**Mohammad Sadiq Warsi, 001232055**

BSc (Hons) in Computer Science (Data Science)

**Provisional Title:** Machine Learning and Sentiment Analysis based Portfolio Management

**Introduction**

Recently, researchers and financial practitioners have focused on the integration of Machine Learning (ML) into portfolio management. (Zhang, et al, 2024) The necessity to explore innovative approaches to portfolio management is based on the increasing complexity of financial markets, in addition to the growing influence of social media and news on investor behaviour, and because traditional financial models often struggle to account for the rapid shifts in market sentiment. This makes the exploration of sentiment analysis particularly relevant (Baker, Bloom & Davis, 2023).

**Machine Learning** (ML) enables the development of predictive models to analyze large financial datasets**.** These models can identify patterns and trends that are not immediately apparent, providing valuable insights for portfolio management. Researchers find that ML models, such as LSTM and Support Vector Machines, have demonstrated their effectiveness in predicting financial market trends (Priyadarshini & Cotton). **Sentiment Analysis**, in this context, a subset of ML, involves analysing textual data from sources such as news articles and social media to gauge market sentiment. This information can be used to predict market movements and inform investment decisions. **Financial Modelling** traditionally involves using historical data to forecast future market behaviour and optimize investment strategies. (Diaz de Arce, 2024) We can create more adaptive and responsive models that better reflect current market conditions, by integrating sentiment analysis with financial modelling (Anjum et al, 2024). **Portfolio Management** involves selecting and managing a mix of investment assets to achieve specific financial goals while minimizing risk.

**Purpose:** The purpose of this project is to develop an interactive dashboard that leverages ML and sentiment analysis to optimize portfolio management. By integrating real-time sentiment data from financial news and social media with historical market data, the dashboard aims to provide financial analysts and individual investors with valuable insights to make informed investment decisions. This approach allows for a potentially better decision-making and better risk management. This project seeks to bridge the gap between traditional financial modelling and modern data-driven approaches, enhancing the decision-making process in portfolio management.

**Target Users**: The final product, an interactive dashboard, is designed for financial analysts and individual investors. It will allow users to visualize sentiment data, track portfolio performance, and make informed investment decisions based on real-time market insights.

**Objectives**

* **Develop a Real-Time Sentiment Analysis Tool**: Creating a tool that captures and analyses sentiment data from financial news and social media in real time. The tool will combine sentiment data with historical market data to create a comprehensive dataset for analysis. Based on the evaluation success of the proposed product before time, can consider integrating a multi-dimension sentiment scoring.[[1]](#footnote-2)
* **Create Predictive Models for Portfolio Optimization**: Develop and validate machine learning models that predict market trends and optimize portfolio allocations based on sentiment and market data.
* **Design an Interactive Dashboard**: Building an interactive dashboard that allows financial analysts and individual investors to visualize sentiment data, track portfolio performance, and make informed investment decisions.

**Problem Domain**

The finance and investment sector faces several key challenges that have received significant interest lately. These include market volatility, the influence of public sentiment on stock prices, and the difficulty of optimizing portfolios under uncertain conditions. Addressing these issues is crucial for improving decision-making processes and enhancing the overall stability of financial markets.

#### **Impact of Market Volatility**

Market volatility refers to the rapid and unpredictable changes in asset prices. Various factors, including economic data releases, geopolitical events, and changes in investor sentiment, drive this volatility. For investors, market volatility presents significant challenges, such as financial losses and increased risk aversion (Baker, Bloom & Davis, 2023).

Sudden price fluctuations can result in substantial financial losses for investors who are unable to react quickly to market changes. High volatility can also increase risk aversion, leading to reduced market participation and liquidity. During the COVID-19 pandemic, stock markets experienced extreme volatility, causing significant losses for many investors (Baker, Bloom & Davis, 2023).

