Artificial Stock Analyst: Using Machine Learning to Create a Holistic Approach to Stock Analysis

The Use and Ethics of Algorithmic Trading and Artificial Intelligence in the Stock Market

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Nathaniel Gleberman

October 27, 2023

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Rider Foley, Department of Engineering and Society

Briana Morrison, Department of Computer Science

INTRODUCTION

From February to April 2020, the stock market plummeted with the largest daily drop in the global markets reaching nearly 13%. In mere weeks, billions of dollars in savings and income seemingly disappeared as markets reached new lows. Events like pandemics that turn the market upside down are referred to as “Black Swan” events given their difficulty to predict. The prospect of savings rapidly disappearing is frightening and might discourage investing in stock markets. Yet, by August 16^{th}, 2021, the S&P 500 had doubled from that pandemic bottom, an achievement that typically takes nearly one thousand days. Buying at the bottom could have earned investors years’ worth of income (Li & Rattner, 2021).

Understanding when to buy and sell stock is a learnable and important skill that can help people retire faster. The market is complex as stock price movement can rarely be attributed to one factor (Harper, 2022). This complication cannot be mistaken for pure chance. Algorithms can penetrate this complication and create educated estimates for price movement, although perfect accuracy is never guaranteed. Trading with an algorithm could help increase the probability of positive returns.

Learning effective trading strategies can be a time consuming and arduous challenge for many. According to a 2019 study, wealthier Americans tend to directly own stock yet only 52.6% of American families own stock (USAFacts, 2021). This serves as a strong indicator that to reach the wealthier quartiles, owning stock is a necessity yet most families are absent from the stock market. By putting free and powerful tools into the hands of people looking to diversify their savings, trading could become a manageable and encouraging pursuit for anyone regardless of socioeconomic status or coding background.

To clarify, the U.S. Securities and Exchange Commission (SEC) defines the term *algorithmic trading* as computerized strategies for stock trading and not broker algorithms used to execute trades (SEC, 2020). The first instance of computerized trading was in 1976 when the New York Stock Exchange unveiled the Designated Order Turnaround (DOT) system for routing orders from traders to specialists. As technology advanced, DOT adapted to accept electronic trading (Chen, 2022). By 2022, algorithmic trading accounted for 60-73% of equity trading in the U.S and is currently valued at 14.42 billion dollars and expected to grow to 23.74 billion in the next five years (Mordor, 2022). As artificial intelligence (AI) has grown in popularity and possibility, AI trading has become a subset of algorithmic trading that uses machine learning (ML) development services to proactively react to the market (Lezginov, 2022). With such a large chunk of the market controlled by algorithms, human traders who cannot respond as quickly or have access to less data are at a disadvantage leaving them at the whims of the algorithms in terms of rapid crashes or final judgements on a company.

By analyzing the history and variations of trading algorithms through the lens of the social construction of technology, the development of a free, powerful, and accountable supplemental stock trading AI is revealed to be a necessity in improving accessibility to the stock market. By relying on traditional technical indicators and limiting the AI tool to the S&P 500, the program looks to replicate overall market health while escaping the loop of development. This implementation can be done in an ethical manner that promotes overall economic health, sustainable practices, prevention of discrimination, and better mental health.

TECHNICAL TOPIC

To increase accessibility for powerful trading and investing tools, I propose building a comprehensive model that can be quickly developed, freely launched, and minimally maintained.

To accomplish this, a web application will be developed and hosted on Firebase, a service offered through Google that allows hosting web servers for free (Google, 2023). The front end features a simple user interface that grants access to stock price predictions for the S&P 500. The backend will consist of two major components: the database engine and the stock predictor engine. The stock predictor engine consists of technical indicators, a separate stock prediction ML model, and the overarching ML model. The database engine maintains local databases inside the GitHub repository and includes an updating script that must be run once a week by a local machine.

The user interface will feature a simple dropdown menu of companies from the S&P 500. For each company, the web application will output a weekly stock percent change prediction. This weekly prediction is derived from models built with TensorFlow, an open-source ML library developed in Python from engineers at Google. TensorFlow has a strong focus on deep neural networks while maintaining a robust array of other models including Random Forests (Abidi et al., 2016). TensorFlow will create an overarching Random Forest model which will take as input three categories of data. The first category is technical indicators which are quantifiable signals produced by stock price, volume, or open interest indicating potential movement for an equity (Chen, 2021). The next category is economic data, specifically data from the Federal Reserve Economic Data representing a weekly economic index consisting of "ten daily and weekly indicators of real economic activity, scaled to align with the four-quarter GDP growth rate" (*Weekly Economic Index,* n.d.). The last category is the prediction outputted by an artificial neural network (ANN) for the weekly price change prediction.

