Survey on Algorithmic Trading Using Data Science

Akarsh N
UG Student, Dept. Of CSE,
Sir M. Visvesvaraya Institute of Technology, Bengaluru, India

Ms. Dhivya V
Associate Professor,
Dept. of CSE,
Sir M. Visvesvaraya Institute of Technology,
Visvesvaraya Technological Institute, Bengaluru, India

Arjun A
UG Student, Dept. Of CSE,
Sir M. Visvesvaraya Institute of Technology, Bengaluru, India
8792850727a@gmail.com

Akhil S
UG Student, Dept. Of CSE,
Sir M. Visvesvaraya Institute of Technology, Bengaluru, India
akhil173@gmail.com

M R Harikrishna
UG Student, Dept. Of CSE,
Sir M. Visvesvaraya Institute of Technology, Bengaluru, India
harimr486@gmail.com

Abstract - One aspect of stock trading is the buying and selling of shares in a certain firm. You can acquire ownership of a piece of a firm by purchasing specific stocks and shares. A person who trades stocks on behalf of a financial institution is known as a stock trader. Stock traders may be categorized into three groups: knowledgeable, uninformed, and intuitive. Algorithmic trading is the process of executing orders using automated and pre-programmed trading instructions that take into account variables such as price, time, and volume. An algorithm is a set of instructions for solving a problem. Computer algorithms progressively deliver the complete order to the market in smaller portions. Algorithmic trading makes use of complex formulas and Data Science models to decide whether to purchase or sell financial securities on an exchange using computer algorithms and human oversight. High-frequency trading technology, which enables a company to execute tens of thousands of deals per second, is frequently used by algorithmic traders.

Index Terms **- Data Science, Mathematical models, High-Frequency trading technology, Algorithmic trading**

I. INTRODUCTION

The group of people who buy and sell stocks, also known as shares, which stand for ownership holdings in businesses, is referred to as a stock market, sometimes known as an equity market or share market. These securities, such as shares of private companies that are made available to investors through equity crowdfunding platforms, may be listed on a public stock market or solely traded privately.

When computerized trading systems were introduced to the American financial markets in the 1970s, the use of algorithms in trading increased. The Designated Order Turnaround (DOT) system, which directs trading orders to experts on the trading floor, was initially introduced to the New York Stock Exchange in 1976. The ability of exchanges to conduct electronic trading improved over the ensuing decades, and by 2009, upwards of 60% of all transactions in the U.S. were handled by computers.

Large brokerage firms and institutional investors are the main users of algorithmic trading to reduce trading expenses. Large order sizes, which might account for as much as 10% of all trading activity, are said to benefit particularly from algorithmic trading, according to study. Market makers commonly utilise algorithmic trading to provide liquidity.

Faster execution times and lower costs are two benefits of algorithmic trading, but it also has the potential to accentuate unfavourable market trends by resulting in flash crashes and a sudden lack of liquidity.

II. RELATED WORK

Ruppa K. Thulasiram.,2020[1] stated that one of the most noticeable trends in the finance sector has been the explosion of algorithmic trading. For algo trading, a unique data-driven fuzzy Bollinger bands method is suggested. Algorithms are used in algorithmic trading to generate judgements based on sets of instructions (or rules). The majority of transactions are started based on favorable trends, which are simple and straightforward for algorithms to apply. The majority of well-known algorithmic trading systems employ simple moving averages (SMA) and other closely related technical indicators to identify trends. Moving average crossovers are used in a SMA crossover strategy to enter long (buying) or short (selling) positions. X = Corr(X) (measured by the student's t distribution with sign correlation X and finite variance) is the definition of the sign correlation of a random variable X with mean.

