Machine Learning in Finance

Huseyin Sinan Havus
Dept. of Computer Engineering
Bilkent University
Ankara, Turkiye
sinan.havus@bilkent.edu.tr

Mahssa Nassiri
Dept. of Electrical and Electronics Engineering
Bilkent University
Ankara, Turkiye
mahsa.nasiri@bilkent.edu.tr

***Abstract*—Machine learning has become an essential tool in modern finance, offering advanced techniques for data-driven decision making, risk management and predictive analytics. This paper explores the applications of machine learning in financial domains, including high-frequency trading, credit risk assessment, fraud detection, portfolio optimization, and macroeconomic forecasting. Various models such as supervised learning (random forests, gradient boosting), deep learning (LSTMs, CNNs), reinforcement learning (DQN) and unsupervised learning (clustering, anomaly detection) are leveraged to enhance financial predictions and automate complex tasks. By analyzing large datasets and identifying hidden patterns, machine learning improves accuracy and efficiency in financial markets. This study highlights the advantages, challenges, and future potential of machine learning in shaping the financial industry.**

***Index Terms*—Machine Learning, High-Frequency Trading, Credit Risk Assessment, Fraud Detection, Portfolio Optimization, Macroeconomic Forecasting**

I. Introduction

Machine learning has fundamentally transformed the financial sector by enabling data-driven decision-making, automating complex processes, and uncovering insights from vast, multidimensional datasets. Traditional financial models, often constrained by static assumptions and linear relationships, struggle to capture the nonlinear dynamics and hidden dependencies inherent in modern markets [4].

In contrast, machine learning techniques spanning from supervised learning, deep learning, reinforcement learning and to unsupervised methods excel at processing high frequency trading data, optimizing portfolios and detecting fraudulent transactions with unprecedented accuracy. These models leverage heterogeneous data sources, from macroeconomic indicators to real-time market sentiment to enhance predictive capabilities in credit risk assessment, stock price forecasting and macroeconomic trend analysis. The adoption of machine learning in finance is not without challenges. Issues such as model interpretability, data quality inconsistencies, and regulatory compliance require careful consideration, particularly in high-stakes domains like algorithmic trading and risk management [4].

Furthermore, emerging paradigms like decentralized machine learning governance and quantum machine learning hint at future frontiers, promising enhanced computational efficiency and ethical frameworks for financial applications [12]. This survey systematically examines these advancements, evaluating their technical foundations, practical implementation, and transformative potential while addressing critical barriers to adoption.

II. Machine Learning Applications in Finance

A. Exchange Rate Prediction

Exchange rate prediction plays a vital role in finance, impacting global trade, investment strategies, and risk assessment. Machine learning techniques have become essential for forecasting currency fluctuations by analyzing past exchange rate trends, macroeconomic factors, interest rates and market sentiment.

Shen et al. [1] enhance the DBN model by incorporating RBM, demonstrating that it outperforms the random walk algorithm, auto-regressive-moving-average and FNN with lower error rates. Similarly, Zheng et al. [2] evaluate DBN's performance and conclude that it provides more accurate exchange rate estimates than the FNN model, highlighting the significant impact of a small number of layer nodes on DBN's effectiveness. Many researchers of the financial filed suggest that hybrid models can enhance predictive accuracy.

Ravi et al. [3] propose a hybrid approach combining MLP (FNN), chaos theory, and multi-objective evolutionary algorithms, achieving a remarkably low mean squared error of 2.16E-08. Further studies indicate that CNN, leveraging the deep hierarchical structure of DNN, achieves even higher classification accuracy in predicting exchange rate movements.

B. Stock Trading

Machine learning has revolutionized stock trading by enabling data-driven decision making. Traders and investors leverage machine learning models to analyze vast financial datasets, identify patterns and forecast stock price movements. Techniques such as deep learning, reinforcement learning, and sentiment analysis help enhance trading strategies, optimize portfolios and manage risks [4].

Many studies use FNN for stock trading. Sezer et al. [5] combine GA with MLP, while Chen et al. [6] show that a double-layer NN outperforms ARMA-GARCH and single-layer NNs. RNN-based models have shown potential in stock trading. Some researchers have integrated fuzzy learning into RNN [7], claiming optimal results, while others have focused on LSTM, demonstrating improved trading performance [8]. However, profitability fluctuated in later years. To enhance accuracy, advanced techniques like wavelet transforms and autoencoders have been incorporated, further refining LSTM-based predictions [8].

