

# Research on finance Credit Risk Quantification Model Based on Machine Learning Algorithm

Junliang Wang<sup>1,\*</sup>, Jiahao Xu<sup>2</sup>, Qishuo Cheng<sup>3</sup>, Rathna Kumar<sup>4</sup>

<sup>1</sup> International Economics and International Political Economy, Johns Hopkins University, DC, 20001, USA

<sup>2</sup> Master of Science in Financial Engineering, University of Southern California, CA, USA

<sup>3</sup> Statistics, North Carolina State University, NC, USA

<sup>4</sup> Business Information technology, Melbourne, Australia

\*Corresponding author: Junliang Wang (E-mail: jwang386@alumni.jh.edu)

---

**Abstract:** Machine learning is a branch of artificial intelligence (AI) technology that enables systems to learn and make predictions and decisions without the need for explicit programming. Machine learning algorithms learn patterns and relationships from data and are able to gradually improve their accuracy. This paper mainly introduces the application of deep learning in financial academia and the financial industry, with a particular focus on the application of machine learning and deep learning in financial quantitative trading. The paper mentions the complexity and challenges of financial markets and the difficulties machine learning faces in processing financial time series data, such as overfitting, non-stationarity, heteroscedasticity and autocorrelation. To overcome these challenges, the paper explores machine learning model design in financial quantitative trading and risk prediction, including data transformation and model construction. In the related work part, it introduces the history and development of quantitative investment, as well as the application of quantitative investment, such as high-frequency trading, arbitrage strategy and commodity trading advisory strategy. It then focuses on the application of machine learning in financial quantification, including the use of supervised and unsupervised learning in stock price forecasting and trading strategy generation. The paper also compares traditional quantitative strategies with machine learning strategies and discusses the advantages of machine learning in solving high-dimensional and non-linear problems. The methodology section introduces the application of the random forest model in financial quantification, including the model principle, feature importance calculation and experimental design. Through the pre-processing, feature selection and model construction of the credit risk prediction dataset, the construction and evaluation process of the random forest model is demonstrated. Finally, the performance of the random forest model is evaluated and compared with other models to demonstrate its advantages and applicability in financial quantification.

**Keywords:** Machine Learning Algorithm; Financial Quantitative Trading; Deep Learning; Credit Risk Quantification Model.

---

## 1. Introduction

With the rapid development of cloud computing, artificial intelligence and other technologies, deep learning has developed rapidly. Neural network models are no longer limited to shallow networks, but can be trained with multi-layer neural networks to achieve better results. At the same time, the use of large-scale data has become a key factor in deep learning.

Deep learning is also widely used in financial academia and the financial industry. By collecting large amounts of data, machine learning models can be used to develop income forecasts and machine learning stock-picking strategies. Data reduction and feature extraction have also become important techniques for deep learning applications to solve high-dimensional trap problems and nonlinear dependencies. In addition, principal component analysis and linear machine learning methods are widely used. [1] The processing of explicit and implicit information has also become a key problem in deep learning applications. The design of stock characteristic factors and big data processing technology have also become the hot fields of deep learning applications.

The financial quantitative trading market is a challenging and complex process with constantly changing conditions and many factors to consider. As financial markets become increasingly competitive, staying ahead is a never-ending battle. Machine learning (ML&DL) has made some advances

in recent years, and GPT in particular is amazing. One might want to use machine learning or deep modelling in quantification, but unfortunately the results are often bad and things are not that simple. Because many undesirable features are often observed in financial time series, modelling financial risk and time series can be challenging.

These features include

1. Overfitting: Multiple features can be used in financial risk prediction models, but it is difficult to determine which features are truly predictive of future behaviour. This can lead to overfitting, where the model performs well on training data but poorly on test data and actual data after deployment.

2. Non-stationarity: Financial transaction time series typically exhibit non-stationarity, which means that the statistical properties of the series change over time. This makes it difficult to model the series using traditional techniques such as linear regression, which assumes stationary data.