Jithin Mathews., 2018[2] suggested There are many various ways to manipulate taxes, each with its own goals and levels of complexity. This essay outlines a specific method of tax evasion called "circular trade." The commercial tax dataset of the Indian state of Telangana is used to test the suggested algorithm. A business dealer is required to pay a specific tax amount on the items acquired from dealerA to dealerB. Similar to this, dealerB will receive some taxes on the offered items when dealerC purchases these goods from dealerB. The flow of money in a value-added transaction is shown in Figure 1. DealerA will make up a lot of imaginary dealers in this article utilising the personal information of his/her reliable friends. The tax owed on these fraudulent transactions is practically nonexistent.

Chia-Han Chou, 2018[3] concludes that this project aims to produce a visual interface that enables traders to create programs using the straightforward Drag & Draw method. According to Greenwich Associates, Inc., 95% of major U.S. traders use algorithmic trading technology to carry out their trading plans. The term "algorithmic trading" refers to the type of trading technology that enables traders to turn quantitative formulas, technical indicators, and fundamental indicators into computer programs.

Aaron Wray., 2020[4] wrote for a tradeable asset, it uses input and output data from the Limit Order Book (LOB) or Ladder of the market. Instead of making explicit price predictions, it hones its skills on LOB snapshot/quote pairs. The first computer programme to mimic or outperform a human trader's style of trading is called "DeepTrader." The gadget is a "black box" that accepts all data input from the user's screen and audio/voice lines. The box begins to automatically send a stream of orders to the market after a given amount of time. The human trader's "sensory inputs" and "motor outputs" are irrelevant to us; all that matters is how well she interprets them. In supervised learning, where an algorithm is used.

Md. Erfanul Hoque, 2020[5] We explore a data-driven method for choosing parameters that maximize the Sharpe ratio (SR) and produce the best trading signals. For dynamic state space models, non-Gaussian maximum informative filtering techniques are proposed. In 2018, up to 50% of all stock transaction activity in Canada was generated by algorithms. The hedging ratio, or regression coefficient 1, is used in pairs trading to specify how much of one security to buy or sell for each unit of the other. Additionally, data were used to develop an estimate of the standard deviation. The vector parameter case and multiple trading are added to the combined estimating function technique. The use of smoothing, rolling forecasting, and filtering is discussed in part II along with other dynamic trading strategies.

In relation to a particular dataset that they took into consideration, Wei Wang, 2019[6] suggested a framework for machine learning-based trading algorithms that use fictitious bids in the power markets. To produce precise predictions for the variation in power price between day-ahead and real-time markets, a mixed density network model was created. Results from back casting using market data from ISO New England show that the recommended method outperforms the standard online learning strategy. DECs might be used by IPPs and LSE to protect themselves against hazards like as forced generator outages, increasing RT unstable LMPs and RT electric load. Proprietary trading firms utilize virtual bids to arbitrage the electricity market. With the use of DECs, energy may be exchanged across the RT and DA markets at the same price node. In the United States' five biggest electricity markets

You Liang, 2020[7] stated in this work, combined Using algorithmic trading, volatility and stock price predictions are put into practice. Generalized double exponential smoothing and a data-driven exponentially weighted average are used to construct interval stock price projections (DD-EWMA). For both the short and long timeframes in our analysis, we use SMA trend indicators. The SMA is the average of the Pt, t=1, D adjusted closing prices for the preceding D periods. Moving average crossovers are used in the SMA technique to modify a portfolio's weights. The conceptual strength of the suggested strategy and procedures is strong, and the distribution behind them is adequately explained. Index trading automation is advised as a trading method. The proposed trading approach is used to determine the profits of various stock indexes.

Zimo Zhu, 2020[8] proposes one of the most notable recent changes in the banking business has been the rise of algorithmic trading. KF has recently attracted more attention as a pair trading strategy. Results indicate that the data-driven innovation volatility forecast (DDIVF) trading approach is superior to the widely utilized KF-based model. Lasso estimations can be thought of as L 1 - penalized least squares estimations with penaltyJ 2 = 1. The answers to maximum formative estimating equations yield filters. The ideal EF based on the observed process and the prior EF of the prior process are combined to form an unbiased Bayesian regularized EF. The unbiased filtering equation for state space models might be solved to get the best linear predictor.

