Machine Learning in Finance

Elavalagan Balasubramani, Baskaran Priyanka

Rajalakshmi School of Business

Chennai - Bangalore Highway,

Kuthambakkam Post, Kanchipuram - Chennai Rd, Chembarambakkam, Tamil Nadu 600124

ABSTRACT

Machine Learning (ML) is a component of artificial intelligence that processes large data sets to detect patterns, generate predictions and recommendations, and improve effectiveness over time. The use cases of machine learning in finance are likely to evolve with the finance sector investing in AI and AI adding value to the services. As we move into the AI-powered digital age, machine learning has become one of the most vital needs for the financial sector. The profound amount of data generation in finance is also proving to be a valuable training environment for AI. There are several uses of machine learning in finance like risk modelling, algo trading, Portfolio Management and Robo-Advisors, fraud detection, financial chatbots, underwriting process, risk management, asset pricing prediction, derivative pricing, sentimental analysis, trade settlement. In the days to come, the financial industry will show increasingly more reliance on machine learning and artificial intelligence-based emerging methods and models to leverage competitive advantages.

INTRODUCTION

The paradigm of machine learning and artificial intelligence has pervaded our everyday life in such a way that it is no longer an area for esoteric academics and scientists putting their effort to solve a challenging research problem. The evolution is quite natural rather than accidental. With the exponential growth in processing speed and with the emergence of smarter algorithms for solving complex and challenging problems, organizations have found it possible to harness a humongous volume of data in realizing solutions that have far-reaching business values.

Financial services, banking, and insurance remain one of the most significant sectors that has a very high potential in reaping the benefits of machine learning and artificial intelligence with the availability of rich data, innovative algorithms, and novel methods in its various applications. While the organizations have only skimmed the surface of the rapidly evolving areas such as deep neural networks and reinforcement learning, the possibility of applying these techniques in many applications vastly remains unexplored. Organizations are leveraging the benefits of innovative applications of machine learning in applications like customer segmentation for target marketing of their newly launched products, designing optimal portfolio strategies, detection, and prevention of money laundering and other illegal activities in the financial markets, smarter and effective risk management is credit, adherence to the regulatory frameworks in finance, accounts, and other operations, and so on. However, the full capability of machine learning and artificial intelligence still remains unexplored and unexploited. Leveraging such capabilities will be critical for organizations to achieve and maintain a long-term competitive edge.

While one of the major reasons for the slow adoption of AI/ML models and methods in financial applications is that the algorithms are not well known and there is an inevitable trust deficit in deploying them in critical and privacy-sensitive applications, the so-called “black-box” nature of such models and frameworks that analyses their internal operations in producing outputs and their validations also impede faster acceptance and deployment of such models in real-world applications.

REVIEW OF LITERATURE

Goodell JW, Kumar S, Lim WM, Pattnaik D (2021) using both co-citation and bibliometric-coupling analyses, infer the thematic structure of AI and ML research in finance for 1986–April 2021, also they highlight trends and research directions regarding AI and ML in finance research.

Dixon MF, Halperin I, Bilokon P (2020) covers multiple machine learning approaches with advanced technical exposition and is therefore especially suitable as an academic reference point, especially on Reinforcement Learning and also introduce a unified theory of financial decision making that combines supervised and reinforcement machine learning.

Rundo F, Trenta F, di Stallo AL, Battiato S (2019) proposes a review of some of the most significant works providing an exhaustive overview of recent machine learning (ML) techniques in the field of quantitative finance showing that these methods outperform traditional approaches and also presents comparative studies about the effectiveness of several ML-based systems.

Ghoddusi H, Creamer GG, Rafizadeh N (2019) reviewed a large body of literature on the energy economics/finance applications of various ML methods and the review identifies applications in areas such as predicting energy prices (e.g. crude oil, natural gas, and power), demand forecasting, risk management, trading strategies, data processing, and analyzing macro/energy trends.

Culkin R, Das SR (2017) survey how and why AI and deep learning can influence the field of Finance and also offer a brief introduction to neural networks and some detail on the various choices of hyper-parameters that make the model as accurate as possible.

MACHINE LEARNING AND ITS TYPES

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

IBM has a rich history with machine learning. One of its own, Arthur Samuel, is credited for coining the term, “machine learning” with his research (PDF, 481 KB) (link resides outside IBM) around the game of checkers. Robert Nealey, the self-proclaimed checkers master, played the game on an IBM 7094 computer in 1962, and he lost to the computer. Compared to what can be done today, this feat seems trivial, but it’s considered a major milestone in the field of artificial intelligence.