Neural networks like ANNs are models meant to simulate the neurons of human beings. As the number of hidden layers data goes through increases beyond three, the neural network becomes a subset of ML known as deep learning which further approaches AI. This anthropomorphic design creates a close link between the human developers and the models themselves. Like the human brain, deep learning is not a fully understood topic with many experts unsure as to how their models can be so accurate (Borch & Min, 2022). This is a concern as widespread usage of these opaque algorithms could perpetuate gender, racial, and other biases as they become difficult or nearly impossible to identify and mitigate. In looking to replicate the human mind, ML and AI have the scary potential to harm others and discriminate or endanger different social groups.

ANALYZING CONCERNS

Machine learning, algorithms, and artificial intelligence are powerful terms that can be easily tossed around. Algorithms are interwoven into our everyday lives from the fingerprint sensors on phones to the inner workings of morning alarm clocks. The history of computers reveals how tightly algorithms and social structures are coupled, often hiding the immense complexity encapsulated by the single word. There exists a disparity between the understanding of algorithms from social groups with influence over digital tools and marginalized individuals who are ultimately affected by new digital technologies without a complete understanding of the underlying technology (Burrell & Forcade, 2021). Disparities between the "coding elite" and the "cybertariat" or marginalized social groups become increasingly apparent with algorithms in the financial realm.

The social construction of technology (SCOT) helps in understanding the relationship between these algorithms and the social groups surrounding them. SCOT states that technology is not deterministic. Rather, social groups advance and shape the technology as people ultimately give meaning and determine the adoption or rejection of that artifact (Pinch & Bijker, 2012). As evidence of this, the bicycle started out as a giant front wheel and slowly evolved into the modern bicycles seen today. Different social groups led to different adaptions, modifications, or rejections to solve their respective and unique problem. For example, safer bicycle models were produced for parties not attuned to the high seats and large front wheels that could create so much speed. Using a multidirectional model, the relevant social groups and their impacts can reveal the complicated narrative between everyday artifacts.

SCOT can be applied to algorithms in the stock market to better understand the connections between the code and humans. Different groups see different approaches to algorithmic trading, so thousands of variations of these financial systems exist. With these developments come accessibility concerns expressed by the terms "coding elite" and "cybertariat." Better models can be created by more technologically and financially inclined social groups which in turn creates disparity between those with a programming background and those without. In a stock market dominated by algorithms, social groups unable to access or create the proper tools for viable competition with powerful models can become the bearer of heavy losses, barely incentivizing gains, or the victim of steep barriers to entry.

Moreover, these models and intelligent trading agents are designed to chase profit—a potentially dangerous final goal. Not all profitable companies are good companies by a standpoint that promotes sustainability and a healthier economy. AI in the stock market is a rapidly growing field with special interest from retail investors, asset management companies, and quant funds. With this massive increase in AI's presence comes huge quantities of data and servers that need to be built and maintained (Vellaiparambill & Natchimuthu, 2022). The development of AI in the stock market needs to be conducted in a sustainable manner while promoting the creation of models that consider more than just profit as a guide.

Not only do these algorithms create strain on the environment, but the competition to create better models places further stress on the human actors behind the development. As models continue to advance, older models become less effective. Companies spend thousands of dollars to develop new strategies which are quickly replicated by competitors, eliminating any brief advantage the company held. As technical advancements continue, a paradox of substantial resources is created as companies looking to compete create an endless loop of resource exhaustion where competition made outperforming the market harder to accomplish. In response to this, the ultimate winner of algorithmic trading could ironically be the company with the best human to human service as positive service is a proven retainer of customers (Pozen & Kothari, 2023).

As companies continue to push this technology, the government has done little to intervene. The SEC has created more ways to provide oversight for algorithmic trading, but as a constantly evolving field, they simply lack the resources to look over the shoulder of every retail investor and corporation trading with algorithms. Relying on statistics to notice discrepancies, the SEC does little to limit algorithmic trading (SEC, 2020). With a government still adjusting to new technologies, a public discussion and agreement of ethics surround AI trading becomes a necessity. AI needs to focus on the relevant systemic effects of usage. Current ethics of AI turns to transparency, fairness, non-maleficence, privacy, and accountability, but fails to capture the big picture. Widespread integration of AI technology in the realm of finance can potentially magnify pre-existing risks. This includes the potential for causing systemic harm and endangering the investments of individuals who do not occupy the upper echelons of the financial hierarchy. As evidence, liquidity crunches, price collapses, and severe market disruptions have all occurred as the result of AI overextending its reach into the market

(Svetlova, 2022). These challenges that exist between technology and humans highlight the importance of approaching AI carefully and thoroughly to better lift all social groups involved in the stock market while promoting overall economic health.

RESEARCH QUESTION AND METHODS

High-frequency trading (HFT) is an algorithmic trading strategy that utilizes the speed of a computer to ride the increase in price of large institutional trades that are broken into multiple timeframes (CFA Institute, 2019). HFTs are often given a bad reputation for this seemingly exploitive strategy, yet data indicates HFTs promote a healthier marketplace by increasing liquidity and sacrificing monetary gains on actual stock positions for monetary gains on associated transaction rebates (Pavlus, 2019).