C. Macroeconomic Prediction

Machine learning has emerged as a powerful tool in macroeconomic prediction, offering significant improvements over traditional econometric models. Studies have shown that machine learning algorithms can effectively predict GDP by incorporating vital macroeconomic variables such as energy prices, unemployment rates and exchange rates [9]. Techniques like Support Vector Regression, Random Forest and Gradient Boosting Machines have been employed with SVR often performing best in GDP prediction after hyperparameter tuning [9]. Additionally, machine learning is used in predicting other economic indicators, such as stock prices and bond yields, by integrating economic indicators and market sentiment [10].

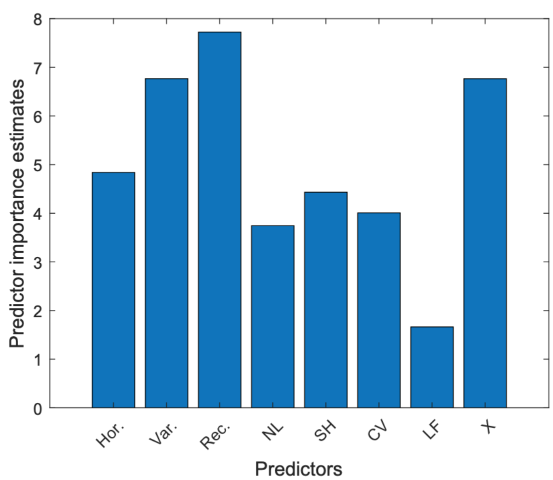

Fig. 1. This figure presents predictive importance estimates. for a random forest predictor. NL, SH, CV and LF stand for nonlinearity, shrinkage, cross-validation and loss function features respectively. [24]

The application of machine learning in macroeconomic prediction extends beyond GDP and stock markets. It also involves data fusion methods to improve the accuracy of macroeconomic data integration [11]. Furthermore, because of machine learning models ability to capture complex temporal dependencies, making them superior in forecasting tasks compared to traditional methods [10]. The use of machine learning in macroeconomic prediction not only enhances forecasting accuracy but also provides valuable insights into economic trends, helping policymakers and investors make informed decisions.

D. Credit and Risk Scoring

Machine learning has also become integral to modern credit and risk scoring, offering enhanced accuracy and efficiency compared to traditional statistical methods. Current approaches leverage diverse algorithms, including Support Vector Machines, Random Forest, LightGBM and hybrid architectures combining multiple models [4].

For instance, SVM demonstrates strong performance in credit classification, achieving an F-measure of 0.86 when integrated with feature selection and Synthetic Minority Oversampling Techniques [13]. Hybrid models, such as those merging Random Forest, Multi-Layer Perceptron and LightGBM, show improved classification accuracy on datasets like the South German Credit dataset (45,211 records), validated through metrics like ROC curves and sensitivity analysis [14].

E. Portfolio Managment

Machine learning models also emerged as a pivotal tool in portfolio management, offering significant advancements in optimizing returns while minimizing risk. Recent academic research highlights the potential of machine learning algorithms in this domain. For instance, a study on the impact of machine learning on portfolio optimization and risk management in sustainable investing found that models like Lasso regression can effectively enhance portfolio performance by leveraging ESG scores from major companies [15].

Also some projects have utilized historical financial data from prominent stock exchanges to train models that can provide insights into future market trends [16]. These models often combine sentiment analysis with advanced neural networks, such as long short-term memory (LSTM) networks, to improve predictive accuracy. The integration of machine learning in portfolio management is seen as essential for handling the computational complexity involved in selecting optimal asset combinations.

The application of machine learning in financial markets is not limited to portfolio optimization alone. It also extends to predicting short-term market fluctuations with positive outcomes. However, challenges remain, such as the non-normal distribution of historical returns, which necessitates the use of more sophisticated models [16].

F. High-Frequency Trading

High-frequency trading is a form of algorithmic trading in finance that involves executing trades at extremely high speeds, with rapid turnover rates and a high ratio of orders to completed trades, utilizing real-time financial data. High-frequency trading has seen significant advancements through machine learning innovations, particularly in reinforcement learning and deep neural architectures [18].

The Trading Deep Q-Network algorithm, adapted from DQN, has emerged as a prominent reinforcement learning approach for optimizing HFT strategies. It maximizes the Sharpe ratio by training agents on artificial market trajectories derived from limited historical data, achieving robust performance across diverse stock markets. Further advancements integrate uncertainty estimation into RL models, enabling dynamic risk capital allocation. For Eurodollar futures trading, deep learning models incorporating prediction uncertainties improved Sharpe ratios by 18.4% compared to uncertainty-agnostic strategies [17].