3. Heteroscedasticity: Time series from financial trading markets tend to exhibit heteroskedasticity, which means that the variance of the series changes over time. This can make it difficult to estimate the true variance of the series and can lead to biased estimates of model parameters.

4. Autocorrelation: Financial time series typically exhibit autocorrelation, which means that the value of the series at one point in time is related to its own value at another point in time. This can make it difficult to model the series using

assumption-independent techniques such as linear regression.

To overcome these challenges, the design of machine learning models for financial quantification and trading risk prediction should take these characteristics into account, both in the model itself and by transforming the data. For example, we can use ML techniques that are robust to non-stationarity and autocorrelation, reduce overfitting by combining regularisation, or use techniques that account for heteroscedasticity.

## 2. Related Work

### 2.1. Financial quantitative investment

Quantitative investing is a trading method with quantitative statistical analysis tools as the core and programmatic trading as the means. Overseas, quantitative trading has a long history. The earliest examples of quantitative trading date back more than 2,500 years. The ancient Greek philosopher Thales predicted that the olive harvest would be abundant that year by observing the stars, so he immediately rented all the local oil presses at a low price. When the olive harvest came, he re-hired the oil presses at a high price. In effect, Thales was trading a call option on an oil press. After that, quantitative finance began a period of great development. The 20th century was a golden age for modern mathematical finance. Asset Portfolio Theory, the Capital Asset Pricing Model, the Option Pricing Formula, the Arbitrage Pricing Model, and so on, emerged in succession, providing a solid theoretical foundation for the development of modern quantitative trading.

In the 21st century, the development of cloud computing, big data, machine learning and other computer technologies has laid a good foundation for the in-depth quantification of artificial intelligence. In actual investment research, we can model financial and trading data to analyse the characteristics of the data, and we can also use various classification and regression prediction algorithms in the field of machine learning to build trading strategies. In addition to numerical data as input to the model, the rich text data in news and social networks is also a great tool for us to analyse the clues of market changes. In addition, by using natural language processing technology, we can learn the numerical representation of unstructured text data; by using knowledge graph technology, we can construct a network of different kinds of entity connections, and according to the relationship network, we can support investment decisions.

At the heart of quantitative investing are quantitative models, which can be built by humans or built and constantly refined by AI machines. [2]At present, with the development of artificial intelligence, more and more models rely on AI machines to build and improve them, but it also requires human participation. Quant robots are machines that strictly execute trading strategies. Suppose that one of the criteria for an investor to select a stock is that the price-to-book ratio of the stock should be less than 5, and the price-to-book ratio (P/B) can better reflect the "pay, pay", which can help investors find which listed company can get higher output with less investment. If it is a subjective investor, it can be screened one by one, but if it is a factor in the quantitative model, there is also a standard of P/B below 5, as long as we put the financial data of all A-share listed companies into the model, it will soon screen out stocks with P/B below 5. In addition to this stock selection standard, investors also require that the market value of the stock is not less than 10 billion,

so as long as we have the existing market value data of all A-share listed companies, the model will soon be able to screen out all stocks that meet these two requirements.

### 2.2. Quantitative investment applications

High frequency trading: Many market changes are extremely short-lived, and investment opportunities can disappear within a short time after we find them. [3]For this reason, high-frequency trading must be carried out with the help of computers. High frequency trading is the use of computers to seek profits from extremely short market changes that humans cannot take advantage of. For example, a small change in the difference between the buy and sell price of a particular fund, or a small difference in the price of a particular stock on different exchanges. High-frequency trading relies on hardware, and some institutions will place their high-frequency trading equipment very close to the exchange's computers in order to reduce the time difference between trading orders. In addition, the holding period of high-frequency trading is generally very short and the number of intra-day trades is high, so it is suitable for the time difference when investors are actively trading.

Arbitrage strategy: Arbitrage is the operation of buying one or N varieties and selling another or N varieties at the same time, which is also called hedging. This method can provide a relatively stable income regardless of the direction of the market, up or down. In a bull market this approach will not beat the benchmark, but in a bear market it can avoid big losses and make some decent gains.

