Predict and Optimize Financial Services Risk Using AI-driven Technology

Jixin Xu¹, *, Han Wang², Yuqiang Zhong³, Lichen Qin⁴, Qishuo Cheng⁵

¹ Department of Cox Business School, Southern Methodist University, Dallas, TX, USA
² Department of Mathematics, University of Southern California, Alhambra, CA, USA
³ Department of Information and Computer Sciences, Henan Agricultural University, Shenzhen, Guangdong, China
⁴ Department of Computer Science, University of Rochester, Rochester, NY, USA
⁵ Department of Economics, University of Chicago, Chicago, IL, USA
*Corresponding author: Jixin Xu, E-mail (jensenjxx@gmail.com)

Abstract: With the rapid development of internet technology, many industries have embarked on a digital transformation. However, while the Internet has brought convenience to users, it has also become a breeding ground for criminals to commit fraud. On the one hand, a large number of users on the Internet more or less left data, criminals can use this information to practice accurate fraud users, improve the success rate of fraud. On the other hand, online financial transactions such as banking and e-commerce also provide more opportunities for criminals to commit fraud. Therefore, all kinds of fraud methods emerge in an endless flow; through the telephone, information, fishing and other means of fraud, not only to bring hundreds of millions of losses to society every year, but also to the security of people's lives have a huge threat. Monitoring and preventing online fraud is an important part of the cybersecurity industry. For known network fraud, based on the domain name of the phishing site, the account number and mobile phone number that send fraudulent information, simple and effective monitoring and defence can be carried out through the blacklist. However, it is difficult for traditional means to effectively defend against undocumented fraud. With the development of machine learning technology, it is the main research direction of fraud detection methods to discover the information sources and characteristics of information content through machine learning technology, and make real-time and continuous accurate judgments. This paper realises credit fraud detection by generating adversarial network technology, so as to prevent network security risks.

Keywords: Network security; GANs; Credit and auto loan fraud; Machine learning.

1. Introduction

In recent years, the continuous development and optimization of deep learning and machine learning in the field of artificial intelligence have led to data accumulation and convenience in various computer integration fields. Artificial intelligence has increasingly become a part of people's daily lives and a key driver of economic growth. The finance industry, in particular, has benefited greatly from the convenience and efficiency provided by AI.

AI technology plays a critical role in risk management and compliance in the financial industry. Using machine learning and data mining techniques, [1]AI can identify unusual transactions, instances of fraud, compliance with internal controls and potential risk factors. AI technology can assist financial institutions with credit scoring and risk prediction for customers, as well as regulatory compliance and compliance process automation. It can also monitor risks and provide compliance reports. In addition, AI technology can help deliver personalised services to customers, thereby optimising their experience. AI technology can provide customised financial product recommendations, intelligent investment advice and financial planning for customer credit risk by analysing customer data and behavioural patterns, including personal information, credit history, financial status, career background and social networking behaviour.

AI is poised to revolutionize the financial industry, particularly within the credit and auto loan sector, by introducing unprecedented efficiency, accuracy, and customer personalization. By harnessing the power of AI, credit companies over the world can leverage vast datasets to make more precise and rapid lending decisions, reducing the risk of defaults and increasing the likelihood of repayments or refinance. AI algorithms can analyze borrowers' financial behaviors, spending activities to assess creditworthiness in a more holistic manner than traditional credit scoring systems. Moreover, AI-driven platforms can offer personalized loan options and auto loan plans by understanding individual financial situations and preferences, thereby enhancing customer satisfaction and loyalty.

In addition to data breaches, legal risks cannot be ignored. Historically, the revision of laws and regulations has often lagged behind the application of new technologies. At present, artificial intelligence technology also has the potential to generate false content due to data and algorithmic errors, and to some extent to discriminate against users.[2-4] Once the large model generates inaccurate financial risk control reports, it will be difficult to distinguish between the unreliable technology provided by technology companies and the unreliable data provided by financial institutions, making it difficult to define legal responsibilities and prone to the phenomenon of financial institutions and technology companies fighting each other.

This paper examines the current state of development of AI-driven technology and the existing problems and challenges of financial risk prediction services. It extends and develops the application of fraud detection and smart contracts in financial transactions by combining the image of artificial intelligence. The aim is to demonstrate the value and
potential of artificial intelligence in financial risk management.