Hybrid models combining CNNs and LSTMs have demonstrated superior performance in price prediction, achieving 71.4% accuracy ($\pm 2.3\%$) and annual returns of 18.4% in backtesting [17]. These architectures process high-frequency limit order book data to forecast short-term price movements, with CNNs excelling at spatial feature extraction and LSTMs capturing temporal dependencies. Comparative studies show CNNs outperform traditional ML models like logistic regression and SVMs in volatility-rich environments [17].

While deep learning and hybrid models are useful for feature extraction and synthetic data generation, they are primarily leveraged for strategy development rather than direct trading execution. Bayesian models, such as Kalman Filters, are effective for probabilistic inference but are typically used as complementary tools rather than primary trading models [19].

G. Fraud Detection

Financial fraud poses significant risks to businesses, financial institutions, and consumers, leading to substantial economic losses and eroding trust in financial systems. Traditional fraud detection methods, such as rule-based systems and manual inspections, often fail to detect complex and evolving fraudulent activities. As a result, machine learning (ML) techniques have emerged as effective tools for detecting fraudulent transactions by analyzing large-scale datasets and identifying patterns that deviate from normal behavior. Various ML algorithms, including decision trees, random forests, neural networks, and anomaly detection methods, have demonstrated high accuracy in identifying fraudulent financial activities.

Recent studies highlight the effectiveness of different ML approaches in fraud detection. Decision Tree Classifiers, especially when combined with oversampling techniques such as SMOTE, have shown exceptional accuracy in detecting fraudulent transactions, reaching up to 99% accuracy in some cases [20]. Similarly, Random Forest models have been found to be particularly robust, achieving nearly perfect accuracy in distinguishing between legitimate and fraudulent banking transactions [21]. Additionally, deep learning techniques, such as convolutional and recurrent neural networks, have enhanced fraud detection capabilities by capturing complex transactional patterns [22].

Advanced methods, such as graph neural networks (GNNs), have also been explored to detect fraud within financial networks. These models excel in capturing intricate relationships between entities in financial transactions, outperforming conventional fraud detection techniques.

By harnessing AI-driven predictive analytics, financial institutions can significantly enhance fraud detection efficiency, minimize false positives, and proactively mitigate risks associated with fraudulent activities. However, the implementation of machine learning (ML)-based fraud detection systems faces key challenges, including data privacy concerns, model interpretability, and the constantly evolving nature of fraud tactics [23].

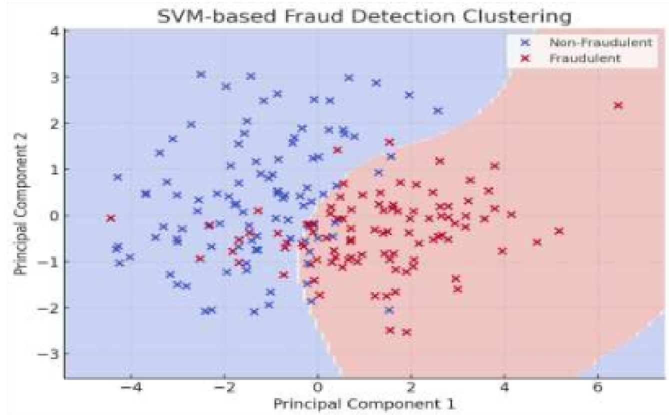

Fig. 2. Sentiment features and structured billing data.

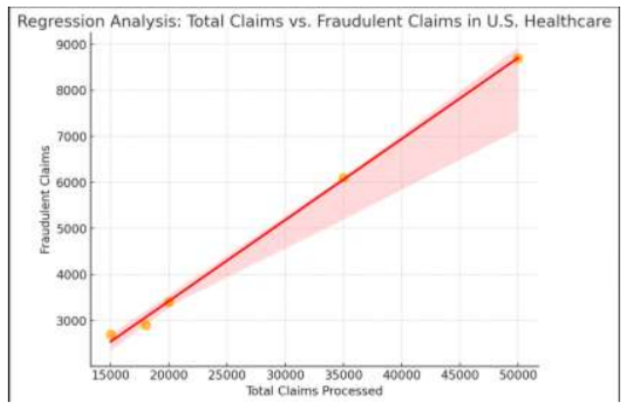

Fig. 3. Regression Analysis Chart

Deep learning techniques, such as convolutional and recurrent neural networks (CNNs and RNNs), have further strengthened fraud detection capabilities by capturing intricate transactional patterns and anomalies. The *AI-Driven Machine Learning for Fraud Detection and Risk Management in U.S. Healthcare Billing and Insurance* study explores the application of advanced ML techniques in identifying fraudulent healthcare claims. The research highlights the effectiveness of Support Vector Machines (SVMs) with sentiment-based features, which excel at distinguishing between fraudulent and legitimate transactions.

