tool defined by a set of trendlines plotted two standard deviations (positively and negatively) away from a simple moving average (SMA) of a security's price, but which can be adjusted to user preferences**.**

3.4 Back Testing the Strategy

For back testing the strategy, a python module named *IbPy* is used. *IbPy* is an unaffiliated third-party python wrapper for Interactive trader Workstation API. IbPy implements functionality that the Python programmer can use to connect to InteractiveBrokers, request stock ticker data, submit orders for stocks and futures, and more.

3.5 Data Visualization

For Data visualization, a popular platform named *TradingView* is used. *TradingView* is a visualization tool with a vibrant open-source community. It’s entirely web-based, and allows users to visualize data, whether the data is the result of paper trading or algorithmic back-testing.

3.6 Implementing the strategy in production

The strategies are implemented on one of India’s most popular trading platform named *Upstox.* Buying and selling of stocks are done through this online platform. Upstox provides real-time reports on gains and losses. One can take advantage of industry insights, benchmark comparison reports, and suggestions to improve online stock trading practices.

4. Results and Discussion

The following tables and figures shows the various results based on various strategy implemented in our work.

	Ticker	Stock Price	Market Capitalization	Number of Shares to Buy
0	A	150.46	43819635848	6
1	AAL	15.06	9571684649	66
2	AAP	186.17	11052414766	5
3	AAPL	154.76	2378149616999	6
4	ABBV	156.16	279462331273	6
...	...	...	...	...
95	CINF	112.91	17293599884	8
96	CL	77.24	63816188001	12
97	CLX	145.96	18366065798	6
98	CMA	74.27	9670026269	13
99	CMCSA	34.19	152336814191	29

Table 1. Output corresponding to S&P 500 Equal Weight Index

	Open	High	Low	Close	Adj Close	Volume
Date						
1996-01-01	15.859429	15.944529	15.754989	15.917452	10.509538	48051995
1996-01-02	15.878770	15.956133	15.677626	15.793671	10.427812	77875009
1996-01-03	16.052837	16.783918	15.878770	15.913584	10.506985	96602936
1996-01-04	15.762726	15.813012	15.553845	15.766594	10.409933	100099436
1996-01-05	15.704703	15.704703	15.522900	15.658285	10.338421	76935930

Table 2. Output corresponding to LSTM

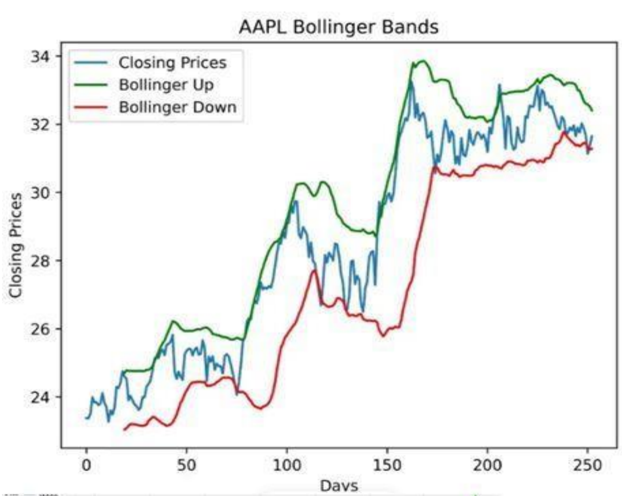


Fig 6. Bollinger Band Chart

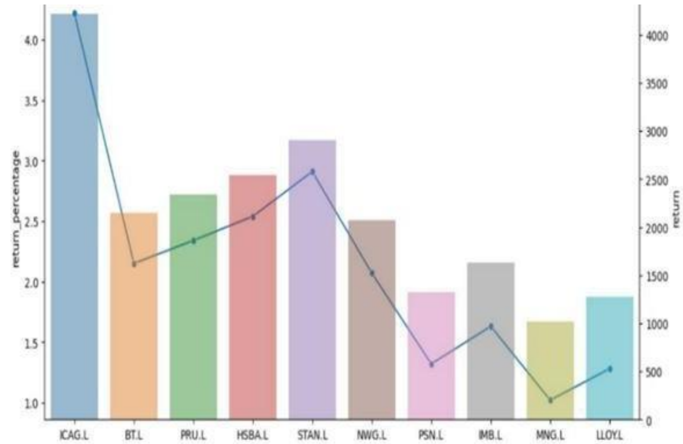


Fig 4.Graph of return % using Quantitative Momentum Investing strategy

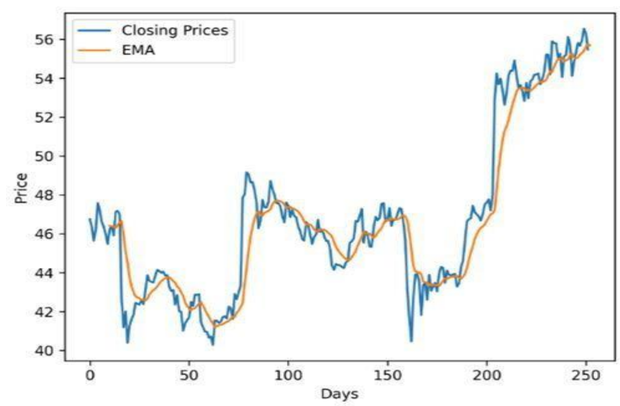


Fig 5.Exponential Moving Average chart

Fig 7. Moving Average Convergence-Divergence Chart

Fig 8. Relative Strength Index Chart

Fig 9. Average Directional Index Chart

In the study, **S&P 500 equal-weight index** tells us how many shares of each S&P 500 constituent one should purchase to get an equal-weight version of the index fund. Whereas in **Quantitative momentum Investment** goal is to work with volatility by finding buying opportunities in short-term uptrends and then selling when the securities start to lose momentum. In **Qualitative momentum Investment** returns are generated with relatively low volatility and low asset turnover.

