Machine Learning in Finance:

Foundations and Practical Applications

Alexander J Zevin

The University of Texas at Austin

Graduate Program, College of Natural Sciences

Case Studies in Machine Learning

Instructor: Dr. Junfeng Jiao

December 2024

Machine Learning in Finance

Introduction

The financial industry is the crossroads of relentless innovation and established practices which often create barriers to innovation. These sectors – banks, insurance companies, asset management, payment systems, and capital markets – are often plagued by regulatory barriers, risk minimization, and aging computer systems that can stunt adoption of new computational applications. For example, in response to the Great Depression, financial institutions within the United States became increasingly compelled to comply with legal guidelines resulting from the Securities Act of 1933, which introduced the Securities and Exchange Commission (SEC). While government oversight is argued to be critical in avoiding economic calamities, regulations can cause friction in novel technology adoption owing to compliance costs, resource allocation, and slower time-to-market (Merton, 1995). Financial institutions can also be characteristically risk averse. Tasked with safeguarding customer funds and data, it is understandable that institutions can be hesitant to stir the waters by experimenting with emerging tools. Further demonstrating stagnation in innovation, many established banks have continued to operate legacy systems that have aged considerably since inception. Strikingly, even today, over 43% of banking systems and 95% of ATMs operate on computers running COBOL, a programming language developed in the 1950s (Arslanian & Fischer, 2019).

At the same time, the financial industry's objective to maximize profits and gain competitive advantages often propels it to become an innovation pioneer. In 2024, there were nearly 29,955 FinTech startups (up nearly 147% from 2018) spurred to leverage technology to generate revenue, provide new services, and disrupt traditional offerings (Alt et al., 2024). Incumbent institutions pour substantial capital in applying new innovations too. In 2023, global technology spending in banking grew at an average annual rate of 9% reaching approximately $650 billion (*Unlocking Value from Banking Technology Investment | McKinsey*, n.d.).

Machine learning is one critical tool adopted by these companies. A 2024 Congressional Research Service report notes “the financial industry’s adoption of artificial intelligence (AI) and machine learning (ML) is evolving as financial firms employ ever greater levels of technology and automation to deliver services (Tierno, n.d.). Building on this context, this paper examines the influence of machine learning within finance through its historical progression and current applications using specific models and techniques. Additionally, it discusses challenges machine learning faces within the financial industry.

Historical Context

Today, machine learning has become synonymous with a modern catchphrase for innovation, having “morphed into an all-purpose marketing buzzword that vendors slap onto products to make them sound more sophisticated” (Kobielus, 2018). However, far from an overnight miracle, it is necessary to appreciate that machine learning concepts are grounded in longstanding mathematical principles. Statistical techniques such as regression analysis and maximum likelihood estimation trace back to as early as the 19th century, and even those are predicated on mathematical foundations from long before. For example, the method of least squares, a form of regression analysis, was first introduced by Adrien-Mariet Legendre back in 1805 with the publication of “Nouvelles méthodes pour la détermination des orbites des comètes [New methods for determining the orbits of comets]”, which provided an algebraic procedure for finding best-fitting lines (Legendre, 1805/2018). In the early 20th century, Ronald A. Fisher formalized the maximum likelihood estimation approach, necessary to statistically infer model parameters (Aldrich, 1997). Within the field of finance, French mathematician Louis Bachelier introduced the idea of modeling stock prices as a random walk in his 1900 thesis "Théorie de la speculation [The Theory of Speculation]" (Bachelier, 1900). This stochastic process paved the way for modeling uncertainty in financial markets, becoming a foundation for modern quantitative financial techniques. Though these types of mathematical techniques are the foundation for analytical modeling, it would not be until the advent of digital computers in the mid 19th century that these mathematical principles could be utilized at scale in finance.

Early Computational Finance

Theoretical models in finance began to transform into practical tools in the mid-20th century. The emergence of the digital computer allowed computationally intensive calculations to be performed on large datasets resulting in new techniques. One such application was the Monte Carlo Simulation developed by Stanislaw Ulam and John von Neumann in the 1940s for solving problems involving uncertainty with random sampling. Monte Carlo has been used in finance to model probabilistic outcome such as future price movement and portfolio risk exposure (Metropolis & Ulam, 1949; Hammersley & Handscomb, 1964).