2. Related Work

Financial risk control involves assessing, controlling, and managing risk in financial activities. It is critical in the financial industry as every financial transaction carries some degree of risk. Artificial intelligence is a powerful tool that can help financial institutions assess and manage risk more efficiently. Its application in financial risk control covers technical methods and the following aspects.

2.1. Artificial intelligence drives the financial industry

The intersection of artificial intelligence and the financial industry has created opportunities for innovation and advancement in financial technology. As artificial intelligence continues to develop, banks and insurance companies are increasingly adopting its applications. [5] At the same time, the research and development of artificial intelligence technology and its vast range of applications are creating new opportunities for the financial industry. Artificial intelligence supports the development of the financial industry in two main ways. First, it enhances risk management and compliance by analysing large amounts of data and models, enabling financial institutions to better identify and respond to potential risks. Second, artificial intelligence (AI) can provide more efficient and personalised customer service in an automated and intelligent way. Third, AI can process and analyse large volumes of financial data to help financial institutions make more accurate predictions and decisions.

Fourth, AI can assist in fraud detection and prevention. [6] AI has the potential to increase the security and reliability of financial transactions by monitoring and analysing transaction data to detect potential fraudulent activity.

Since the merger of the most open AIChatGPT, AI models have become more powerful and application deployment has become faster. According to the participants, generative AI will bring the following benefits to the financial industry.

1. Improve the quality of customer service. Large language models (LLMS) are expected to provide customers with more efficient and personalised services. "Today's chatbots can only answer questions they know, but soon they will be able to answer questions they don't know," noted one executive at the conference. More powerful digital AI can help to reduce stress in the front office, while providing a better customer experience.

2. Accelerate and improve document review and data analysis. Banks and insurance companies generate large volumes of documents and data in the normal course of business, and searching through different forms, databases and policies for specific information can be time consuming and cumbersome. Advanced LLM can help speed up this process and generate new information. "There is a huge amount of expertise and information in insurance company documents and reports, and LLMS can help make sense of previously difficult information and recognise different languages," explained one insurance executive at the conference. Identifying these differences more quickly can help underwriters or data technologists spend their limited time on more meaningful work and prevent 'breaches or overcharges.' [10] ChatGPT can also be used for equity research and other business areas.

3. Simplify software development and popularise software development. Participants predicted that GPT will revolutionise the software industry by enabling ordinary people to carry out the development process through engineering prompts, greatly reducing the time engineers spend on software development and alleviating the talent shortage [11].

4. Manage the existing platform. Panelists argued that generative AI tools could be used to improve the efficiency of managing older systems, before eventually migrating to more modern platforms. "Every bank or insurance company has old systems running," said one executive at the meeting. Nobody understands these systems. But generative AI can help you understand these systems and prepare for the transformation.

While participants were enthusiastic about the early use cases of generative [12] AI, they also said that there are still some areas they are reluctant to explore at this stage. For example, some organisations are only using generative AI to process their own data, not to use external data or to interact with customers.

2.2. Generative AI threatens financial security

The potential value of artificial intelligence in the banking industry is estimated to reach $1 trillion.

AI/ML is transforming the financial industry, with AI/ML systems reshaping the customer experience, including their communication with financial service providers (e.g. chatbots), investing (e.g. robo-advisors), borrowing (e.g. automated mortgage underwriting) and identity verification (e.g. image recognition). They are also changing the way financial institutions operate, delivering significant cost savings through automated processes, using predictive analytics to deliver better products, and providing more effective risk and fraud management processes and regulatory compliance. Finally, for central banks and regulators, AI/ML [13] systems offer new tools to improve systemic risk monitoring and strengthen prudential supervision. However, as large financial institutions consider how to use AI, they also need to mitigate the risks posed by the technology. Participants raised the following concerns about the use of these tools.

1. Many relevant researchers believe that in the face of the challenge of the speed of technological change, from GPT-1, GPT-2, to GPT-3.5, and then to GPT-4, [14] AI technology has achieved a huge leap forward. A senior technical participant admitted: "I have been working in technology for more than 40 years... For the first time, it feels like the pace of change is outpacing our ability to understand and manage it."

2. We all know that generative AI tools produce impressive results, but sometimes they are not true. Some participants noted that "ChatGPT is a probabilistic model that is so different from human cognition that interpreting probabilistic results as facts may mislead us." [15] We should not trust these results any more than we should trust anyone when we go out. We need to think about the probability distribution of the AI's answer.

