**d Project Summary**

**Student:** Jack D. Kincannon

**Faculty Mentor:** Dr. Karl Schubert

**Institution:** University of Arkansas

**Classification:** Senior

**Grade Point Average:**

**Area of Study:** Data Science

**Project Title:** Automated Portfolio Selection and Execution Using Deep Learning

**Abstract:** Automated trading systems execute trade orders based on programmed logic and account for the majority of US equity trades. These systems utilize cutting-edge statistical and computational methods to make decisions. My research will focus on the design and implementation of an automated trading system. The design of the system will balance annualized return, downside risk, and operating costs. Present research has a narrow scope, focusing on which methods produce the best empirical returns.I will instead build a holistic system utilizing current best practices. This system will provide a framework upon which future research can be built. I will program in Python and operate the system as a containerized application in an Azure cloud environment. Both quantitative and qualitative data will be used, and predictions will be generated by combining quantitative finance with modern statistical methods. Only Exchange Traded Funds (ETFs) will be considered for inclusion in the portfolio; advanced algorithms will aid in portfolio construction and trade execution.

**Automated Portfolio Selection and Execution Using Deep Learning**

**Background and Significance**

Modern finance has been transformed by the democratization of computing resources, information, and investing itself. Cloud services provide low-cost computing resources (Rountree & Castrillo, 2014), live data Application Programming Interfaces (APIs) can be accessed for no cost, and trades can be executed for free. These resources can be combined to create an automated trading system (Huang et al., 2019). Automated trading systems account for nearly 3 quarters of all US equity trades (Mordor, 2022). They are powerful tools that actively manage a portfolio using modern statistical methods (Chen, 2022). Active management requires the model to constantly make decisions on how to select assets, optimize the mix of assets, and acquire the assets.

Fundamentally, automated trading systems must generate time-series predictions. This is a challenging problem due to the stock market’s tendency towards being a nonlinear, dynamic, noisy, and chaotic system (Deboeck, 1994; Wang et al., 2011). Predicting prices of individual assets within the stock market complicates the situation further. Prices are affected by a myriad of factors. These factors include political events, corporate policies and news, economic situations, interest rates, and investor sentiments (Wang et al., 2011). Fortunately, data science techniques simplify the problem of stock price prediction. Advanced data collection methods allow researchers to capture features used in predicting the future performance of an asset. Once the future performance of a group of assets has been predicted, portfolio optimization techniques can be used to select a mix of assets that balance risk and return (Markowitz, 1952). Now that a portfolio has been created, the assets included must be acquired. The acquisition of assets occurs through the execution of buy orders. The success of an acquisition is highly dependent upon the time of day and factors like the ones listed above. To adapt to variability, selection, optimization, and execution functions must run constantly during market operation.

The selection of academic literature surrounding this topic is vast. Currently, the frontier focuses on what methods produce the best and most consistent returns. This leads to the research having a narrow scope. I will fill a gap by integrating current best practices and building a holistic automated trading system backed by accessible tools. Most integrated systems, such as this, exist inside organizations with little to no transparency. I want to build a framework for researchers that can be adapted to new problems and improved upon over time.

**Research Objectives**

The objective of this research is to create an application that actively manages a portfolio – within an automated trading system – for as little operating cost as possible. The success of this research project will be determined by the portfolio’s performance – relative to a comparable index — after deducting operating costs.

**Methodology**

Data science and finance methods will be used alongside various software platforms and open-source tools to build an automated trading system. TD Ameritrade will be used as my quantitative data source and brokerage platform. TD Ameritrade holds an advantage over other zero-transaction-fee brokers, like Robinhood, because of their superior API (TD Ameritrade, 2022). The system will be programmed in Python. This language was the best choice, first due to my experience, but also due to its versatility. I will run the system as a Docker container deployed on the Azure cloud platform (Docker, 2022).

This automated trading system will only consider Exchange Traded Funds (ETFs). ETFs revolutionized investing by providing the desirable features of mutual funds, with the added ability of trading like a stock. Definitionally, an ETF is a pooled investment security that contains multiple underlying assets. ETFs may hold all types of investments: stocks, commodities, or bonds. Most ETFs are structured around a specific asset class, or investment strategy (Chen, 2022). Valuations on ETFs are different from those on individual stocks. The value of an ETF is determined by the value of the assets it holds.

To begin, I will create a screener that searches for ETFs fitting certain parameters. From the roughly 3000 ETFs listed on the NYSE; this screener will select ETFs that are undervalued or properly valued, liquid, and trending upwards (New York Stock Exchange, 2022). To determine a valuation the screener will consider five main factors: Net Asset Value, total and estimated cash, Intraday Indicative Value, shares outstanding, and accrued dividends (Abner, 2022). To measure liquidity, the screener will analyze the trading volume of the securities held by the ETF and the trading volume of the ETF itself (Artzberger, 2022). The performance of an ETF will be measured by its historical return and risk, and its relative performance. In this stage, ETFs are evaluated through fundamental analysis considering three main factors: valuation, liquidity, and past performance. This analysis serves to decrease my list of available securities by a factor of around 15.

After screening ETFs, I can begin to develop relevant beliefs about the future performance of the selected securities (Markowitz, 1952). To create these beliefs, sentiment analysis, and technical financial analysis will be conducted using a two-stage deep learning model (Jing et al., 2021; Yun et al., 2020). Sentiment analysis uses natural language processing (NLP) to quantify subjective information like people’s opinions or appraisals (Zhang et al., 2018). Technical analysis is a method used to forecast future performance, through the evaluation of statistical trends in price and volume data. In the first stage of the model, I will quantify investor attitudes toward selected ETFs using data from the Twitter API (Twitter, 2022). In the second stage, sentiment predictions will be used alongside technical indicator data to predict risk and return over time. The predictions generated by the two-stage model serve as relevant beliefs about future performance. Once these predictions are generated, I will have completed the first stage of portfolio selection (Markowitz, 1952).