Later frameworks such as Markowitz's Portfolio Theory and the Black-Scholes Model would also carry the field towards employing machine learning tools. Portfolio theory, introduced by Harry Markowitz in 1952, formalized a risk-return optimization approach by calculating expected returns, variances, and covariances of assets (Markowitz, 1952, 1959). This financial technique paved the way for machine learning models to predict returns and classify asset behaviors. Similarly, the Black-Scholes Model, which was introduced in 1973, provided a mathematical formula for option pricing using variables such as volatility and asset maturity (Black & Scholes, 1973). These implementations, often computationally challenging and expensive, would necessitate new mathematical techniques.

Modern Machine Learning in Finance

In the 1990s and early 2000s new techniques such as Support Vector Machine (SVM) emerged and gained influence in the financial industry because of its ability to solve problems involving high-dimensional data. Introduced by Vladimir Vapnik and Corinna Cortes, SVM is a supervised machine learning algorithm used to maximize the margin between classes in datasets by identifying an optimal hyperplane (Cortes & Vapnik, 1995). SVMs became relevant for applications such as algorithmic trading, and could be applied to predict market trends, for example, by evaluating results of text mining (Xie & Jiang, 2019). Predictive financial modeling became more robust using Gradient Boosting Machines (GBMs). Developed by Jerome Friedman in the 2000s, GBMs could combine multiple weak predictive models to create stronger ones through a method called boosting (Friedman, 2001). This allowed for more sophisticated financial models that used computational resources more efficiently.

Deep Learning emerged in the 2010s and found practical applications in finance. Algorithms like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) allow traders to analyze large amounts of historical market data to find patterns, influencing trading decisions. A study published in *Applied Intelligence* highlights this advantage: "…we use two NN-based models (CNN and RNN) to deal with spatial and temporal information in order to refine the portfolio strategy… results indicate that the proposed [PMS] scheme in conjunction with the reward function of Sharpe ratio can outperform that with the reward function of trading returns, resulting in a 39% improvement in returns and a 13.7% reduction in drawdown." (Wu et al., 2021). Leveraging historical data through deep learning techniques is also practical in financial fraud detection by uncovering and interpreting patterns more effectively than traditional rule-based programming. A study published in *SN Computer Science* found that "traditional fraud detection methods frequently depend on present rules and patterns that skilled scammers can easily circumvent. Machine learning and deep learning algorithms have surfaced as promising methods for detecting fraud in order to handle this problem" (Gandhar et al., 2024).

BloombergGPT, a large language model (LLM) tailored for the financial industry was introduced in 2023 by Bloomberg, one of the world's largest financial and media companies. Boasting over 50-billion parameters and trained on vast amounts of financial data, it has been deployed by Bloomberg and clients for uses such as risk assessment, real-time event detection alerts, and market trend analysis (Wu et al., 2023). Investments in new machine learning infrastructure such as BloombergGPT demonstrate the financial industry's keen interest in the field. It is evident that machine learning will continue to influence the financial industry as more powerful computing resources become available and data continues to grow exponentially.

Evaluation of Current Applications

Actionable financial applications of machine learning are becoming increasingly apparent, including in high-frequency trading, credit risk assessment, and fraud prevention.

Support Vector Machine for High-Frequency Trading

Although high-frequency trading (HFT) lacks a legal or formal definition, the Congressional Research Service (CRS) defines HFT as "the trading in financial instruments, such as securities and derivatives, transacted through supercomputers executing trading within microseconds or milliseconds (or, in the technical jargon, with *extremely low latency*)" (Miller, n.d.). In short, HFT leverages a combination of algorithms, high speed networks, and modern computer infrastructure to perform market transactions at high speed. As a result, firms can profit from short-term market inefficiencies by engaging in financial strategies such as market making by placing buy and sell positions to profit from the bid-ask spread, and statistical arbitrage, which exploits pricing differences between related securities (Cartea, 2015).