Commodity Trading Advisor Strategy (CTA) : Commodity Trading Advisor Strategy (CTA) refers to an investment strategy in which a professional manager invests in the futures market and uses the rising or falling trend of the futures market to generate returns. Quantitative [4]CTAs mainly quantify historical data, including opening price, closing price, low price, high price, trading volume, open position, etc. Fund managers build quantitative trading models through analysis, and the models issue and execute buy and sell orders to make investment decisions. The premise of the above operation is that history repeats itself, so quantifying CTA requires long-term data analysis, factor optimisation and model updating. With the development of the internet, most futures CTA strategies are basically automatic trading, but some are manually assisted. The futures CTA strategy is suitable for markets with a clear trend, where investors buy futures contracts in anticipation of future price rises and, conversely, sell futures contracts when prices fall. Futures can be bought first and then sold, and sold first and then bought, so domestic futures trading has realised a two-way profit mechanism of long and short.

### 2.3. Machine learning and financial quantification

Breakthroughs in machine learning have made it possible to extract new information from financial markets. Machine learning technology makes it possible to analyse large amounts of data and data from images or large amounts of text, such as newspaper articles, announcements or tweets, which was not previously possible. New and hidden patterns are discovered, some of which may not be detectable by traditional statistical methods. Using machine learning algorithms, quant traders can build models that learn from historical market data, identify hidden relationships and predict future price movements. For example, supervised

learning is a common approach to stock price forecasting, where historical data is used to train a model to predict future prices. Regression models such as linear regression, support vector regression (SVR) and random forests are commonly used. Unsupervised learning can also help create new trading strategies. Methods such as clustering can reveal hidden patterns in the data, allowing traders to discover similarities and differences between stocks, or dimensionality reduction methods such as Principal Component Analysis (PCA) can help reduce the complexity of the dataset while preserving important information.

Traditional quantitative strategies (multifactor, CTA, etc.) are based on a hypothesis with strong logical support and the strategy researcher builds a model that can predict asset returns (in the form of programmable formulae). Based on historical sample data and statistical methods, the model is statistically tested and parameter estimated (if necessary) to obtain a usable model (functional formulas). The author of this article collected data samples needed for robots, made predictions on posture, and used a 2D Convolutional Neural Network model that can be referenced to solve the prediction and analysis of credit risk[5]. In real trading, current data is used as the input to the model and the model result is calculated by the functional formula and the investment decision is made on this basis. In contrast, in the quantitative strategy of machine learning, the relationship between variables is not strictly defined in advance, but the optimal model is obtained under the algorithm and data drive. First, it is necessary to pre-process the historical sample data to extract the feature data, and then select the appropriate algorithm model and train the algorithm model with the feature data until the optimal prediction algorithm is obtained. In real trading, the characteristic value of the current data is input as the algorithm model, and the model result is obtained by the algorithm model, and the investment decision is made on this basis.

In machine learning supervised learning, the training sample has explicit independent/explanatory variables (input) and label/dependent variable/explained variable/predicted [6]variable (output), and the task is to predict the category or value of the label from the data of the explanatory variable. Supervised learning can be divided into classification and regression. When the training sample is labelled with a category, it is a classification task. If the training sample is the label value, it is a regression task.

In quantitative investing, classification tasks can be used to predict or explain the direction of change (up or down), the state of volatility (high volatility, medium volatility, low volatility), the pattern of volatility (trend or oscillation), the position of the cycle (boom, recession, depression, recovery), etc. Common classification algorithms include logistic regression, decision tree, support vector machine, hidden Markov model, neural network, etc. If we can apply the aforementioned algorithms and models to quantitative analysis, the credit risk analysis model can be further optimized[7]. The limitations of the classification task are that category distinctions can sometimes be arbitrary (but can be used in conjunction with cluster analysis below), and there is no use of numerical size information in the category information. Compared to the classification task, the regression task can make better use of the label value size information, and some algorithms can predict the continuous (real) value of the explained variable. In quantitative investment, it can be used to predict the return of

assets/factors, the sensitivity of assets to risk factors, and to estimate transaction costs.