With help of **LSTM,** we have predicted future prices of shares using past data. With help of a **Simple Moving Average** when we apply **Bollinger Band,** we observed an accuracy of 95% when the observed price action fell within the predicted range. An **Exponentially Weighted Moving Average** reacts more significantly to recent price changes than a **Simple Moving Average**, and because of this EMA follows prices more closely than a corresponding SMA. The **Relative Strength Index (RSI)** aims to signal whether a market is overbought or oversold in relation to recent price levels. Whereas **Moving Average Convergence-Divergence (MACD)** focuses on the level and direction of the MACD/signal lines compared with preceding price movements in the security at hand.

6. Conclusion

In this work, we have combined many popular strategies those are arbitrage, index fund rebalancing, mean reversion, and market timing. Other strategies are scalping, transaction cost reduction, and pairs trading in our Algo trading model. Because previous works have been done using only one strategy which may result in more chances of losing but in our model, we have merged several strategies to give our trading more chances of success so if one of the strategies is not working for a given condition another strategy would be automatically implemented. Unlike previous projects, we have taken the dynamic nature of the market into consideration, a few factors like elections, new government policies, and budget allocations that impact the stock exchange directly or indirectly. After completing the project layer by layer, we got a favorable outcome. Although we have done greater progress in this vast field still as the nature of the market keeps on changing and so do the strategies, there will be new variables and new aspects to look upon to keep growing in this field of Algorithmic Trading.

References

1. A SURVEY ON COMPUTER AUTOMATED TRADING IN INDIAN STOCK MARKETS(ACADEMIA)BY K Sudhakar, S Naganjaneyulu, Y Rama Mohan
2. Automated Trading System Statistical and Machine Learning methods and hardware implementation: a survey by Boming Huang, Yuxian Huan, Li Da Xu
3. Algo-Trading using statistical Learning And Optimizing Sharpe Ratio And DrawdownBy Neeraj Kasheety, Hanumanula Sravya, C Amarnath Reddy,Jaypal Medida
4. Building Winning Algorithmic Trading Systems_ A Trader's Journey from Data Mining to Monte Carlo Simulation to

Live Trading, Kevin J. Davey, WILEY

5. Achelis, Steven B. 2001. Technical Analysis from A to Z, 1st ed.; New York: McGraw Hill

6. Allen, Franklin, and Risto Karjalainen. 1999. Using genetic algorithms to find technical trading rules. Journal of Financial Economics

7. Ellis, Craig A., and Simon A. Parbery. 2005. Is smarter better? A comparison of adaptive, and simple moving average trading strategies.

8. Abu-Mostafa YS, Atiya AF (1996) Introduction to financial forecasting

9. Adegboye A, Kampouridis M (2021) Machine learning classification and regression models for predicting directional changes trend reversal in FX markets.

10. Aloud M (2020) An intelligent stock trading decision support system using the genetic algorithm. Int J Decis

11. Chang JF, Huang YM (2014) PSO-based time series models applied in exchange rate forecasting for business performance management. Electron Commerce Res

12. Ye A, Chinthalapati VR, Serguieva A, et al (2017) Developing sustainable trading strategies using directional changes with high-frequency data. In: 2017 IEEE international conference on big data (Big Data), 2017. IEEE, pp 4265–4271

13. Spero JEE, Hart JA (2009) The politics of international economic relations. Cengage Learning, Boston Thornton C, Hutter F, Hoos HH et al (2013) Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining, 2013. ACM, pp 847–855

14. Jones, C. M., Kaul, G., & Lipson, M. L. (1994). Transactions, volume, and volatility. Review of Financial Studies, 7(4), 631–651.

15. Aggarwal, N., & Thomas, S. (2014). The causal impact of algorithmic trading on market quality. Working paper, IGIDR Mumbai – WP-2014–23

16. Dubey, R. K., Chauhan, Y. K., & Syamala, S. R. (2017). Evidence of algorithmic trading from Indian equity market: interpreting the transaction velocity element of financialization. Research in International Business and Finance, 42, 31–38.

17. Dubey, R. (2016). Algorithmic trading: Is the machine taking over our equity markets? (pp. 28–34). IUP Publications

18. Sharpe, W. F. (1998). The Sharpe ratio. Streetwise–the Best of the Journal of Portfolio Management, 169-185

19. Suresh, A. A Study on Fundamental and Technical Analysis. Int. J. Mark. Finance. Serv. Manag. Res. 2013, 2, 44–59.

20. Adam, K., Marcet, A., & Nicolini, J. P. (2016). Stock market volatility and learning. The Journal of Finance, 71(1), 33-82.

21. Rundo, F. (2019). Deep LSTM with reinforcement learning layer for financial trend prediction in FX high-frequency trading systems. Applied Sciences, 9(20), 4460.

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

23. Sadique, S.; Silvapulle, P. Long-term memory in stock market returns international evidence. Int. J. Financ. Econ. 2001, 6, 59–67, doi:10.1002/IJFE.143.