SVMs are one tool used to find these opportunities. As a supervised learning model, SVMs leverage historical data to allow prediction of stock movements. For example, research published in 2023 proposed an SVM-based approach for mid-price (average of bid and ask prices) prediction in HFT environments by identifying three modeling strategies: 1) recovering information lost during data thinning, 2) sampling combined with ensemble learning, and 3) applying functional principal component analysis (FPCA) to compress and incorporate long-term price trends (Zhang et al., 2023).

First, the researchers recovered lost information in their dataset by introducing new windowed variables, such as mean, variance, and volatility of mid-price changes. By recovering valuable information, the team achieved an F1 score (a measure of predictive performance) increase of up to 0.056. Next, they combined random sampling with ensemble learning, training multiple SVM models on random subsets of data and aggregated their predictions through a voting mechanism to minimize individual model biases. This second strategy led to the largest boost in accuracy, with an F1 score improvement of up to 0.087. Finally, the team applied FPCA to compress historical price trajectories into a smaller set of predictors, allowing the model to integrate long-term movements with short-term variations. This third strategy led to an F1 score improvement of 0.016. Results of these strategies demonstrate the effectiveness of complementing machine learning techniques like SVM with other statistical tools such as FPCA to improve predictive accuracy in HFT. This allows traders to make more informed decisions in time-sensitive trading environments.

Random Forest for Credit Risk Assessment

Credit Risk Assessment is another pivotal domain in finance where machine learning has been applied. Its importance is highlighted by the 2007-2008 financial crisis where, among other factors, poor risk management and mispriced mortgage-backed securities led to defaults and economic instability (Rodini, 2023). Accurate credit risk assessment through methods like credit scoring is critical for financial institutions because it minimizes default rates, ensuring financial stability. Random forest, an ensemble machine learning method, is one effective machine learning tool that addresses these challenges.

Random forests are highly effective at enhancing predictive accuracy and controlling overfitting by constructing multiple decision trees during training, providing either classification or regression outputs. These machine learning methods are particularly powerful for credit risk assessment. A study conducted in Lessmann, et al. (2015) benchmarked 41 machine learning algorithms on real credit scoring datasets and found that random forest consistently outperformed (Lessmann et al., 2015). Random Forest was compared to other traditional techniques including logistic regression and decisions trees using metrics such as area under receiver operating characteristics curve (AUC-ROC), which plots the true positive rate. In the study, random forest attained AUC-ROC of 0.93, while logistic regression and decision trees achieved AUC-ROC scores of 0.85 and 0.88, respectively. Random forest’s higher AUC-ROC score on test data highlights its effectiveness in distinguishing between borrowers with high and low risk profiles. Further, because random forest is an ensemble method that aggregates outputs of multiple decision trees, random forest is also particularly resilient to overfitting. This characteristic along with its performance makes it a favorable and practical tool for financial institutions in optimizing lending decisions.

Deep Learning to Prevent Fraud

Innovative computational countermeasures are continuously developed and refined because fraudulent activities such as credit card fraud, identity theft, and money laundering can lead to irreversible reputational damage and financial loss. Traditional rule-based systems are often ill-equipped to evolving fraud due to their rule-based nature which are static and lacking contextual awareness, and thus have provided a new opportunity for machine learning techniques (Fraud detection in finance, IBM, n.d.).

Recurrent Neural Networks (RNNs), and particularly Long Short-Term Memory (LSTM) networks are among the machine learning techniques used to detect fraudulent credit card transactions. For example, a study conducted by Jurgovsky et al. (2018) demonstrated that an LSTM-based model is effective at detecting anomalies in user and behavior and transactional data (Jurgovsky et al., 2018). By learning typical transactional flow, including the order and timing of transactions across users, LSTM could flag potentially abnormal activities. This LSTM-based approach achieved an F1 score improvement of 10% compared to traditional methods such as Random Forest classifier.

Due to its effectiveness, a variety of financial institutions actively use LSTM for fraud detection. Partnering with Nvidia, American Express has used LSTM networks to detect fraud, leading to accuracy enhancements of up to 6% in certain segments, operating within latency requirements of under two milliseconds (NVIDIA, n.d.). At scale, these gains in accuracy enable financial institutions to confidently respond to fraud.