Traditional approaches to sentiment analysis include lexicon-based methods, machine learning methods, and deep learning methods. Lexicon-based methods are straightforward and easy to implement but struggle with context and nuance, failing to adapt to new words or phrases (Sharma, Ali & Kabir, 2024). This causes issues in analysing important Portfolio data. Machine learning models, such as Support Vector Machines (SVM) and Naive Bayes, can handle large datasets and capture complex patterns in the data. However, they require extensive labelled data for training and struggle with interpretability (Sharma, Ali & Kabir, 2024). Deep learning techniques like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks excel at capturing intricate patterns in text data but are computationally intensive and require significant resources for training (Priyadarshini & Cotton, 2021), which may not be readily available. Sharma et al(2024) find that by integrating machine learning models like SVM and Random Forests, investors can analyze large datasets and capture and predict market volatility more accurately compared to traditional models.

**Influence of Public Sentiment**

Public sentiment, shaped by social media and news outlets, plays a crucial role in influencing stock prices. This is because positive or negative sentiment can drive investor behaviour, leading to price movements that may not be justified by fundamental factors.

In other words, investors may be influenced by prevailing sentiment, leading to herd behaviour and irrational decision-making. Sentiment-driven price movements can create profit opportunities and increase the risk of losses if the sentiment shifts unexpectedly. For instance, Elon Musk's tweets about Tesla have caused its stock price to rise or fall sharply, demonstrating the power of public sentiment (Cristescu et al., 2023).

Researchers find that models like FinBERT[[2]](#footnote-3) are particularly effective in capturing nuanced sentiments related to market movements (Cristescu et al., 2023).

**Limitations of Traditional Models in Portfolio Management**

Traditional models, such as the mean-variance optimization (MVO), often rely on historical data and assume fixed parameters such as asset returns are normally distributed and that correlations between assets are constant over time. These assumptions often do not hold in real-world markets, leading to suboptimal portfolios (Fama, 1998). These models may not adequately capture the dynamic nature of financial markets and the influence of sentiment. Relying on static models can result in suboptimal portfolios that do not adapt to changing market conditions. Capital Asset Pricing Model (CAPM) offers a simple way to estimate the expected return of an asset based on its risk relative to the market but relies on assumptions that may not be realistic, such as the idea that investors hold diversified portfolios and that markets are efficient. Alternative models like the Black-Litterman model incorporate investor views and market equilibrium, providing more flexibility and realism in terms of actual performance. However, these models can be complex to implement and require subjective inputs, which can introduce bias (Idzorek, 2024). Moreover, failure to account for sentiment and other dynamic factors can increase the risk of significant losses. For example, during the 2008 financial crisis, many traditional portfolio models failed to predict the market crash, leading to significant losses for investors.

**Gaps in Current Solutions**

Current solutions often fail to integrate real-time sentiment analysis with financial modelling, limiting their ability to adapt to rapid market changes (Shen and Zhang, 2024). Traditional models do not account for the influence of public sentiment, leading to suboptimal investment decisions (Islam et al., 2024). There is a need for more adaptive and responsive models that can leverage real-time data to optimize portfolio management. Despite advancements, there are still challenges in integrating real-time sentiment analysis with financial modelling. Issues such as data quality, model interpretability, and computational requirements need to be addressed (Islam et al.,2024)

**Proposed Project**

The project aims to help investors manage market volatility by providing real-time sentiment analysis. The dashboard would allow users to monitor market sentiment and make informed decisions based on current conditions by integrating sentiment data from financial news and social media. The approach leverages FinBERT, a model specifically designed for financial sentiment analysis, and LSTM networks. This combination is chosen as it aims to capture nuanced sentiments related to market movements more effectively than traditional methods.

The proposed approach integrates sentiment analysis with machine learning-driven predictive modelling, this is done to create a more adaptive and responsive portfolio management system. The approach will incorporate real-time sentiment data and advanced risk metrics like CVaR, and better align portfolios with current market conditions and investor sentiment (Schulmerich et al, 2014), to address the limitations of static models. However, the complexity of integrating multiple data sources and models can be challenging (Odeyemi, et al, 2024). Ensuring the accuracy and reliability of sentiment data is crucial, as errors in sentiment analysis can lead to poor investment decisions (Adelakun, 2023).

**Methodology**

The project employs an approach that integrates sentiment analysis with machine learning-driven predictive modelling to optimize portfolio management.