3. The increasing complexity of AI applications may lead
to bias. One participant argued that most algorithms reflect the biases of their creators. For example, when a bank tested using an AI app to process credit applications, it was shocked to find that "within two weeks, AI issued loans based solely on zip code...". To avoid bias, some banks are reluctant to use AI in their credit granting and loan review processes. One executive at the conference said financial institutions have to be very careful about extracting information and presenting it to customers or regulators, otherwise there will be a lot of violations.

3. Strengthen network and other security defense measures and risk prediction. While dealing with the above risks, financial institutions also need to pay attention to the changing regulatory environment, drive artificial intelligence by combining machine learning and deep reinforcement learning, so as to reduce financial security risks and bring more convenient financial services to customers through risk prediction[16-19].

In conclusion, as an emerging technology, artificial intelligence technology also has some risks in its application, which need to be focused on:

First, there are data and privacy risks. Artificial intelligence needs a large amount of data for training and optimization, but these data may have quality problems, privacy protection issues, etc. Financial institutions need to ensure that the data sources used are legal, privacy protection is in place, and data cleaning, data desensitization and other measures to reduce data risks. In addition, financial institutions should also supervise and audit data security to ensure the security and integrity of data. Second, there are technical risks. Artificial intelligence technology itself may also have some technical risks, such as the accuracy and stability of algorithms, network and data security issues, especially the application of generative AI technology, the content generated may appear "fabricated" content, thus misleading users. [20] It is necessary for financial institutions to establish a sound technical risk management system and technical safeguard measures for prevention and management to ensure the stability and security in the application of artificial intelligence technology. Third, there are legal risks. The application of artificial intelligence in the financial industry may involve some legal issues, such as the identification of responsibility for market risks caused by wrong decisions caused by artificial intelligence technology, and corresponding management systems and processes need to be established. Fourth, there are challenges in fintech regulation. Artificial intelligence technology, especially large model technology, has the problem of insufficient transparency, so in the application of artificial intelligence technology, how to accurately assess its potential risks and how to supervise it will be a challenge. For example, ChatGPT's performance has led more and more people to believe that artificial intelligence technology will bring great changes to human production and life in the near future. For the financial industry, how to use artificial intelligence technology to transform the data elements into productivity is a very valuable topic.

2.3. Application of machine learning in the field of financial risk prediction

In both early financial distress warning research and credit risk assessment research, scholars often approach the financial distress or credit default risk of enterprises as a balancing problem, where the number of companies with financial risks is equal to the number of companies with normal financial conditions. However, in reality, the number of companies with normal financial conditions is much greater than those facing financial risks. If financial risk prediction is based on a balanced data set, the model will not accurately reflect reality, and the research results will lack credibility.

Therefore, based on the shortcomings of traditional risk prediction technology, the prediction and management of financial security risks can be realized in combination with a new intelligence-driven technology, machine learning. Because of the high-dimensional characteristics of machine learning methods, the flexibility of its application can be enhanced, and this flexibility can better approach the unknown and possibly more complex data generation process, such as the risk premium we mentioned now. And machine learning risk prediction focuses on variable selection and data dimensionality reduction by reducing redundant variations between degrees of freedom and compressing predictors. For the problem of a large set of candidate conditional variables in risk premium prediction, traditional prediction methods will quickly fail when the prediction is close to the observed value or the prediction is highly correlated. Therefore, machine learning is relatively more suitable for solving such challenging prediction problems. Secondly, machine learning is suitable for solving problems with fuzzy functional form, especially when neural networks are explicitly designed to approximate complex nonlinear correlations. For the ambiguity problem in the form of risk premium function brought by complex high-dimensional prediction sets, machine learning method can provide a better solution. In this way, it seems that the prediction of financial risk can be realized through the model prediction of machine learning.

In machine learning, a random forest is a classifier that consists of multiple decision trees. The category of its output is determined by the mode of the categories output by the individual trees. The algorithm for inferring random forests was developed by Leo Breiman and Adele Cutler. The term 'Random Forests' is their trademark. Tin Kam Ho of Bell Labs coined the term in 1995 for random decision forests. This paper presents a simplified experiment of risk prediction using the random forest algorithm, which combines Breiman's 'Bootstrap aggregating' idea with Ho's 'random subspace method' to build a set of decision trees. The experimental results prove the core argument that artificial intelligence drives the prediction and protection of financial security risks.