Limitations and Challenges

While machine learning remains a powerful tool in finance, its adoption is hindered by several challenges. Robust models are contingent on the availability and quality of their underlying data. Additionally, computational complexity and the cost to process immense amounts of data is a substantial barrier for both firms and researchers. Addressing these challenges is critical to fully leverage machine learning applications in finance.

Data Quality and Availability

The effectiveness of machine learning models is dependent on data integrity. Research indicates that "incomplete, erroneous or inappropriate training data can lead to unreliable models that produce ultimately poor decisions" (Budach et al., 2022). This requirement is particularly relevant in finance, where inaccurate or incomplete data can lead to erroneous prediction that may result in financial loss or damage to institutional reputation. Therefore, as noted by Wasserbacher & Spindler, "particular care must be taken to avoid the pitfalls of using [machine learning models] for planning and resource allocation" (Wasserbacher & Spindler, 2021).

Datasets often contain missing or incorrect values. This can occur for a variety of reasons, including human error, system glitches, or delayed reports. When a dataset containing missing or incorrect values is used as an input for machine learning models, there may be biases which do not reflect real-world scenarios. In one credit rating study conducted by Tsai and Chen, they reported many financial datasets contain missing values, affecting the performance of traditional machine learning models including decision trees, artificial neural networks (ANNs), support vector machines (SVMs), and logistic regression (Tsai & Chen, 2020).

Noisy data is another challenge for machine learning models. Financial markets are influenced by unpredictable systematic and unsystematic changes, such as global events, regulatory changes, shifts in consumer sentiment, and asset price fluctuations. Noise obscures underlying patterns and trends machine learning models are designed to capture. As one study puts it, “Noise is a fundamental problem in machine learning theory with huge effects in the application of Machine Learning (ML) methods, due to real world data tendency to be noisy.” (Ibias et al., 2024)

Financial datasets often contain outliers caused by factors like market shocks and irregular transactions. Even a small quantity of outliers poses challenges to models, disproportionately skewing predictions and reducing accuracy. One study on maximum likelihood states this concern, “Outliers or noisy labels in training data result in degraded performances as well as incorrect estimation of uncertainty.” (Nair et al., 2022)

Availability of financial data is another challenge. Financial data is often proprietary, which limits access to smaller firms, researchers and individuals (Einav & Levin, 2014). Even financial datasets which are available are often costly to acquire. For example, Bloomberg Terminal, a leading financial data tool, can cost as much as $32,000 per year for a single user (*What Is a Bloomberg Terminal (BT),* n.d.). A report by Neudata shows that some institutions often pay over $500,000 for yearly subscriptions to some financial data sets (*How Much Are Managers Paying for Data*, 2020). This poses a particular challenge for academic researchers budgeting research expenditure.

Data privacy and confidentiality can also decrease access to data. For example, in the European Union, regulations like the General Data Protection Regulation (GDPR) impose strict guidelines on the use of personal financial data. Facing non-compliance penalties of up to 20 million euros or 4% of annual global turnover, many firms are understandably cautious to share data, affecting its availability for machine learning purposes (Fines / Penalties, n.d.; Voigt & Bussche, 2017).

To mitigate data quality and accessibility challenges, machine learning researchers employ a variety of tools. Procedures such as data preprocessing, cleaning, and sometimes augmentation are standard practice during research (Mastering Data Cleaning & Data Preprocessing, n.d.). Missing data is often imputed by substituting values with measures of central tendency (mean, median, and mode), or frequency (in the case of categorical data). In some cases, instances with missing values may be excluded from datasets. Clustering and statistical methods, such as Z-score and Interquartile Range (IQR), are effective at identifying outlier values, which can also be excluded from datasets. By addressing data-related challenges using these methodologies, researchers can improve model performance and unlock the potential of machine learning in finance.

Computational Complexity and Cost

The computational complexity of machine learning algorithms applied to financial datasets increases exponentially with the number of features, a phenomenon known as the "curse of dimensionality". This challenge affects both computing resource requirements and model results. As one study by Pes and Lai highlights, "Besides posing severe requirements in terms of computational resources, high dimensionality may have a negative impact on the predictive performance of machine learning algorithms" (Pes & Lai, 2021). In portfolio management, for instance, it is common for the number of features to be much greater than the number of individual stocks, further increasing computational cost (Bobbitt, 2021).