Starting with sentiment analysis, which is a crucial component of our methodology. Sentiment analysis involves using Natural Language Processing (NLP) techniques to analyze and interpret the sentiment expressed in textual data. Models like FinBERT[[3]](#footnote-4), specifically designed for financial sentiment analysis, capture nuanced sentiments related to market movements (Priyadarshini & Cotton, 2021). FinBERT is fine-tuned for financial text, making it more accurate for this specific domain, while LSTM networks can capture temporal dependencies, providing a more dynamic analysis of sentiment trends over time (Requejo, 2024). However, this approach may still face challenges with interpretability and computational requirements, relying on the availability of large datasets for training and validation. ML models like Recurrent Neural Networks (RNN) and Long Short-Term Memory [[4]](#footnote-5)(LSTM) networks are prevalent for processing sequential data (Footnotes has the sample codes for the models). These models excel in capturing temporal dependencies in sentiment data, which allows for the analysis of sentiment trends over time (Cristescu et al., 2023).

For predictive modelling, ML models like Support Vector Machines (SVM) and Random Forests. SVM is chosen because it is a powerful algorithm for classification tasks, making it suitable for predicting market trends based on sentiment data. Random Forests Model could also be used, based on the results of the initial model, as Random Forests combine multiple decision trees to improve prediction accuracy and robustness (Huang et al., 2024). Based on the evaluation, other models like Gradient Boosting and Bagging, which are ensemble methods[[5]](#footnote-6), could be used to further enhance prediction accuracy by combining multiple models, thus improving performance and helping against overfitting.

Sentiment data will be sourced from Twitter and financial news APIs, providing a continuous stream of information. Market data, including stock prices and trading volumes, will be obtained from reliable financial databases such as Bloomberg or Yahoo Finance. Integrating these sources allows us to create a comprehensive dataset that captures both market movements and public sentiment.

The project’s workflow consists of several key steps:

1. **Data Collection: (Test Bed has the details)**
2. **Data Preprocessing:** The data will be cleaned and pre-processed to ensure it is suitable for analysis, and to address reliance on large datasets. This includes removing noise from social media data and normalizing market data. Then, data will be transformed accordingly (e.g. scaling, rotation) to create new samples.
3. **Model Training**: The FinBERT and LSTM models will be trained on the sentiment data, and the SVM, Random Forest, and other models on the combined sentiment and market data. Would fine-tune them on our specific dataset to reduce the need for a large volume of data.
4. **Model Validation**: The models will be validated using cross-validation techniques and back-testing against historical data to ensure accuracy and reliability.
5. **Integration:** The trained models will be integrated into an interactive dashboard that provides real-time sentiment analysis and predictive insights for portfolio management.

**Test-Bed**: Historical stock market data and sentiment data from financial news and social media will serve as the test bed for model training and validation. Alpha Vantage API[[6]](#footnote-7), Twitter API[[7]](#footnote-8), Financial News Dataset (FNSPID)[[8]](#footnote-9), Google News API[[9]](#footnote-10), et cetera will be used.

**Data Sources and Their Usage in the Project**

We will use the Yahoo Finance API to gather historical stock prices, trading volumes, and financial metrics. This data will help analyze past market trends, train predictive models, and create features like moving averages and volatility indices.

The Google Finance Dataset [[10]](#footnote-11) provides financial data like daily closing prices, trading volumes, and historical performance metrics for individual stocks. It also includes market indices, sector performance, and economic indicators. This dataset will help us understand broader market trends, validate predictive models, and visualize key metrics on the dashboard.

For real-time data, the Alpha Vantage API will deliver up-to-date stock prices, forex, and cryptocurrency data. This real-time data is crucial for providing current market insights, updating the dashboard, and dynamically adjusting models based on market conditions.

The IEX Cloud API [[11]](#footnote-12) offers both real-time and historical market data. We will combine past and present data for analysis, train and validate models for market trends and display trends on the dashboard.

Using the Twitter API, we will collect real-time tweets related to financial markets to analyze public sentiment. This sentiment data will train models to predict market movements based on public opinion and provide sentiment insights on the dashboard.