Sevi Baltaoglu, 2018[9] suggested that a virtual trader's job is to arbitrage the price differences between power day-ahead and real-time. An online algo-learning algorithm is created to optimize the cumulative reward over a finite number of trading sessions. The recommended strategy surpassed standard benchmarks and the S&P 500 index within the same time frame. The difference between the Daily Average (DA) and Real Time Average (RT) power rates might be caused by a number of factors, including different levels of demand and supply, outages, and weather. The American power markets were first exposed to virtual trading in the beginning of the 2000s.

Hongyong Sun 2019[10] stated that in this study, neural networks incorporating reinforcement learning are used in high-frequency currency trading. Learn-\sing. The idea is that a neural network-based agent would be able to recognize the temporal pattern in data and execute trades on its own based on historical data and the current situation of the market. To increase overall profit, we suggest two strategies: action shaping and advantage function shaping. Since it is expected that the agent trades fixed position sizes in a single security, the action shaping is utilized to prevent the agent from acting unethically. To raise the likelihood of activities that produce greater profit, benefit function shaping is advocated. On the currency market, the suggested system has undergone back testing. The outcomes show that our approach works well in most circumstances.

III. METHODOLOGY

The System Design Mainly Consist of

A. Fetching Data
B. Handling Missing Data
C. Scaling and Normalizing Data
D. Training Model
E. Prediction Using Live Data
F. Sending API Calls based on Predictions

A. Fetching Data

Algorithmic trading involves fetching and processing market data to make informed trading decisions. This includes collecting data from various sources, such as financial news, stock market feeds, and economic indicators. The data is then analyzed using statistical and mathematical techniques to identify patterns and trends. Specialized software programs and programming languages are used for this data analysis. Once the relevant data has been collected and analyzed, the algorithmic trading system will use this information to make decisions on when to buy or sell securities based on predefined trading rules and strategies. The accuracy and quality of the data used are crucial to the success of the trading system, and algorithmic trading firms invest heavily in data collection and processing infrastructure to ensure they have access to reliable and up-to-date information.

B. Handling Missing Data

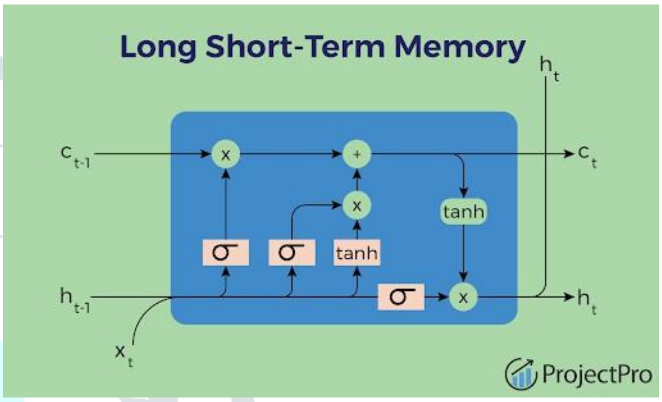
In stock price prediction, missing data needs to be handled appropriately. There are several methods to handle missing data, including imputation, forward and backward filling, and interpolation. Imputation involves filling in missing values using statistical techniques such as using the mean or median value. Forward filling uses the last available value, backward filling uses the next available value, and interpolation involves estimating missing values using a mathematical function. Choosing the best method to handle missing data is crucial as it can have a significant impact on the accuracy of the stock price prediction model.

C. Scaling and Normalizing Data

Scaling and normalizing are important data preparation techniques in stock price prediction. Scaling involves transforming different feature values in the dataset to a common scale to avoid features with disproportionate impact on the model's predictions. Methods used for scaling include Min-Max scaling, where values are scaled to a range between 0 and 1, and Standardization, where values are scaled to have a mean of 0 and a standard deviation of 1. Normalization, on the other hand, involves transforming data to a normal distribution since many statistical models used in stock price prediction assume the data is normally distributed. Common methods for normalization include the Box-Cox transformation and the Z-score normalization. Selecting the best method depends on the specific data and modeling approach being used, and it's important to consider their impact on the accuracy of the stock price prediction model.