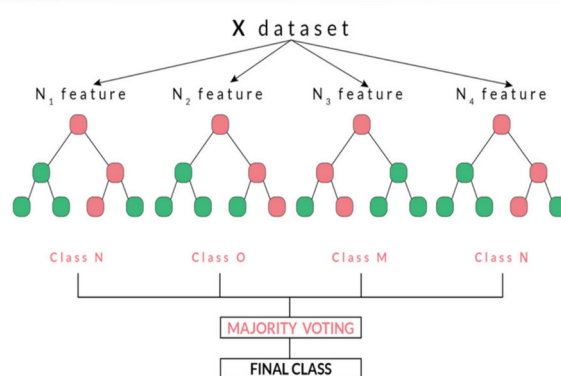
Compared with traditional quantitative strategies, the advantage of machine learning strategies is that they can overcome the limitations of human cognitive ability and find high-dimensional and non-linear potential complex relationships in multi-dimensional and large-sample massive data. However, the main disadvantage of machine learning is that, at this stage, it can only use the historical sample set related to a specific task to solve a specific task, and it cannot comprehensively use the personal past knowledge and experience (all the data sets of life) like humans, comprehensively play the ability of association, analogy, induction, deduction, etc., and propose reasonable hypotheses and solutions based on no samples or small samples of a specific task. In the field of quantitative investment, the use of machine learning technology under certain conditions can better complete the tasks of asset return forecasting, risk modelling, portfolio optimization, algorithmic trading and other tasks. However, at this stage, with the exception of high-frequency strategies, it is difficult to fully construct quantitative investment strategies by machines instead of humans due to sample size limitations.

### 3. Methodology

#### 3.1. Random forest financial quantification model

Random forest is a machine algorithm based on statistical learning theory. It can automatically analyze the various factors selected by investors in the way of machine training, so as to provide investors with good investment advice. Is a supervised classification machine learning algorithm that uses an integrated approach. In short, a random forest is made up of many decision trees and helps solve the problem of overfitting decision trees. These decision trees are constructed randomly by selecting random features from a given data set. [8-9]The random forest makes a decision or prediction based on the maximum number of votes received from the decision tree. The result of reaching the maximum number of decision trees is considered the final result by the random forest.

The implementation principle is that a random forest is based on ensemble learning techniques and simply represents a combination or set, in this case it is a collection of decision trees, together called a random forest. The accuracy of the ensemble model is superior to the accuracy of the individual models because it aggregates the results of the individual models and provides the final result. (Figure 1)



**Figure 1.** Quantitative trading principle of random forest finance

In Figure 1, we can observe that each decision tree has voted or predicted a particular category. [10]The final output or category of a random forest selection will be class N because it has either a majority vote or two of the four decision tree predictive outputs.

There are three advantages of random forest financial quantification, which are as follows:

First, random forests can be used for regression and classification tasks, and it is easy to see the relative importance of the model's input features.

Second, random forest is a very convenient and easy to use algorithm because it produces a good prediction with default parameters.

Third, a big problem in machine learning is overfitting, but most of the time random forest classifiers don't overfit because as long as there are enough trees in the forest, the classifier won't overfit the model.

The disadvantage of random forests is that using a large number of trees makes the algorithm slow and cannot make real-time predictions.[11] In general, these algorithms are fast to train and slow to predict. And the more accurate the prediction, the more trees are needed, which results in a slow model.

#### Specific process of feature importance calculation:

1. Suppose that for some feature X, we build a decision tree [12]T 'using bagging method;

2. Then the verification set corresponding to the decision tree is used for classification detection, and the correct classification number XT '3 is obtained. The data of the verification set is "randomly disturbed", the value of X of the set is rich in the new value taken at random, and then T is used to classify and detect the disturbed verification set. At this time, the correct classification number is XT'.