Over the last couple of decades, the technological advances in storage and processing power have enabled some innovative products based on machine learning, such as Netflix’s recommendation engine and self-driving cars.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

Machine learning algorithms are typically created using frameworks that accelerate solution development, such as TensorFlow and PyTorch.

The three types of machine learning, are supervised learning, unsupervised learning, and reinforcement learning.

Supervised

The main goal in *supervised learning* is to train a model from labeled data that allows us to make predictions about unseen or future data. Here, the term *supervised* refers to a set of samples where the desired output signals (labels) are already known. There are two types of supervised learning algorithms: classification and regression.

Classification

Classification is a subcategory of supervised learning in which the goal is to predict the categorical class labels of new instances based on past observations.

Regression

Regression is another subcategory of supervised learning used in the prediction of continuous outcomes. In regression, we are given a number of predictor (explanatory) variables and a continuous response variable (outcome or target), and we try to find a relationship between those variables that allows us to predict an outcome.

Unsupervised

Unsupervised learning is a type of machine learning used to draw inferences from datasets consisting of input data without labeled responses. There are two types of unsupervised learning: dimensionality reduction and clustering.

Dimensionality reduction

Dimensionality reduction is the process of reducing the number of features, or variables, in a dataset while preserving information and overall model performance. It is a common and powerful way to deal with datasets that have a large number of dimensions.

Clustering

Clustering is a subcategory of unsupervised learning techniques that allows us to discover hidden structures in data. The goal of clustering is to find a natural grouping in data so that items in the same cluster are more similar to each other than to those from different clusters.

Why consider machine learning in finance?

Despite the challenges, many financial companies already take advantage of this technology. The figure below shows that financial services' execs take machine learning very seriously, and they do it for a bunch of good reasons:

1. Reduced operational costs thanks to process automation.
2. Increased revenues thanks to better productivity and enhanced user experiences.
3. Better compliance and reinforced security.

MACHINE LEARNING APPLICATIONS IN FINANCE

With the increasing availability and declining cost for complex models executing on high-power computing devices exploiting the unlimited capacity of data storage, the financial industry is geared up to exploit the benefits of machine learning to leverage a competitive business edge. While some of the use cases have already found their applications in the real world, others will need to overcome some existing business and operational challenges before they are deployed. Some of the applications are mentioned below.

Risk Modeling

One of the major applications of AI/ML models and algorithms is in the extensive domain of risk modeling and management. While on one hand, the risk modeling credit and market is a critical application of machine learning, on the other hand, a non-finance application such as operational risk management, compliance, and fraud management is also quite important. The majority of the classification approaches and modeling techniques in machine learning such as binary logistic regression, multinomial logistic regression, linear and quadratic discriminant analysis, and decision trees, etc., are the foundational building blocks of applied modeling in the real world. However, in data science applications, the availability of data and its richness play a pivotal role. Hence, in data-rich applications such as credit risk modeling and scoring, designing mortgage schemes, the AI/ML models have already made substantial inroads in comparison to scenarios such as low default credit portfolios for well-known parties that lack the availability of data. Fraud analytics remains another intensive area of AI/ML applications in the non-financial domain.

Algorithmic Trading

Algorithmic trading (or simply algo trading) is the use of algorithms to conduct trades autonomously. With origins going back to the 1970s, algorithmic trading (sometimes called Automated Trading Systems, which is arguably a more accurate description) involves the use of automated preprogramed trading instructions to make extremely fast, objective trading decisions.

Machine learning stands to push algorithmic trading to new levels. Not only can more advanced strategies be employed and adapted in real time, but machine learning-based techniques can offer even more avenues for gaining special insight into market movements. Most hedge funds and financial institutions do not openly disclose their machine learning-based approaches to trading (for good reason), but machine learning is playing an increasingly important role in calibrating trading decisions in real time.

Portfolio Management and Robo-Advisors

Asset and wealth management firms are exploring potential artificial intelligence (AI) solutions for improving their investment decisions and making use of their troves of historical data.

One example of this is the use of *robo-advisors*, algorithms built to calibrate a financial portfolio to the goals and risk tolerance of the user. Additionally, they provide automated financial guidance and service to end investors and clients.

A user enters their financial goals (e.g., to retire at age 65 with $250,000 in savings), age, income, and current financial assets. The advisor (the *allocator*) then spreads investments across asset classes and financial instruments in order to reach the user goals.