High-frequency financial data, such as trading data, can also pose a computationally intensive challenge for machine learning algorithms. Financial data generated rapidly and in large volume is costly both in terms of computational power and memory for machine learning algorithms such as neural networks. Additionally, in finance this data may be time-sensitive and require real-time processing and analysis. For example, one study by Dixon found significant demands and costs associated when using RNNs to classify sequences in limit order books in real-time high-frequency trading settings (Dixon, 2017).

Another computationally complex challenge in the industry arises from the interconnectedness and the vast scale of financial networks. Financial institutions are linked together through a variety of financial instruments and obligations, forming networks difficult to map. Battiston et al. (2016) note that modeling these systems requires significant resources due to the high dimensionality and interdependency of variables. As the complexity of financial systems increase so does the computational cost of modeling the data. This further illustrates that even when an opportunity to leverage machine learning arises within the industry, it may not always be straightforward implementing it.

Adversarial Behavior and Market Adaptation

A foundational concept in financial market theory, particularly in game theory, is that the market is composed of intelligent agents who form strategies in response to anticipated actions of others. Implementation of machine learning models may become less effective as market participants adapt, leading to diminishing benefit over time. It has been found that the profitability of many technical trading strategies tends to decline as market participants adapt and market efficiency improves particularly within small-cap stocks (Cakici et al., 2023).

Adversarial attacks also pose a challenge to machine learning models. A study conducted by Goldstein, et al. (2021) found that adversarial attacks can be used to manipulate behavior of machine learning systems, especially in high-frequency trading which can lead to significant financial losses. New methods to deal with these risks are evolving, such as adversarial reinforcement learning (ARL) which “can be used to produce market marking agents that are robust to adversarial and adaptively chosen market conditions” (Spooner & Savani, 2020). These adversarial behaviors demonstrate the continued need for model evaluation and adaptation, along with financial strategies that anticipate market changes.

Conclusion

Machine Learning is continuing to emerge as a transformative force in finance, redefining traditional approaches found throughout the industry. The McKinsey Global Institute (MGI) highlights that generative AI alone, through increased productivity, will contribute between $200 billion and $340 billion in value annually across the global banking sector (*The Future of AI in Banking | McKinsey*, n.d.). This enormous economic potential is underscored by consensus among industry leaders, evidenced by a recent Deloitte survey which found that 86% of financial services executives say that AI “will be very or critically important to their business’s success in the next two years” (Deloitte, 2023)”. This profound and generational shift is aptly described by Financial Times which has stated, “AI in finance is like moving from typewriters to word processors” (Murray, 2024). As the financial industry continued to adopt machine learning techniques, it will be crucial to be mindful of machine learning’s limitations and ethical considerations. With thoughtful implementation, machine learning will continue to deliver lasting value and solve some of the financial sector’s greatest challenges.

References

Aldrich, J. (1997). R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science, 12*(3). https://doi.org/10.1214/ss/1030037906

Alt, R., Fridgen, G., & Chang, Y. (2024). The future of fintech—Towards ubiquitous financial services. *Electronic Markets, 34*(1), 3. https://doi.org/10.1007/s12525-023-00687-8

Arslanian, H., & Fischer, F. (2019). Incumbent financial institutions and their response to fintechs. In H. Arslanian & F. Fischer (Eds.), *The Future of Finance: The Impact of FinTech, AI, and Crypto on Financial Services* (pp. 57–67). Springer International Publishing. https://doi.org/10.1007/978-3-030-14533-0_4

Bachelier, L. (1900/2011). *The theory of speculation*. In P. H. Cootner (Ed.), *The random character of stock market prices* (pp. 17–78). Princeton University Press. (Original work published 1900)

Battiston, S., Caldarelli, G., May, R. M., Roukny, T., & Stiglitz, J. E. (2016). The price of complexity in financial networks. *Proceedings of the National Academy of Sciences of the United States of America*, 113(36), 10031–10036. https://doi.org/10.1073/pnas.1521573113

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637–654.