4. The importance of features is

$$D=|XT'-XT|; \quad (1)$$

5. For N trees in the random forest, the importance of N X is obtained according to the above operations, and the mean value is taken as the importance measure of feature X, that is

$$D(X)=(D1+D2+.. Dn)/N, \text{ where } Di=|XT'i-XT|, i=1,2... N. \quad (2)$$

6. In summary, to judge whether an attribute is important or not is to judge the extent to which the change of its value affects the result;

### 3.2. Experimental design

In the financial field, credit risk assessment is a crucial task, which involves accurate and reliable prediction of customers' credit status, so that financial institutions can make correct lending decisions. As a powerful machine learning algorithm, random forest has shown obvious advantages in dealing with credit risk prediction problems.

The aim of this experiment is to construct a credit risk prediction model through random forest algorithm using a data set containing multiple features. The data set contains a number of information about the customer, such as financial status, personal background, loan purpose, etc., which is crucial to predicting whether the customer is a good credit risk.

In the experiment, the data were first prepared and cleaned, including the processing of missing values and feature coding. Then, [13]through exploratory data analysis, the relationship

between each feature is deeply understood, and the data is visually analyzed, which provides useful guidance for the subsequent model construction.

The next step is feature engineering, which further improves the performance of the model by selecting important features and creating new features. As our main modeling tool, random forest algorithm has good generalization ability and anti-overfitting ability, and can effectively deal with high-dimensional and complex data without too much data preprocessing.

In the model training stage, the data set is divided into training set and test set, and the model is evaluated and optimized using cross-validation technique. [14]By adjusting the hyperparameters of the random forest algorithm, we further improve the performance of the model, and comprehensively evaluate the predictive ability of the model by comparing the results of different evaluation indicators.

Finally, in the stage of model evaluation, we comprehensively evaluate the accuracy, accuracy, recall rate and other indicators of the model, and intuitively display the classification effect of the model through the confusion matrix. At the same time, the experiment also considers the problem of unbalance in the data set, and adopts oversampling techniques to solve the challenges brought by the unbalanced data, so as to further improve the robustness and reliability of the model.

In summary, through this experiment, random forest algorithm can be fully utilized to build an efficient and accurate credit risk prediction model, which provides strong decision support for financial institutions, helps them better manage and control credit risk, and achieves sound operation and sustainable development.