The system then calibrates to changes in the users goals and real-time changes in the market, aiming always to find the best fit for the users original goals. Robo-advisors have gained significant traction among consumers who do not need a human advisor to feel comfortable investing.

Fraud Detection

Fraud is a massive problem for financial institutions and one of the foremost reasons to leverage machine learning in finance.

There is currently a significant data security risk due to high computing power, frequent internet use, and an increasing amount of company data being stored online. While previous financial fraud detection systems depended heavily on complex and robust sets of rules, modern fraud detection goes beyond following a checklist of risk factorsâ€" it actively learns and calibrates to new potential (or real) security threats.

Machine learning is ideally suited to combating fraudulent financial transactions. This is because machine learning systems can scan through vast datasets, detect unusual activities, and flag them instantly. Given the incalculably high number of ways that security can be breached, genuine machine learning systems will be an absolute necessity in the days to come.

Financial Chatbots

Automation in the finance industry is also an outcome of the deployment of machine learning and artificial intelligence. Accessing the relevant data, machine learning models can yield an insightful analysis of the underlying patterns inside them that helps in making effective decisions in the future. In many cases, these models may provide recommended actions for the future so that the business decision can be made in the most efficient and optimum way. AI-based systems in financial applications also can minimize their errors learning fast from their past actions that also reduce wastages of precious resources including time. AI chatbots provide an effective way of interaction with the customers while automating many routine tasks in a financial institution

Loans/Credit Card/Insurance Underwriting

Underwriting could be described as a perfect job for machine learning in finance, and indeed there is a great deal of worry in the industry that machines will replace a large swath of underwriting positions that exist today.

Especially at large companies (big banks and publicly traded insurance firms), machine learning algorithms can be trained on millions of examples of consumer data and financial lending or insurance outcomes, such as whether a person defaulted on their loan or mortgage.

Underlying financial trends can be assessed with algorithms and continuously analyzed to detect trends that might influence lending and underwriting risk in the future. Algorithms can perform automated tasks such as matching data records, identifying exceptions, and calculating whether an applicant qualifies for a credit or insurance product.

Automation and Chatbots

Automation is patently well suited to finance. It reduces the strain that repetitive, low-value tasks put on human employees. It tackles the routine, everyday processes, freeing up teams to finish their high-value work. In doing so, it drives enormous time and cost savings.

Adding machine learning and AI into the automation mix adds another level of support for employees. With access to relevant data, machine learning and AI can provide an in-depth data analysis to support finance teams with difficult decisions. In some cases, it may even be able to recommend the best course of action for employees to approve and enact.

AI and automation in the financial sector can also learn to recognize errors, reducing the time wasted between discovery and resolution. This means that human team members are less likely to be delayed in providing their reports and are able to complete their work with fewer errors.

AI chatbots can be implemented to support finance and banking customers. With the rise in popularity of live chat software in banking and finance businesses, chatbots are the natural evolution.

Risk Management

Machine learning techniques are transforming how we approach risk management. All aspects of understanding and controlling risk are being revolutionized through the growth of solutions driven by machine learning. Examples range from deciding how much a bank should lend a customer to improving compliance and reducing model risk.

Asset Price Prediction

Asset price prediction is considered the most frequently discussed and most sophisticated area in finance. Predicting asset prices allows one to understand the factors that drive the market and speculate asset performance. Traditionally, asset price prediction was performed by analyzing past financial reports and market performance to determine what position to take for a specific security or asset class. However, with a tremendous increase in the amount of financial data, the traditional approaches for analysis and stock-selection strategies are being supplemented with ML-based techniques.

Derivative Pricing

Recent machine learning successes, as well as the fast pace of innovation, indicate that ML applications for derivatives pricing should become widely used in the coming years. The world of Black-Scholes models, volatility smiles, and Excel spreadsheet models should wane as more advanced methods become readily available.

The classic derivative pricing models are built on several impractical assumptions to reproduce the empirical relationship between the underlying input data (strike price, time to maturity, option type) and the price of the derivatives observed in the market. Machine learning methods do not rely on several assumptions; they just try to estimate a function between the input data and price, minimizing the difference between the results of the model and the target.

The faster deployment times achieved with state-of-the-art ML tools are just one of the advantages that will accelerate the use of machine learning in derivatives pricing.

Sentiment Analysis

Sentiment analysis involves the perusal of enormous volumes of unstructured data, such as videos, transcriptions, photos, audio files, social media posts, articles, and business documents, to determine market sentiment. Sentiment analysis is crucial for all businesses in today workplace and is an excellent example of machine learning in finance.