Bobbitt, Z. (2021, February 10). *What is High Dimensional Data? (Definition & Examples)*. Statology. https://www.statology.org/high-dimensional-data/

Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2022). The effects of data quality on machine learning performance. *arXiv*. https://arxiv.org/abs/2207.14529

Cakici, N., Fieberg, C., Metko, D., & Zaremba, A. (2023). *Predicting Returns with Machine Learning Across Horizons, Firms Size, and Time* (SSRN Scholarly Paper No. 4552395). Social Science Research Network. https://papers.ssrn.com/abstract=4552395

Cartea, Á. (with Jaimungal, S., & Penalva, J.). (2015). *Algorithmic and High-Frequency Trading* (1st ed). Cambridge University Press.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Deloitte. (2023). *AI transforming the future of banking.* Retrieved from https://www2.deloitte.com

Dixon, M. F. (2017). *Sequence Classification of the Limit Order Book using Recurrent Neural Networks* (No. arXiv:1707.05642). arXiv. https://doi.org/10.48550/arXiv.1707.05642

Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science (New York, N.Y.)*, *346*(6210), 1243089. https://doi.org/10.1126/science.1243089

Fines / Penalties. (n.d.). *General Data Protection Regulation (GDPR)*. Retrieved November 24, 2024, from https://gdpr-info.eu/issues/fines-penalties/

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gandhar, A., Gupta, K., Pandey, A. K., & Raj, D. (2024). Fraud detection using machine learning and deep learning. *SN Computer Science, 5*(5), 453. https://doi.org/10.1007/s42979-024-02772-x

Goldblum, M., Schwarzschild, A., Patel, A. B., & Goldstein, T. (2021). *Adversarial Attacks on Machine Learning Systems for High-Frequency Trading* (No. arXiv:2002.09565). arXiv. https://doi.org/10.48550/arXiv.2002.09565

Gupta, K., & Wang, Y. (2023). Redefining efficiency: Computational methods for financial models in Python. *International Journal of Computer Trends and Technology, 71*(10), 114–121. https://doi.org/10.14445/22312803/IJCTT-V71I10P113

Hammersley, J. M., & Handscomb, D. C. (1964). *Monte Carlo methods.* Methuen.

How Much Are Managers Paying for Data? (2020, July 9). Institutional Investor. https://www.institutionalinvestor.com/article/2bsx5x6m2pew8jzgavmyo/corner-office/how-much-are-managers-paying-for-data

Hu, L., Chen, J., Vaughan, J., Yang, H., Wang, K., Sudjianto, A., & Nair, V. N. (2020). Supervised machine learning techniques: An overview with applications to banking (No. arXiv:2008.04059). *arXiv.* https://doi.org/10.48550/arXiv.2008.04059

Ibias, A., Capala, K., Varma, V. R., Drozdz, A., & Sousa, J. (2024). *Improving Noise Robustness through Abstractions and its Impact on Machine Learning* (No. arXiv:2406.08428). arXiv. https://doi.org/10.48550/arXiv.2406.08428

IBM. (n.d.). *Fraud detection in finance: Safeguarding assets and ensuring compliance*. Retrieved from https://www.ibm.com/topics/fraud-detection

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234–245. https://doi.org/10.1016/j.eswa.2018.01.037

Kobielus, J. (2018, March 20). *Machine learning is becoming a meaningless buzzword*. InfoWorld. https://www.infoworld.com/article/3264290/machine-learning-is-becoming-a-meaningless-buzzword.html

Legendre, A.-M. (2018). *Nouvelles méthodes pour la détermination des orbites des comètes.* Wentworth Press. (Original work published 1805)

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

Liu, Q., Chan, K.-C., & Chimhundu, R. (2024). Fintech research: Systematic mapping, classification, and future directions. *Financial Innovation, 10*(1), 24. https://doi.org/10.1186/s40854-023-00524-z

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments.* John Wiley & Sons.