This capability is visually demonstrated in Figure 1, which presents an SVM-based Fraud Detection Clustering model. The chart effectively separates fraudulent claims (red) from non-fraudulent claims (blue) based on sentiment features and structured billing data, showcasing the model's ability to establish a clear decision boundary.

Furthermore, Figure 2 illustrates a Regression Analysis Chart, which reveals a strong positive correlation between the total number of claims processed and the number of fraudulent claims detected. This insight underscores how fraud risk increases with transaction volume, reinforcing the need for scalable AI-driven fraud detection solutions.

Overall, the findings emphasize the transformative role of AI in fraud prevention, enhancing financial security, improving detection accuracy, and optimizing operational efficiency in healthcare billing systems [22].

III. Challenges in Machine Learning for Finance

While machine learning offers transformative capabilities in finance, its implementation comes with significant challenges that must be addressed to ensure reliable and ethical outcomes. These challenges span technical, regulatory, and operational domains, posing complex barriers to the widespread adoption of machine learning techniques.

One of the most critical challenges in applying machine learning to finance is ensuring access to high-quality and comprehensive datasets. Financial data is often fragmented, noisy, and inconsistent due to differences in data collection methodologies and market conditions. Moreover, acquiring proprietary financial data can be expensive and legally restrictive, limiting model training and evaluation. Inconsistent data quality can lead to biased or inaccurate models, reducing their effectiveness in real-world applications.

Another significant challenge is model interpretability and transparency. Machine learning models, particularly deep learning approaches, often operate as "black boxes," making it difficult to interpret and explain their predictions. In the financial sector, where decision accountability is crucial, the lack of interpretability can hinder trust in automated systems. This opacity raises concerns when making critical decisions like credit approvals, trading strategies, or risk assessments.

Financial markets are dynamic and subject to rapid changes driven by economic events and geopolitical factors. Machine learning models trained on historical data may overfit to past patterns, limiting their ability to adapt to new conditions. To address these challenges, robust validation techniques, regular model retraining, and ensemble approaches are essential for improving generalization and minimizing overfitting risks. Additionally, the computational complexity of advanced techniques like deep reinforcement learning poses significant challenges. These methods require substantial resources for both training and deployment, especially in real-time applications like high-frequency trading, where low-latency solutions and scalable infrastructure are crucial. Balancing computational efficiency with model accuracy remains a critical concern as data volumes continue to grow.

And lastly, machine learning systems in finance also face security and privacy risks. They are vulnerable to adversarial attacks, data breaches, and model inversion techniques that can expose sensitive information. Ensuring data privacy and maintaining robust security protocols is vital to safeguard customer trust and market integrity. The risk of model manipulation further underscores the need for continuous monitoring and secure frameworks. Addressing these multifaceted challenges requires a multidisciplinary approach that blends technical innovation, regulatory awareness, and ethical considerations. Collaboration between financial experts, data scientists, and policymakers is crucial to developing resilient and trustworthy machine learning systems in the financial sector.

IV. Conclusion and Discussion

Machine learning has revolutionized finance, significantly enhancing decision-making in exchange rate prediction, stock trading, credit risk assessment, portfolio management, and high-frequency trading. By analyzing vast amounts of data and uncovering hidden patterns, machine learning improves accuracy, efficiency and automation in financial markets.

Despite its advantages, challenges persist, including model interpretability, data quality inconsistencies, regulatory constraints, and ethical concerns. Deep learning models often operate as black boxes, making it difficult to explain predictions. Additionally, reliance on historical data can introduce biases and regulatory frameworks struggle to keep pace with rapid advancements. Addressing these issues is essential for the responsible deployment of AI-driven financial tools.

Future advancements in computing, machine learning and decentralized learning hold promise for improving financial applications. The integration of more transparent and robust models, combined with ethical considerations, will be crucial in fostering trust and regulatory acceptance. As machine learning continues to evolve, ongoing research and collaboration between financial experts and machine learning practitioners will be key to harnessing its full potential while mitigating risks. With responsible adoption, machine learning will play an increasingly vital role in shaping a more efficient, secure and intelligent financial ecosystem.

References

[1] Shen, F., Chao, J., Zhao, J. (2015). Forecasting exchange rate using deep belief networks and conjugate gradient method. Neurocomputing, 167, 243–253.

[2] Zheng, J., Fu, X., Zhang, G. (2017). Research on exchange rate forecasting based on a deep belief network. Neural Computing and Application, 31, 573–582.