The second stage of portfolio selection begins with beliefs about future performance and ends with an investor choosing a portfolio (Markowitz, 1952). To choose a portfolio I will perform mathematical optimization with custom constraints following ideas derived from Modern Portfolio Theory (MPT) (Rom & Ferguson, 1993). MPT attempts to build diversified portfolios that optimize the balance of risk and return (Markowitz, 1952).

Now that the system has selected a portfolio, I must submit buy and sell orders. Under certain time conditions, the success of a buy or sell order is heavily dependent on external events, and time of day (Ma et al., 2021; Wang et al., 2020). To predict the best time to submit orders, I will employ a deep learning model for time series classification (Ismail Fawaz et al., 2019). This model will generate trading signals fueled by, macro, technical and sentiment analysis (Jing et al., 2021; Yun et al., 2020). Once the model generates a buy or sell signal, the system will place orders through the TD Ameritrade API. Placing trades completes the process of selecting assets, optimizing the mix of assets, and acquiring the assets. Completing this process achieves the goal of creating a holistic automated trading system backed by accessible tools. In doing so, I will have built a framework for researchers that can be adapted to new problems and improved upon over time.

**Research Timeline and Dissemination**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Jan. | Feb. | Mar. | Apr. | May |
| Data Collection |  |  |  |  |  |
| Model Development |  |  |  |  |  |
| System Tests |  |  |  |  |  |
| Model Deployment |  |  |  |  |  |
| Documentation |  |  |  |  |  |

Through the 2021-2022 academic year, I completed multiple projects related to this research. Project titles include: “Computational Portfolio Optimization Employing MPT” and “Ethereum Price Prediction Using Tree-Based ML.” I plan to present my work at the Open Data Science Conference (OSDC) East 2023. The conference will take place from May 9th to 11th, 2023, in Boston, MA. OSDC welcomes students, academics, and professionals to present research, train, and learn from experts in the field (Open Data Science Conference, 2022).

**References**

Abner, D. J. (2016). *The ETF Handbook: How to Value and Trade Exchange Traded Funds*. John Wiley & Sons.

*Algorithmic Trading Market Size, Statistics, Trends | 2022—27 | Industry Growth*. (n.d.). Mordor Intelligence. Retrieved September 9, 2022, from <https://www.mordorintelligence.com/industry-reports/algorithmic-trading-market>

Artzberger, W. (2022, January 31). *ETF Liquidity: Why It Matters*. Investopedia. <https://www.investopedia.com/articles/exchangetradedfunds/08/etf-liquidity.asp>

Chen, J. (n.d.). *What Is an Exchange Traded Fund (ETF)?* Investopedia. Retrieved September 9, 2022, from <https://www.investopedia.com/terms/e/etf.asp>

Chen, J. (2022, June 11). *Active Management Definition*. Investopedia. <https://www.investopedia.com/terms/a/activemanagement.asp>

Deboeck, G. J. (1994). *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*. John Wiley & Sons.

*Deploying Docker containers on Azure*. (2022, September 9). Docker Documentation. <https://docs.docker.com/cloud/aci-integration/>

Huang, B., Huan, Y., Xu, L. D., Zheng, L., & Zou, Z. (2019). Automated trading systems statistical and machine learning methods and hardware implementation: A survey. *Enterprise Information Systems*, *13*(1), 132–144. <https://doi.org/10.1080/17517575.2018.1493145>

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963. <https://doi.org/10.1007/s10618-019-00619-1>

Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019. <https://doi.org/10.1016/j.eswa.2021.115019>

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, *165*, 113973. <https://doi.org/10.1016/j.eswa.2020.113973>

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. <https://doi.org/10.2307/2975974>

*NYSE Arca Quarterly ETF Report*. (2022, June 30). New York Stock Exchange. <https://www.nyse.com/etf/exchange-traded-funds-quarterly-report>

*ODSC East 2023 | Open Data Science Conference*. (n.d.). Open Data Science Conference. Retrieved October 6, 2022, from <https://odsc.com/boston/>

Rom, B. M., & Ferguson, K. W. (1993). *POST-MODERN PORTFOLIO THEORY COMES OF AGE*. 16.

Rountree, D., & Castrillo, I. (2014). Chapter 1—Introduction to the Cloud. In D. Rountree & I. Castrillo (Eds.), *The Basics of Cloud Computing* (pp. 1–17). Syngress. <https://doi.org/10.1016/B978-0-12-405932-0.00001-3>

*TD Ameritrade for developer | APIs*. (n.d.). TD Ameritrade. Retrieved September 9, 2022, from <https://developer.tdameritrade.com/apis>

*Twitter API Documentation*. (n.d.). Twitter Developer. Retrieved October 6, 2022, from <https://developer.twitter.com/en/docs/twitter-api>

Wang, J.-Z., Wang, J.-J., Zhang, Z.-G., & Guo, S.-P. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, *38*(11), 14346–14355. <https://doi.org/10.1016/j.eswa.2011.04.222>

Wang, W., Li, W., Zhang, N., & Liu, K. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications*, *143*, 113042. <https://doi.org/10.1016/j.eswa.2019.113042>

Yun, H., Lee, M., Kang, Y. S., & Seok, J. (2020). Portfolio management via two-stage deep learning with a joint cost. *Expert Systems with Applications*, *143*, 113041. <https://doi.org/10.1016/j.eswa.2019.113041>

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, *8*(4), e1253. <https://doi.org/10.1002/widm.1253>