The Financial News Dataset (FNSPID) includes sentiment scores for news articles, categorized as positive, negative, or neutral, along with historical stock prices linked to news events. This will help gauge market sentiment from news articles, improve model accuracy by integrating sentiment data, and visualize news impact on the dashboard.

NewsAPI will give us news articles from sources like Reuters, Bloomberg, and CNBC. This data will be used for sentiment analysis, providing real-time news updates on the dashboard, and enhancing model predictions with the latest market information.

Finally, the Google News API provides news articles for sentiment analysis relevant to financial markets. This real-time sentiment data will be displayed on the dashboard and incorporated into model training to reflect current market conditions.

**Programming Framework:** Python will be used as the primary programming language, and libraries such as TensorFlow and PyTorch for deep learning, and sci-kit-learn for traditional machine learning algorithms will be used. These tools provide robust support for building and training complex models (Huang et al., 2024). ‘Plotly (Dash)’, ‘Flask’, ‘Pandas’, and other appropriate libraries for the dashboard design. Relevant existing algorithms from sources will be used, and modified (parameter fine-tuning, GridSearch hyperparameter, etc) for our project.

Reinforcement learning[[12]](#footnote-13) will be used to dynamically adjust portfolio allocations based on changing market conditions and sentiment factors. This is so because this methodology has proven effective in optimizing asset allocations in response to market dynamics, and so it will help manage the impact of market volatility. Other models (links mentioned earlier) like SVM, and ensemble methods like the Random Forests model will be used based on the evaluation success. According to Anese et al. (2023), the integration of machine learning models with traditional financial models resulted in a 15% improvement in portfolio returns and a 20% reduction in risk compared to traditional models alone.

For sentiment analysis models like FinBERT and LSTM will be used at first and evaluated for performance. Then other models could be considered.

After the success of the models and the dashboard, will consider integrating advanced risk metrics like Conditional Value at Risk (CVaR) and Expected Shortfall with traditional metrics which could enhance our decision-making process. This is because these metrics help capture tail risks[[13]](#footnote-14) and provide a more comprehensive view of potential risks like risks of liquidity, credit, and operational. Additionally, incorporating real-time sentiment data into models allows for adaptive portfolio management that responds to market shifts and are effective risk management method. (Baker, Bloom & Davis, 2023).

**Simulator**: A simulation environment will be created to back-test portfolio strategies against historical data, allowing for the evaluation of performance metrics like the Sharpe ratio and cumulative returns. Here’s a good code example[[14]](#footnote-15) of the simulator our project will draw from.

1. **Environment Setup**: The state space will include features such as current portfolio allocation, market sentiment scores, and historical price data. The action space will include possible adjustments to the portfolio, such as buying or selling specific assets.
2. **Simulation Engine**: A financial market simulation engine like MarS will be used, to simulate reasonably realistic market conditions and order flows. This engine allows for the generation of target scenarios and modelling of market impacts.
3. **Training and Testing**: Reinforcement learning algorithms will be used to train the RL agent in the simulated market environment. The agent will learn to optimize portfolio allocations based on the simulated market conditions and sentiment data.
4. **Evaluation**: The performance of the RL agent will be evaluated using metrics such as returns, risk (e.g., Conditional Value at Risk - CVaR), and alignment with market sentiment.

**Real Device**: Developing an interactive dashboard tailored for financial analysts and individual investors will enhance usability and accessibility because the focus on user experience ensures that the insights generated by the model can be effectively utilized in decision-making processes. The dashboard will allow users to visualize sentiment data, track portfolio performance, and make informed investment decisions.

**Dashboard Development Steps**:

* + **Design Layout**: the main content of the dashboard will be headings, a market trends section, a sentiment analysis section, and a portfolio performance section. Interactive elements like dropdown menus to select stocks, periods, etc, sliders to adjust date ranges etc, and buttons to refresh data can be added based on needs.
  + **Data Integration**: The pre-processed and cleaned data will be integrated into the dashboard. This includes real-time sentiment data from Twitter and financial news, as well as historical and real-time market data. In addition to the predicted portfolio performance, sentiment information, etc.
  + **Visualization**: Interactive visualizations such as line charts, bar charts, and heatmaps will be used to display sentiment trends, market performance, and portfolio analytics (from the ML models). **Interactivity**: This will implement interactive features such as dropdown menus, sliders, and buttons to allow users to customize their views and interact with the data.
  + **Deployment**: Will deploy the dashboard on a web server using Flask (example code)[[15]](#footnote-16), making it accessible to users via a web browser

The project will be managed on Azure DevOps(link)[[16]](#footnote-17) appropriate permissions have been given to the supervisor as of now.