D. Training Model

Training a model in stock price prediction involves using historical data to train a machine learning algorithm that can make predictions about future stock prices. The process involves selecting an algorithm, such as linear regression or neural networks, and providing it with historical data that is split into training and testing sets. The algorithm learns from the data and creates a model that can be used to make predictions on new, unseen data. The model's accuracy is evaluated using metrics such as mean squared error and R-squared, and once it has been trained and evaluated, it can be used to make predictions about future stock prices. However, stock price prediction is complex and unpredictable, and other factors can influence stock prices beyond historical data, so it's important to use predictions as a tool rather than a sole basis for trading decisions.

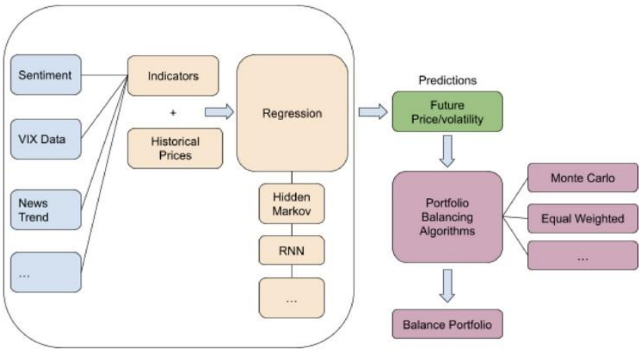


E. Prediction Using Live Data

Prediction using live data in stock price prediction involves using real-time data from various sources such as news articles, social media, and financial statements to make predictions about future stock prices. The data is continuously fed into a machine learning model, which adjusts its predictions based on the latest information. To use live data, a robust data pipeline is required that can collect, clean, and process the data in a timely manner. However, live data-based predictions face the same limitations as historical data-based prediction due to the complex and unpredictable nature of the stock market. Therefore, live data predictions should be used as a tool to inform investment decisions rather than relying solely on them to make trading decisions.

F. Sending API Calls based on Predictions

Sending API calls based on predictions in stock price prediction involves using the output of a prediction model to automatically trigger trades or other actions through an API. The prediction model is integrated with an API that can execute trades or actions based on its predictions, and when a prediction is generated, an API call is sent to the trading platform for automatic execution. However, this process requires careful consideration and testing, as automated trades based on predictions can be subject to the same risks as manual trading. Regular evaluation of the model's accuracy and risk management strategies are necessary to mitigate potential losses.



IV. RESULT

The Proposed system, Algorithmic Trading using Data Science and the machine learning model like LSTM playing a crucial role in predicting stock price prediction gives a root mean square error of 22.56 in predicting the stock price and was able to successfully call the API to connect to the broker and place the trades accordingly.

V. CONCLUSION

Trading assets on the stock market has long been a popular activity. Trading is now easier and involves a lot less human interaction as a result of the effectiveness and promise of algorithmic trading mixed with various machine learning algorithms like LSTM. The most recent development in the area is the idea of HFT (High Frequency Trading), which combines algorithmic trading with low latency hardware to significantly increase trading speed.

VI. REFERENCES

[1] Thavaneswaran, Y. Liang, Z. Zhu and R. K. Thulasiram, "Novel Data-Driven Fuzzy Algorithmic Volatility Forecasting Models with Applications to Algorithmic Trading," *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1-8, doi: 10.1109/FUZZ48607.2020.9177735.

[2] J. Mathews, P. Mehta, C. S. Babu and S. V. K. V. Rao, "An algorithmic approach to handle circular trading in commercial taxation system," *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, 2018, pp. 67-75, doi: 10.1109/ICBDA.2018.8367653.

[3] C. -H. Chou and A. Y. Chang, "Visual Development Platform for White-Box Algorithmic Trading," *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2010, pp. 712-715, doi: 10.1109/IIHMSP.2010.180.