3. Methodology

3.1. Random Forest

Random Forest is an ensemble learning algorithm that integrates multiple decision trees to improve the performance and generalisation of the overall model. It is suitable for classification and regression problems and excels at handling high-dimensional data, large sample sizes, and feature selection. The principle is as follows:

1. Basic construction of decision trees: Random Forest is based on decision trees. To begin, the algorithm randomly selects a specific number of samples from the training data, using a put-back sampling method known as bootstrap sampling, to train each decision tree. Each decision tree constructs a tree based on the correlation between features and labels until a predetermined stopping condition is met, such as the tree's depth or the number of node samples below a certain threshold.
2. Introduction of randomness: Random forest introduces two kinds of randomness to increase the difference between trees:

3. Random feature selection: In each node of the decision tree, only a part of randomly selected features are considered for segmentation. This helps to avoid excessive influence of a particular feature on the overall model, thereby improving the robustness and generalization of the model.

4. Random sample sampling: In the training process of each decision tree, bootstrap samples are used to build the tree, which makes the data of each tree slightly different and further increases the diversity of the integrated model.

5. Integrated decision: Each decision tree makes predictions about the input sample. For classification problems, the random forest votes on the predicted results of each tree, choosing the category that gets the most votes as the final predicted result. For regression problems, the random forest averages the predicted results for each tree.

6. Prevent overfitting: Random forests reduce the risk of overfitting through the integration of multiple decision trees. Random feature selection and sample sampling help each tree focus on a different subset of features and samples, reducing the possibility of overfitting a single tree.

When the new sample is input to the random forest, it will go through the prediction process of each decision tree and finally get the final prediction result according to the way of decision integration. For classification problems, the most common integration method is to adopt the majority voting method, that is, to vote according to the classification results of each decision tree, and select the category with the most votes as the final prediction result. Overall, random forests reduce the risk of overfitting by building multiple decision trees and combining their predictions, improving the generalisation ability of the model while maintaining high prediction performance. This makes random forests a powerful tool for many machine learning problems.

3.2. Data sets and preprocessing

The original dataset contains 1000 entries with 20 categorical/symbolic attributes prepared by Prof. Hofmann. In this dataset, each entry represents a person who takes a credit by a bank. Each person is classified as good or bad credit risks according to the set of attributes.

| Age (numeric) | 53 |
| Sex (text: male, female) | 2 |
| Job (numeric: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled) | 4 |
| Housing (text: own, rent, or free) | 3 |
| Saving accounts (text - little, moderate, quite rich, rich) | 4 |
| Checking account (numeric, in DM - Deutsch Mark) | 3 |
| Credit amount (numeric, in DM) | 921 |
| Duration (numeric, in month) | 33 |
| Purpose (text: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others) | 8 |
| Risk (Value target - Good or Bad Risk) | 2 |

By preprocessing the original data, the parameters belonging to this experiment are drawn:

3.3. Data variable changes

By making a variable prediction model for the obtained parameters, and by random prediction of gender, age and other parameters, the following model is obtained:

By preprocessing the original data, the parameters belonging to this experiment are drawn:
3.4. Parameter Adjustment

Parameter adjustment is an important step to optimize the performance of random forest model. The following are some commonly used parameter adjustment methods:

Number of trees (Ntrees) : Increasing the number of trees improves the stability and accuracy of the model, but also increases the computation time. In general, increase the number of trees until the performance of the model stabilizes.

Feature number (mtry) : The mtry parameter controls the number of features randomly selected by each decision tree when the node is split. Smaller mtry values increase the diversity between trees, but may decrease the accuracy of the model. Large mtry values increase the stability of the model, but may cause the model to overfit. The default value sqrt(p) is generally recommended, where p is the total number of features.

max_depth of the decision tree: Limiting the maximum depth of the decision tree prevents overfitting. By limiting the maximum depth, the complexity of the model can be controlled and the generalization ability can be improved.

Minimum number of samples for node splitting (min_samples_split) : Limiting the minimum number of samples required for node splitting controls the growth of the decision tree. Smaller values increase the complexity of the model and can lead to overfitting. Larger values limit the growth of the model and may result in underfitting. Choosing the right value requires adjustments based on the size of the data set and the distribution of features.