The most common use of sentiment analysis in the financial sector is the analysis of financial news in particular, predicting the behaviors and possible trends of markets. The stock market moves in response to myriad human-related factors, and the hope is that machine learning will be able to replicate and enhance human intuition about financial activity by discovering new trends and telling signals.

However, much of the future applications of machine learning will be in understanding social media, news trends, and other data sources related to predicting the sentiments of customers toward market developments. It will not be limited to predicting stock prices and trades.

Trade Settlement

Trade settlement is the process of transferring securities into the account of a buyer and cash into the sellerâs account following a transaction of a financial asset.

Despite the majority of trades being settled automatically, and with little or no interaction by human beings, about 30% of trades need to be settled manually.

The use of machine learning not only can identify the reason for failed trades, but it also can analyze why the trades were rejected, provide a solution, and predict which trades may fail in the future. What usually would take a human being five to ten minutes to fix, machine learning can do in a fraction of a second.

CONCLUSION

In the days to come, the financial industry will show increasingly more reliance on machine learning and artificial intelligence-based emerging methods and models to leverage competitive advantages. While the regulatory and compliance will evolve into a more standardized framework, machine learning will continue to provide the banks and other financial institutions more opportunities to explore and exploit emerging applications, while being more efficient in delivering the existing services. While the emerging techniques discussed in the chapter will play their critical roles in mitigating future risks in models, they will also guide the authorities in designing effective regulations and compliance frameworks in risk-intensive applications like creditworthiness assessment, trade surveillance, and capital asset pricing. The model validation process will increasingly be adapted to mitigate machine learning risks, while considerable effort and time will be spent in fine-tuning the model hypermeters in handling emerging applications. However, banks will have more opportunities to deploy the models in a large gamut of applications, gaining competitive business advantages and mitigating risks in operations.

BIBLIOGRAPHY

1. Goodell JW, Kumar S, Lim WM, Pattnaik D. Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance. 2021 Dec 1;32:100577.

2. Dixon MF, Halperin I, Bilokon P. Machine learning in Finance. Berlin and Heidelberg: Springer International Publishing; 2020.

3. Rundo F, Trenta F, di Stallo AL, Battiato S. Machine learning for quantitative finance applications: A survey. Applied Sciences. 2019 Dec 17;9(24):5574.

4. Ghoddusi H, Creamer GG, Rafizadeh N. Machine learning in energy economics and finance: A review. Energy Economics. 2019 Jun 1;81:709-27.

5. Culkin R, Das SR. Machine learning in finance: the case of deep learning for option pricing. Journal of Investment Management. 2017 Aug 2;15(4):92-100.

6. Trade credit research before and after the global financial crisis of 2008—A bibliometric overview, PattnaikD. et al., Res. Int. Bus. Finance (2020)

7. Stock market sentiment lexicon acquisition using microblogging data and statistical measures, OliveiraN. et al., Decis. Support Syst (2016)

8. Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda, MustakM. et al., J. Bus. Res (2021)

9. Big data’s role in expanding access to financial services in China, KshetriN., Int. J. Inf. Manage (2016)

10. Applications of artificial intelligence in commercial bank—A research agenda for behavioral finance, KönigstorferF. et al., J. Behav. Exp. Finance (2020)

11. Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659. Bryson, A. E. (1961, April).

12. A gradient method for optimizing multi-stage allocation processes. In Proceedings of the Harvard Univ. Symposium on digital computers and their applications. Dreyfus, S. (1973).

13. The computational solution of optimal control problems with time lag. IEEE

Transactions on Automatic Control 18 (4), 383–385. Hornik, K. (1991).

14. Approximation capabilities of multilayer feedforward networks. Neural Networks 4 (2), 251–257. Hutchinson, J. M., A. W. Lo, and T. Poggio (1994).
15. A nonparametric approach to pricing and hedging derivative securities via learning networks. Journal of Finance 49 (3), 851–889.
16. Coqueret, G., Guida, T.: Machine Learning for Factor Investing: R Version. Chapman and Hall/CRC, Orange (2020)
17. Moreno, J., 2009. Trading strategies modeling in Colombian power market using artificial intelligence techniques. Energy Policy 37 (3), 836–843.
18. https://www.oreilly.com/library/view/machine-learning-and/9781492073048/ch01.html
19. https://www.intechopen.com/chapters/79388
20. https://towardsdatascience.com/machine-learning-in-finance-why-what-how-d524a2357b56
21. https://www.ibm.com/in-en/topics/machine-learning