**Standards**

We will adhere to the following standards to ensure the reliability, validity, and ethical handling of data:

1. [FINRA Guidelines](https://www.finra.org/rules-guidance) for financial data processing and analysis will be followed to ensure compliance with industry regulations.
2. Will Adhere to ISO standards for data management and information security, specifically [ISO 27001](https://www.iso.org/standard/27001) for information security management and [ISO 9001](https://www.iso.org/standard/27001) for quality management systems.
3. Ensuring compliance with GDPR ([GDPR Guidelines](https://gdpr.eu/)) for data privacy and protection, particularly when handling users’ data from social media sources.
4. Following appropriate practices including model validation, cross-validation, and regularization techniques to prevent overfitting and ensure robust model performance.

The methodology builds on established frameworks, such as the Black-Litterman model[[17]](#footnote-18), which have been empirically validated in the literature. This foundation increases the credibility and reliability of our approach (Idzorek, 2024).

**Evaluation**

The evaluation of our project is to ensure that the models and the interactive dashboard perform effectively and meet the project objectives. This section outlines the methods that will be used to assess the success of the project.

**Evaluation Metrics**

To measure the success of the models and the dashboard, several key metrics will be used:

* **Accuracy**: This metric will measure the percentage of correct predictions made by the model. High accuracy would indicate that the model is reliable in predicting market trends based on sentiment data (Huang et al., 2024).
* **Precision and Recall**: Precision evaluates the model’s ability to correctly identify positive sentiments, while recall measures its ability to identify negative sentiments. These metrics are crucial for understanding the model’s performance (Priyadarshini & Cotton, 2021).
* **F1 Score**: The F1 score combines precision and recall into a single metric, providing a balanced view of the model’s performance, especially in cases of class imbalance (Cristescu et al., 2023).
* **Sharpe Ratio**: This metric assesses the risk-adjusted return of the portfolio strategies generated by the model. A higher Sharpe ratio indicates better performance relative to risk (Baker, Bloom & Davis, 2023).
* **Cumulative Returns**: This metric tracks the total returns generated by the portfolio over time, allowing for comparison against a benchmark index to assess relative performance (Fama, 1998).
* **Maximum Drawdown**: This metric measures the largest peak-to-trough decline in portfolio value, helping to evaluate the risk exposure of the investment strategy (Bailey et al., 2016).

**Model Validation**

To ensure the reliability and accuracy of our models, we will use the following validation methods:

* **Cross-Validation**: We will use cross-validation techniques to assess the performance of our models. This involves dividing the dataset into multiple subsets and training the model on each subset while validating it on the remaining data. This method helps in identifying any overfitting or underfitting issues (Zhang et al., 2020).
* **Back-Testing**: We will back-test our models against historical data to evaluate their performance in real-world scenarios. This process involves simulating the model’s predictions on past data to see how well it would have performed (Lo, 2002).

**User Feedback** (For the device)

We will use the following methods to collect and utilize feedback:

* **Surveys and Interviews**: We will conduct surveys and interviews with users, alpha testers and volunteers, to gather qualitative feedback on the usability and effectiveness of the dashboard. This feedback will help us identify any areas for improvement.
* **Usability Testing**: We will conduct usability testing sessions where users interact with the dashboard. These sessions will help us identify any usability issues and make necessary adjustments to enhance the user experience.

**Performance Evaluation**

The performance of the models and the dashboard will be continuously monitored and evaluated using the following methods:

* **Real-Time Testing**: The performance of the models will be monitored in real-time to ensure they are accurately predicting market trends and providing valuable insights (Huang et al., 2024).
* **Comparison with Benchmarks**: We will compare the performance of our models against standard benchmarks to assess their effectiveness. This comparison will help us understand how well our models perform relative to existing solutions (Fama, 1998).