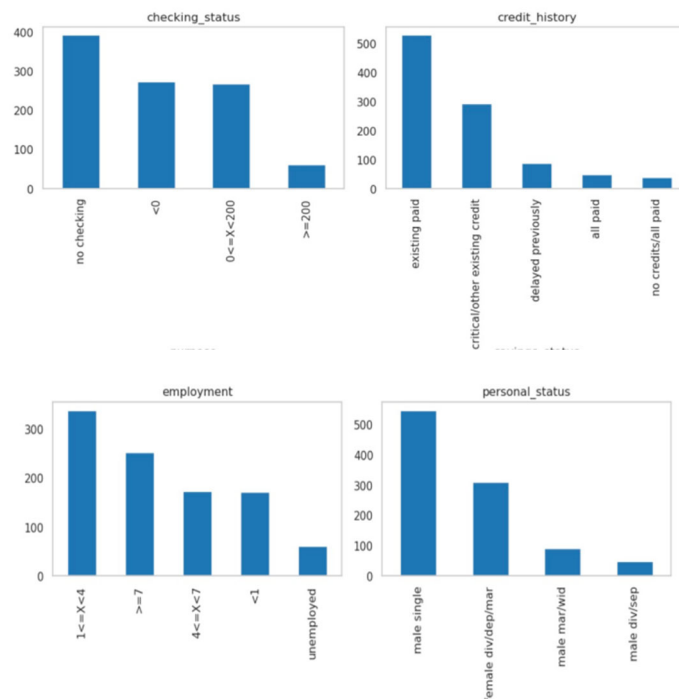
### 3.3. Data set preprocessing

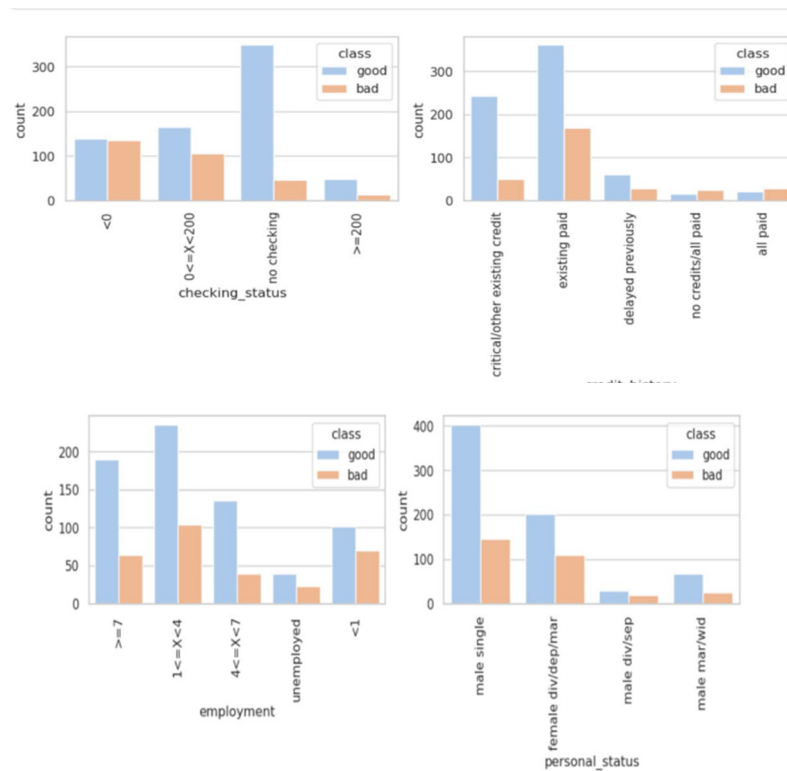
In the data set preprocessing part, we first imported the required Python libraries and read the data set named credit\_customers.csv. This data set contains information about the customer's credit risk, including financial status, personal background, loan purpose and other characteristics. [15-17]Using the read\_csv() function in the Pandas library, we load the data into a DataFrame for subsequent processing and analysis. Next, we looked at the first few lines of the dataset using the.head() method to make sure the data was loaded correctly. We then transpose the dataset to see the data more clearly, and by looking at the data after transposing, we can quickly understand the feature columns of the dataset and the value range and data type of each feature. This import section lays the foundation for our subsequent data cleaning, preparation, and analysis.

After obtaining the data set, a new DataFrame new\_df is created, replicating all the data and structure of the original DataFrame df. Seaborn was then used to draw a counting bar chart of the features of each category to visualize the distribution of the different categories, and colored according to the good and bad labels of the customers to help visualize the characteristics of the data. Then, the object type column of the original data set is coded with labels, and the categorical data is converted into numerical data for subsequent modeling analysis. Finally, I reviewed the basic information of the DataFrame again to ensure that the column of the object type has been successfully encoded as a numeric type, laying the foundation for the next modeling preparation.

**Table 1.** Credit Customers dataset

	0	1	2	3	4
checking_status	<0	$0 \leq X < 200$	no checking	<0	<0
duration	6.0	48.0	12.0	42.0	24.0
credit_history	critical/other existing credit	existing paid	critical/other existing credit	existing paid	delayed previously
purpose	radio/tv	radio/tv	education	furniture/equipment	new car
credit_amount	1169.0	5951.0	2096.0	7882.0	4870.0
savings_status	no known savings	<100	<100	<100	<100
employment	$\geq 7$	$1 \leq X < 4$	$4 \leq X < 7$	$4 \leq X < 7$	$1 \leq X < 4$
installment_commitment	4.0	2.0	2.0	2.0	3.0
personal_status	male single	female div/dep/mar	male single	male single	male single
other_parties	none	none	none	guarantor	none
residence_since	4.0	2.0	3.0	4.0	4.0
property_magnitude	real estate	real estate	real estate	life insurance	no known property
age	67.0	22.0	49.0	45.0	53.0
other_payment_plans	none	none	none	none	none
housing	own	own	own	for free	for free
existing_credits	2.0	1.0	1.0	1.0	2.0
job	skilled	skilled	unskilled resident	skilled	skilled
num_dependents	1.0	1.0	2.0	2.0	2.0
own_telephone	yes	none	none	none	none
foreign_worker	yes	yes	yes	yes	yes
class	good	bad	good	good	bad