HFTs offer a chance to spotlight the discrepancies in human sentiment about algorithms, prompting a need for a deeper comprehension of algorithms' role in the stock market and the ethical considerations surrounding them. Bibliometrics provides a chronological approach to analyzing this development and growth of algorithmic trading. Researching archived and old strategies will illustrate the impact of social groups developing different approaches to finding profitable trades through code. Starting from the DOT era and tracing the evolution up to the coining of the term program trading, a comprehensive investigation will be conducted to unveil the complete narrative encompassing various terminologies leading to the classification of algorithmic trading. In addition, I will conduct interviews with engineers at Quantitative Investment Management, a global investment management firm based in Charlottesville, Virginia focused on predicting short- and medium-term price movements (*Quantitative Investment Management*, n.d.). These interviews will shed light on more current trading strategies that bibliometrics might fall short on sharing and give a developer perspective on the resources and strain that arise in creating these models.

CONCLUSION

Creating free and intelligent trading algorithms is paramount in improving stock market trading accessibility. This technical project outlines a proposal for the quick development of a publicly available tool that could be easily maintained. In creating this tool, the ethics of algorithmic trading need to be analyzed by first observing the history of building code that can intelligently trade stocks. Here, the STS deliverable looks to unveil a better understanding of that history, discuss an appropriate ethical lens to analyze artificially intelligent agents in the stock market, and shed light on the current market conditions of algorithmic trading.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., & Wicke, M. (2016). *TensorFlow: A System for Large-Scale Machine Learning This paper is included in the Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). TensorFlow: A system for large-scale machine learning*. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Borch, C., & Hee Min, B. (2022). Toward a sociology of machine learning explainability: Human–machine interaction in deep neural network-based automated trading. *Big Data & Society*, *9*(2), 205395172211113. https://doi.org/10.1177/20539517221111361

Burrell, J., & Forcade, M. (2021). Society of Algorithms. *Annual Reviews*, *47*, 213–237. https://doi.org/10.1146/annurev-soc-090820-020800

Chen, J. (2021, September 29). *Technical Indicator Definition*. Investopedia. https://www.investopedia.com/terms/t/technicalindicator.asp

Chen, J. (2022). Algorithmic Trading. *Investopedia*. https://www.investopedia.com/terms/a/algorithmictrading.asp

Google. (2023). Google Firebase. https://firebase.google.com

Harper, D. R. (2022, July 22). *What Drives the Stock Market?* Investopedia. https://www.investopedia.com/articles/basics/04/100804.asp

High-Frequency Trading. (2019, October 29). High-Frequency Trading. *CFA Institute*. https://rpc.cfainstitute.org/en/policy/positions/high-frequency-trading

Lezginov, M. (2022, January 4). *AI Trading vs. Algorithmic Trading. What's the Difference?* Scopic. https://scopicsoftware.com/blog/the-future-of-ai-trading/

Li, Y., & Rattner, N. (2021, August 16). *S&P 500 doubles from its pandemic bottom, marking the fastest bull market rally since WWII*. CNBC. https://www.cnbc.com/2021/08/16/sp-500-doubles-from-its-pandemic-bottom-marking-the-fastest-bull-market-rally-since-wwii.html

Mordor Intelligence. (2022). Algorithmic Trading Industry Report. *Mordor Intelligence*. https://www.mordorintelligence.com/industry-reports/algorithmic-trading-market

Pavlus, J. (2019). Do high-frequency traders deserve their bad rap? *Northwestern University*. https://insight.kellogg.northwestern.edu/article/do-high-frequency-traders-deserve-their-bad-rap

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. Social Studies of Science, 14(3), 399-441. https://doi.org/10.1177/030631284014003004

Pozen, R. C., & Kothari, S. P. (2023). The Paradox of AI and Investing. *International Banker*. https://internationalbanker.com/technology/the-paradox-of-ai-and-investing/

Quantitative Investment Management | LinkedIn. (n.d.). Www.linkedin.com. Retrieved October 27, 2023, from https://www.linkedin.com/company/quantitative-investment-management

Securities and Exchange Commission (2020), Staff Report on Algorithmic Trading in U.S. Capital Markets 1–92. https://www.sec.gov/files/Algo_Trading_Report_2020.pdf

Svetlova, E. (2022). Ai Ethics and Systemic Risks in Finance. *AI and Ethics*, *2*(4), 713–725. https://doi.org/10.1007/s43681-021-00129-1

Vellaiparambill, A., & Natchimuthu, N. (2022). Ethical tenets of stock price prediction using Machine Learning Techniques: A sustainable approach. *ECS Transactions*, *107*(1), 137–149. https://doi.org/10.1149/10701.0137ecst

Weekly Economic Index (WEI). (n.d.). www.newyorkfed.org. Retrieved October 27, 2023, from https://www.newyorkfed.org/research/policy/weekly-economic-index#/overview

What percentage of Americans own stock? (2021, March 2). USAFacts. https://usafacts.org/articles/what-percentage-of-americans-own-stock/