**Risk Management**

Effective risk management is crucial for the success of our project. We will use the following strategies to manage risk:

* **Risk Metrics**: We will use advanced risk metrics like Conditional Value at Risk (CVaR) and Expected Shortfall to assess the potential risks associated with our portfolio strategies. These metrics provide a substantial view of potential risks and help in making informed decisions (Baker, Bloom & Davis, 2023).
* **Adaptive Management**: We will adjust our strategies based on real-time data to ensure that our portfolio management approach remains adaptive and responsive to market changes that are happening in real time.(Anese et al., 2023).

**Feasibility and Planning**

**Project Management Plan**

**Introduction** This project management plan outlines the key activities, timelines, resources, and risk management strategies for the development and implementation of the project. The plan ensures that the project meets the specified objectives on time. (Table 1 has the project timeline)

**Project Objectives** **& Scope** : (Defined in the Introduction)

**Project Timeline**

Table The Project outline

| **Phase** | **Task Description** | **Start Date** | **End Date** | **Duration** |
| --- | --- | --- | --- | --- |
| **Phase 1: Planning** | Define project scope, objectives, and requirements | Oct 10, 2024 | Nov 13, 2024 | 34 days |
| **Phase 2: Data Collection** | Collect historical and real-time sentiment and market data | Nov 13, 2024 | Dec 20, 2024 | 37 days |
| **Phase 3: Data Preprocessing** | Clean and preprocess data for analysis | Dec 20, 2024 | Jan 07, 2025 | 17 days |
| **Phase 4: Model Development** | Develop sentiment analysis and predictive models | Jan 08, 2025 | Feb 07, 2025 | 30 days |
| **Phase 5: Model Validation** | Validate models using cross-validation and back-testing | Feb 08, 2025 | Feb 22, 2025 | 14 days |
| **Phase 6: Integration** | Integrate models into the interactive dashboard | Feb 22, 2025 | Mar 24, 2025 | 30 days |
| **Phase 7: User Testing** | Conduct usability testing and gather feedback | Mar 25, 2025 | Apr 09, 2025 | 12 days |
| **Phase 8: Final Adjustments** | Make final adjustments based on feedback | Apr 09, 2025 | Apr 17, 2025 | 8 days |
| **Phase 9: Project Completion** | Final review and project submission | Apr 17, 2025 | Apr 22, 2025 | 5 days |

**Resources Required**

* **Software Tools**: Python, TensorFlow, PyTorch, sci-kit-learn, data visualization tools, and financial data APIs.
* **Hardware**: Reasonable computing resources for model training and testing. A combination of university and personal resources would be sufficient.
* **Data Sources**: (mentioned earlier)

**Risks & Mitigation**

* **Data Quality Issues**: This would ensure data is cleaned and pre-processed to remove noise and inconsistencies.
* **Model Overfitting**: Using cross-validation and regularization techniques to prevent overfitting.
* **Technical Challenges**: Allocating time for troubleshooting and technical support.
* **Timeline Delays**: Monitoring progress regularly and adjusting the timeline as needed to stay on track.
* **User Feedback**: Incorporate user feedback promptly to improve the dashboard's usability and functionality.

**Communication Plan**

* **Weekly Meetings**: Hold weekly meetings with the project tutors to review progress and address any issues.
* **Feedback Sessions**: Conduct feedback sessions with users during the testing phase.