Mastering Data Cleaning & Data Preprocessing. (n.d.). Retrieved November 24, 2024, from https://encord.com/blog/data-cleaning-data-preprocessing/

Merton, R. C. (1995). Financial innovation and the management and regulation of financial institutions. *Journal of Banking & Finance*, *19*(3), 461–481. https://doi.org/10.1016/0378-4266(94)00133-N

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association, 44*(247), 335–341.

Murray, S. (2024, June 16). AI in finance is like 'moving from typewriters to word processors.' *Financial Times*. https://www.ft.com/content/c35ce925-d7b3-4920-a431-c4ca1aa33503

Nair, D. S., Hochgeschwender, N., & Olivares-Mendez, M. A. (2022). *Maximum Likelihood Uncertainty Estimation: Robustness to Outliers* (No. arXiv:2202.03870). arXiv. https://doi.org/10.48550/arXiv.2202.03870

NVIDIA. (n.d.). *American Express prevents fraud and foils cybercrime with NVIDIA AI solutions*. Retrieved from https://www.nvidia.com/en-us/case-studies/american-express-prevents-fraud-and-foils-cybercrime-with-nvidia-ai-solutions/

Pes, B., & Lai, G. (2021). Cost-sensitive learning strategies for high-dimensional and imbalanced data: A comparative study. *PeerJ Computer Science*, *7*, e832. https://doi.org/10.7717/peerj-cs.832

Rodini, L. (2023, March 28). *What Is Credit Risk? Definition, Importance & Examples*. TheStreet. https://www.thestreet.com/dictionary/credit-risk

Smith, D. E. (1959). *A source book in mathematics* (Vol. 2, pp. 576–579). Dover Publications. (Original work published 1929 by McGraw-Hill)

Spooner, T., & Savani, R. (2020). *Robust Market Making via Adversarial Reinforcement Learning* (No. arXiv:2003.01820). arXiv. https://doi.org/10.48550/arXiv.2003.01820

Tierno, P. (n.d.). *Artificial intelligence and machine learning in financial services.*

The future of AI in banking | McKinsey. (n.d.). Retrieved November 24, 2024, from https://www.mckinsey.com/industries/financial-services/our-insights/scaling-gen-ai-in-banking-choosing-the-best-operating-model

Tsai, C. F., & Chen, M. L. (2020). Credit rating by hybrid machine learning techniques. *Applied Soft Computing*, 10(2), 374–380.

Unlocking value from banking technology investment | McKinsey. (n.d.). Retrieved November 22, 2024, from https://www.mckinsey.com/industries/financial-services/our-insights/unlocking-value-from-technology-in-banking-an-investor-lens

Voigt, P., & Bussche, A. V. D. (2017). *The EU General Data Protection Regulation (GDPR)*. Springer International Publishing.

Wasserbacher, H., & Spindler, M. (2021). Machine learning for financial forecasting, planning and analysis: Recent developments and pitfalls. *Digital Finance*, 4, 63–88. https://doi.org/10.1007/s42521-021-00046-2

Wu, M.-E., Syu, J.-H., Lin, J. C.-W., & Ho, J.-M. (2021). Portfolio management system in equity market neutral using reinforcement learning. *Applied Intelligence, 51*(11), 8119–8131. https://doi.org/10.1007/s10489-021-02262-0

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). *BloombergGPT: A Large Language Model for Finance* (No. arXiv:2303.17564). arXiv.

Xie, Y., & Jiang, H. (2019). Stock market forecasting based on text mining technology: A support vector machine method (No. arXiv:1909.12789). *arXiv.* https://doi.org/10.48550/arXiv.1909.12789

Wasserbacher, H., & Spindler, M. (2021). Machine learning for financial forecasting, planning and analysis: Recent developments and pitfalls. *Digital Finance*, 4, 63–88. https://doi.org/10.1007/s42521-021-00046-2

What Is a Bloomberg Terminal (BT)? Functions, Costs, and Inexpensive Alternatives. (n.d.). Investopedia. Retrieved November 24, 2024, from https://www.investopedia.com/terms/b/bloomberg_terminal.asp

Zhang, X., Huang, Y., Xu, K., & Xing, L. (2023). Novel modelling strategies for high-frequency stock trading data. *Financial Innovation, 9*(1), Article 39. https://doi.org/10.1186/s40854-022-00431-9