[3] Ravi, V., Pradeepkumar, D., Deb, K. (2017). Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. Swarm and Evolutionary Computation, 36, 136–149.

[4] El Hajj, M.; Hammoud, J. Unveiling the Influence of Artificial Intelligence and Machine Learning on Financial Markets: A Comprehensive Analysis of AI Applications in Trading, Risk Management, and Financial Operations. J. Risk Financial Manag. 2023, 16, 434.

[5] Sezer, O. B., Ozbayoglu, M., Gogdu, E. (2017). A deep neural-network-based stock trading system based on evolutionary optimized technical analysis parameters. Procedia Computer Science, 114, 473–480.

[6] Hsu, P. Y., Chou, C., Huang, S. H., Chen, A. P. (2018). A market making quotation strategy based on dual deep learning agents for option pricing and bid-ask spread estimation. The proceeding of IEEE international conference on agents (pp. 99–104).

[7] Deng, Y., Bao, F., Kong, Y., Ren, Z., Dai, Q. (2017). Deep direct reinforcement learning for financial signal representation and trading. IEEE Transactions on Neural Networks and Learning Systems, 28(3), 653–664.

[8] Bao, W., Yue, J., Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long- short-term memory. PLoS One, 12(7), e0180944.

[9] Perera, Piyumi and Anil Fernando (2024). Impact of Energy Prices and Macroeconomic Variables on GDP Prediction UK: Machine Learning Approach. Journal of Business and Management Studies.

[10] Li, Z., Wang, B., Chen, Y. (2024). Incorporating economic indicators and market sentiment effect into US Treasury bond yield prediction with machine learning. Journal of Infrastructure, Policy and Development.

[11] Wang, H. (2024). Macroeconomic Data Fusion Method Based on Machine Learning. 2024 3rd International Conference for Advancement in Technology (ICONAT), 1-4.

[12] Alsagheer, D., Xu, L., Shi, W. (2023). Decentralized Machine Learning Governance: Overview, Opportunities, and Challenges. IEEE Access, 11, 96718-96732.

[13] Krishna, K.S., Kumar, B.D., Reddy, M.D., Saketh, C., Venugopalan, M. (2024). Credit Scoring for Loan Applicants Using Machine Learning Models. 2023 4th International Conference on Intelligent Technologies (CONIT), 1-5.

[14] Mendhe, V., Gote, A., Mahant, R. (2024). Comparative Analysis of Machine Learning Models for Credit Scoring: A Case Study on the South German Credit Dataset. International Journal of Advanced Research in Science, Communication and Technology.

[15] Masood, F. (2024). Assessing the impact of Machine Learning Algorithms on Portfolio Optimization and Risk Management in the Era of Sustainable Investing. International Journal of Emerging Multidisciplinaries: Computer Science Artificial Intelligence.

[16] Narale, P. (2024). Investment Portfolio Management System Using Machine Learning. International Journal for Research in Applied Science and Engineering Technology.

[17] Sazu, M.H. (2022). How Machine Learning Can Drive High Frequency Algorithmic Trading for Technology Stocks. International Journal of Data Science and Advanced Analytics.

[18] Gomber, Peter Arndt, Björn Lutat, Marco Uhle, Tim. (2011). High-Frequency Trading. SSRN Electronic Journal. 10.2139/

[19] Manveer Kaur Mangat, Erhard Reschenhofer, Thomas Stark, Christian Zwatz. High-Frequency Trading with Machine Learning Algorithms and Limit Order Book Data[J]. Data Science in Finance and Economics, 2022, 2(4): 437-463. doi: 10.3934/DSFE.2022022

[20] Owoade, A. A. (2025). Identifying Financial Fraud Transactions Using Decision Tree Classifier Algorithm. Dutse Journal of Pure and Applied Sciences, 11(1a), 338-352.

[21] Preciado Martínez, Pedro María, et al. "Comparative analysis of machine learning models for the detection of fraudulent banking transactions." Cogent Business Management 12.1 (2025): 2474209.

[22] Dey, Raktim, et al. "AI-driven machine learning for fraud detection and risk management in US healthcare billing and insurance." Journal of Computer Science and Technology Studies 7.1 (2025): 188-198.

[23] Cheng, Dawei, et al. "Graph neural networks for financial fraud detection: a review." Frontiers of Computer Science 19.9 (2025): 1-15.

[24] Goulet Coulombe, Philippe, et al. "How is Machine Learning Useful for Macroeconomic Forecasting?" Journal of Applied Econometrics 38.1 (2019): 1-15.