**References**

1. Anese, G., Corazza, M., Costola, M., & Pelizzon, L. (2023) 'Impact of Public News Sentiment on Stock Market Index Return and Volatility', *Computational Management Science*. Available at: https://doi.org/10.1007/s10287-023-00454-2 (Accessed: 28 September 2024).
2. Bailey, D. H., Borodin, A., & Kauffman, R. J. (2016) 'Maximum drawdown: A new measure of risk', *Journal of Risk and Financial Management*, 9(2), pp. 1-16. Available at: https://doi.org/10.3390/jrfm9020016 (Accessed: 30 September 2024).
3. Baker, S. R., Bloom, N., & Davis, S. J. (2023) 'Economic Policy Uncertainty and Stock Market Volatility', *Journal of Financial Economics*. Available at: https://doi.org/10.1016/j.jfineco.2023.01.004 (Accessed: 5 October 2024).
4. Cristescu, M. P., Mara, D. A., Nerișanu, R. A., Culda, L. C., & Maniu, I. (2023) 'Analyzing the Impact of Financial News Sentiments on Stock Prices', *Mathematics*, 11(23), p. 4830. Available at: https://doi.org/10.3390/math11234830 (Accessed: 15 October 2024).
5. Fama, E. F. (1998) 'Market efficiency, long-term returns, and behavioural finance', *Journal of Financial Economics*, 49(3), pp. 283-306. Available at: https://doi.org/10.1016/S0304-405X(98)00026-9 (Accessed: 20 October 2024).
6. Huang, S., Cao, L., Sun, R., Ma, T., & Liu, S. (2024) 'Enhancing Portfolio Optimization: A Two-Stage Approach with Deep Learning', *Mathematics*, 12(21), p. 3376. Available at: https://doi.org/10.3390/math12213376 (Accessed: 25 October 2024).
7. Idzorek, T. M. (2024) *A step-by-step guide to the Black-Litterman model*. Zephyr Associates. Available at: https://www.zephyrassociates.com/black-litterman-guide (Accessed: 30 October 2024).
8. Lo, A. W. (2002) 'The statistics of financial market trading', *Journal of Financial Markets*, 5(2), pp. 1-29. Available at: https://doi.org/10.1016/S1386-4181(02)00025-7 (Accessed: 5 November 2024).
9. Nielsen, J. (2012) *Usability engineering*. Morgan Kaufmann. Available at: https://www.elsevier.com/books/usability-engineering/nielsen/978-0-12-518406-9 (Accessed: 10 November 2024).
10. Priyadarshini, I., & Cotton, C. (2021) 'A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis', *The Journal of Supercomputing*. Available at: https://doi.org/10.1007/s11227-021-03838-w (Accessed: 15 November 2024).
11. Rubin, J., & Chisnell, D. (2008) *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley Publishing. Available at: https://www.wiley.com/en-us/Handbook+of+Usability+Testing%3A+How+to+Plan%2C+Design%2C+and+Conduct+Effective+Tests%2C+Second+Edition-p-9780470185483 (Accessed: 25 November 2024).
12. Sharpe, W. F. (1994) 'The Sharpe ratio', *The Journal of Portfolio Management*, 21(1), pp. 49-58. Available at: https://doi.org/10.3905/jpm.v21i1.409501 (Accessed: 30 November 2024).
13. Odeyemi, O., Mhlongo, N.Z., Nwankwo, E.E. and Okoye, C.C. (2024) 'Big data applications in portfolio management: A review of techniques and strategies', *International Journal of Science and Research Archive*, 11(01), pp. 1984-1999. Available at: https://ijsra.net/sites/default/files/IJSRA-2024-0268.pdf (Accessed: 5 December 2024).
14. Zhang, Y. and Li, X., 2024. Enhancing Markowitz's Portfolio Selection Paradigm with Machine Learning. *Annals of Operations Research*, 62(3), pp. 257-276. Available at: https://link.springer.com/article/10.1007/s10479-024-06257-1 [Accessed 01 Dec. 2024]
15. Islam, M.S., Kabir, M.N., Ab Ghani, N., Zamli, K.Z., Zulkifli, N.S.A., Rahman, M.M. and Moni, M.A., 2024. Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. *Artificial Intelligence Review*, 57, article number 62. Available at: https://link.springer.com/article/10.1007/s10462-023-10651-9 [Accessed 15 December 2024].
16. Shen, Y. and Zhang, P.K., 2024. Financial Sentiment Analysis on News and Reports Using Large Language Models and FinBERT. arXiv preprint arXiv:2410.01987. Available at: https://doi.org/10.48550/arXiv.2410.01987 [Accessed 15 December 2024].
17. Diaz de Arce, R., 2024. Financial Modelling and Market Predictions. *Journal of Financial Modelling*, 15(2), pp. 123-145. Available at: https://doi.org/10.1016/j.jfinmod.2024.02.003 [Accessed 15 December 2024].
18. Anjum, A., Patel, R., and Singh, K., 2024. Integrating Sentiment Analysis with Financial Modelling: A Comprehensive Review. *International Journal of Financial Studies*, 12(3), pp. 200-220. Available at: https://doi.org/10.3390/ijfs12030020 [Accessed 15 December 2024].
19. Sharma, R., Ali, S., and Kabir, M., 2024. Advances in Sentiment Analysis for Financial Markets. *Journal of Computational Finance*, 18(1), pp. 50-75. Available at: https://doi.org/10.3905/jcf.2024.18.1.050 [Accessed 15 December 2024].
20. Requejo, J., 2024. Temporal Dependencies in Financial Sentiment Analysis. *Journal of Financial Data Science*, 6(4), pp. 89-105. Available at: https://doi.org/10.3905/jfds.2024.6.4.089 [Accessed 15 December 2024].
21. Schulmerich, M., Leporcher, Y., and Eu, C.H., 2014. Applied Asset and Risk Management: A Guide to Modern Portfolio Management and Behavior-Driven Markets. *Springer*. Available at: https://doi.org/10.1007/978-3-642-54593-2 [Accessed 15 December 2024].
22. Adelakun, A., 2023. Ensuring Accuracy in Financial Sentiment Analysis. *Journal of Financial Analytics*, 11(2), pp. 130-145. Available at: https://doi.org/10.3905/jfa.2023.11.2.130 [Accessed 15 December 2024].
23. Zhang, T., Li, H., and Wang, Y., 2020. Machine Learning Approaches in Financial Market Prediction. *Journal of Financial Technology*, 8(3), pp. 210-230. Available at: https://doi.org/10.1016/j.jfintech.2020.08.003 [Accessed 15 December 2024].