[4] A. Wray, M. Meades and D. Cliff, "Automated Creation of a High-Performing Algorithmic Trader via Deep Learning on Level-2 Limit Order Book Data," *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 1067-1074, doi: 10.1109/SSCI47803.2020.9308266.

[5] Y. Liang, A. Thavaneswaran, N. Yu, M. E. Hoque and R. K. Thulasiram, "Dynamic Data Science Applications in Optimal Profit Algorithmic Trading," *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2020, pp. 1314-1319, doi: 10.1109/COMPSAC48688.2020.00-74.

[6] W. Wang and N. Yu, "A Machine Learning Framework for Algorithmic Trading with Virtual Bids in Electricity Markets," *2019 IEEE Power & Energy Society General Meeting (PESGM)*, 2019, pp. 1-5, doi: 10.1109/PESGM40551.2019.8973750..

[7] Y. Liang, A. Thavaneswaran and M. E. Hoque, "A Novel Algorithmic Trading Strategy Using Data-Driven Innovation Volatility," 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020, pp. 1107-1114, doi: 10.1109/SSCI47803.2020.9308360.

[8] Y. Liang, A. Thavaneswaran, A. Paseka, Z. Zhu and R. K. Thulasiram, "A Novel Dynamic Data-Driven Algorithmic Trading Strategy Using Joint Forecasts of Volatility and Stock Price," 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), 2020, pp. 225-234, doi: 10.1109/COMPSAC48688.2020.00038

[9] S. Baltaoglu, L. Tong and Q. Zhao, "Algorithmic Bidding for Virtual Trading in Electricity Markets," in IEEE Transactions on Power Systems, vol. 34, no. 1, pp. 535-543, Jan. 2019, doi: 10.1109/TPWRS.2018.2862246

[10] H. Sun, N. Sang, J. Wu and C. Wang, "Algorithmic Currency Trading Based on Reinforcement Learning Combining Action Shaping and Advantage Function Shaping," 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019, pp. 1494-1498, doi: 10.1109/ICTAI.2019.00212.

[11] Penumatcha Bharath varma, Neeraj Kasheety, HanumanulaSravya, Chinthapalli Amarnath Reddy, JaypalMedida, “Algo-Trading using Statistical Learning and Optimizing Sharpe Ratio and Drawdown”, 2021 Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP).

[12] Cohen, G. Algorithmic Trading and Financial Forecasting Using Advanced Artificial Intelligence Methodologies. Mathematics 2022, 10, 3302. https://doi.org/10.3390/math10183302

[13] Shubham Horambe, Ketan Khanolkar, Dhairya Dixit, Huzaifa Shaikh1, Prof. Manasi U. Kulkarni,“Algorithmic Approach to Options Trading”.

[14] Goutham Ramkumar, “A Study On The Significance of Algorithmic Trading in Indian Stock Market”, 2018 IJRAR October 2018, Volume 5, Issue 4.

[15] Thomas Spooner, Freddie Gibbs, Hunter S. Thompson, “Algorithmic Trading and Reinforcement Learning”.

[16] Robert L Kissel, “Algorithmic Trading Strategies”.

[17] Michael S. Barr, Elaine Wah, “Computational Models Of Algorithmic Trading in Financial Market”

[18] D. Othalasseril and S. Shaikh, "TradeZilla Using Algorithmic Trading," 2021 IEEE India Council International Subsections Conference (INDISCON), 2021, pp. 1-6, doi: 10.1109/INDISCON53343.2021.9582206.

[19] T. Salkar, A. Shinde, N. Tamhankar and N. Bhagat, "Algorithmic Trading using Technical Indicators," 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021, pp. 1-6, doi: 10.1109/ICCICT50803.2021.9510135.

[20] T. Salkar, A. Shinde, N. Tamhankar and N. Bhagat, "Algorithmic Trading using Technical Indicators," 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021, pp. 1-6, doi: 10.1109/ICCICT50803.2021.9510135.