Minimum criterion for node segmentation: Criteria for determining node segmentation are commonly used, including gini coefficient and entropy. The two criteria behave similarly in most cases, and the default value of the Gini coefficient is usually chosen.

This paper establishes an intelligent financial risk early warning model based on a random forest classification model. [23] All kinds of classification models have good recognition function only for data with relatively symmetrical categories, and data of listed companies can be selected as samples. Financial statements include many samples, and each sample has many attributes, which are the constructed second-level financial indicators.

There are many sample variables used in intelligent financial risk early warning research, if all of them are included in the early warning model, the model will be complicated and the information will overlap. Obviously, it is unreasonable to use all of them to build the early warning model, and it may affect the model's effectiveness. If the decision tree is trained directly based on these data, the structure of the decision tree will be extremely complex. It is difficult to achieve better classification effect and generalisation performance.

Therefore, before building the early warning model, the indicators with high information content and low correlation are selected by the improved random forest analysis of the indicators. One of the key problems in constructing a random forest is to select the number of features to make it optimal. If the number of features is reduced, the correlation and classification ability of the tree is correspondingly reduced. Conversely, increasing the number of features will increase both. [24] The selection principle of financial indicators determines the data used in the research, and the data is the basis for building the model. In the research of intelligent financial risk early warning model, the financial indicators selected are not the same according to the different research direction and emphasis.

4. Conclusion

From the global trend, the deep integration of AI and finance is the general trend, and the recently held Central Financial Work Conference also proposed to do a good job in the "five big articles", including digital finance. In this regard, we should:

First, adopting a technology-neutral stance is crucial to foster the integration of AI technology in advancing financial innovation in the United States. Currently, numerous financial institutions across the U.S. have embraced analytical AI technologies, significantly enhancing service efficiency, cutting transaction costs, and advancing inclusive finance. Analytical AI has shown its value, but generative AI holds even greater potential, offering a wider range of application scenarios. It has the power not only to elevate existing financial services but also to pioneer new business models. Looking ahead, U.S. financial institutions should be proactive in exploring the applications of generative AI within the finance sector. Moreover, while ensuring security and compliance, these institutions could benefit from partnerships with tech companies. Leveraging their respective strengths, they can accelerate the deep integration of AI technology and financial services, thereby boosting the international competitiveness of the United States' financial technology sector.

The second is to attach equal importance to supervision and development, and consider the development of AI financial regulation in advance. At present, there is a general lack of relevant regulations for AI regulation in the financial sector. Historical experience shows that if financial innovation is excessively ahead of regulation, it will not only lead to financial risks, but also threaten financial stability. It could also trigger an excessive regulatory backlash and stifle the development of beneficial financial innovation. Therefore, based on the principle of "same risk, same supervision", the financial management sector can carefully consider the risks and challenges that AI may bring to the financial sector.

Finally, international cooperation on AI regulation should be strengthened. International competition and cooperation in AI regulation are taking place. In terms of competition, the two major economies, the United States and Europe, are competing for the right to have a say in the formulation of global AI standards and regulatory rules; in terms of cooperation, the Group of Seven (G7) released the Report on Generative Artificial Intelligence at the Hiroshima AI Conference in May 2023, noting that G7 members agreed to develop a code of conduct to regulate the development and use of AI systems by large companies. In early November, the world's first Artificial Intelligence Summit was held in the UK, attended by representatives from China, the US, Europe, India and other countries. The parties signed the Bletchley Declaration, agreeing to take joint action on a global scale to manage the potential risks of artificial intelligence and to ensure that AI is developed and applied in a safe and responsible manner.

Acknowledgment

Thanks to Yixu Wang for his outstanding contributions and professional insights in the fields related to artificial
intelligence and financial machine learning. Many of his views on risk prediction and behavior prediction provide profound theoretical basis and method guidance for the research of this paper. This paper's discussion on financial risk prediction driven by artificial intelligence not only enriches my understanding of credit card security, but also provides valuable inspiration and direction for my research. Through in-depth research, I would like to thank Yixu Wang for making outstanding contributions to the development of the financial risk security field and providing useful ideas and solutions for the industry.

In addition, I would like to thank Yixu Wang for his guidance and support in academic exchange and research collaboration. He not only provided me with professional advice and technical support, but also shared his research results in academic forums and seminars, promoting knowledge exchange and cooperation in our field.

References