1. **Multi-dimensional sentiment scoring** is an approach to sentiment analysis that does more than the binary classification of sentiments (positive and negative). It evaluates multiple dimensions of sentiment, such as: **Valence** (The positivity or negativity of the sentiment) **Arousal** (The intensity of the emotion e.g., excited vs. calm), and **Dominance** (The degree of control or influence e.g., powerful vs. weak). [↑](#footnote-ref-2)
2. FinBERT, a transformer-based model fine-tuned for financial text, provides more accurate sentiment analysis in the financial domain. [↑](#footnote-ref-3)
3. <https://github.com/ProsusAI/finBERT> [↑](#footnote-ref-4)
4. <https://github.com/keras-team/keras/blob/master/examples/imdb_lstm.py> [↑](#footnote-ref-5)
5. **Ensemble methods** combine multiple machine learning models to improve overall performance. [↑](#footnote-ref-6)
6. [Alpha Vantage](https://www.alphavantage.co/) [↑](#footnote-ref-7)
7. [Twitter API](https://developer.x.com/en/docs/x-api) [↑](#footnote-ref-8)
8. [FNSPID](https://github.com/Zdong104/FNSPID_Financial_News_Dataset) [↑](#footnote-ref-9)
9. <https://newsapi.org/docs/get-started> [↑](#footnote-ref-10)
10. [Google Finance](https://www.google.com/finance/) [↑](#footnote-ref-11)
11. [IEX Cloud API](https://iexcloud.io/) [↑](#footnote-ref-12)
12. [Relevant example of RL model used similarly](https://github.com/CFMTech/Deep-RL-for-Portfolio-Optimization) [↑](#footnote-ref-13)
13. Tail risks refer to the risk of extreme events that occur at the tails of a probability distribution [↑](#footnote-ref-14)
14. <https://github.com/microsoft/MarS/> [↑](#footnote-ref-15)
15. <https://codemax.app/snippet/developing-a-real-time-dashboard-with-plotly-and-flask-in-python/> [↑](#footnote-ref-16)
16. [Repository Link](https://dev.azure.com/ms0205y/COMP1682%20Final%20Year%20Project) [↑](#footnote-ref-17)
17. This is because the model combines CAPM with investor views to generate more stable and realistic portfolio allocations. [↑](#footnote-ref-18)