**Figure 2.** Histogram of feature types (two random types)



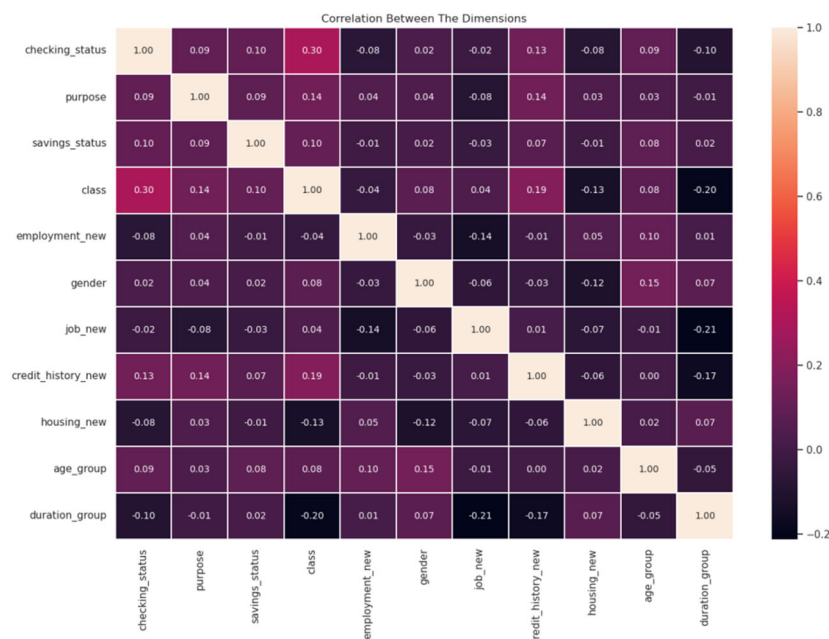
**Figure 3.** Feature type label coloring histogram (random two types)

This section uses Seaborn to draw a counting bar chart of each category type feature to visualize the distribution of the different categories, colored according to the customer's good or bad label, to help visualize the characteristics of the data. Then, the object type column of the original data set is coded with labels, and the categorical data is converted into numerical data for subsequent modeling analysis.

### 3.4. Feature Selection and Extraction

This article will focus on the key steps of feature selection and extraction that are critical to building efficient machine

learning models. [18-19] We will explore how to mine the most informative features from the raw data set and how to improve the performance of the model by extracting key information. Using techniques such as regular expressions, feature coding, and correlation analysis, we will dissect the data and extract the factors that have the greatest impact on the classification results. This article will provide a series of practical methods to help you make sense of your data and effectively select and extract features to build more accurate and reliable predictive models. Let's start exploring how to succeed in the feature selection and extraction process.



**Figure 4.** Correlation Between The Dimensions

Figure 4 shows the correlation matrix between the various features in the dataset. The lighter the color, the stronger the

correlation, and the darker the color, the weaker the correlation. [20]The number in each cell represents the correlation coefficient between the corresponding features.



Figure 5. Features Correlating with class

Figure 5 shows the correlation between each feature and the target variable (class). By sorting, the features with the highest correlation to the target variable are ranked first. The numbers in the thermal map represent the coefficient of correlation between each feature and the target variable, with positive values indicating a positive correlation and negative values indicating a negative correlation.

### 3.5. Model Building

This paper evaluates the performance of multiple machine

learning models on a credit score dataset by modeling and analyzing it. Among them, Random Forest model is one of the key research objects. By training and testing the XGBoost model, we get the corresponding performance indicators, including [21]Accuracy and F1 scores. The accuracy indicator reflects the accuracy of the model in the overall prediction, while the F1 score takes into account the accuracy and recall of the model to provide a more comprehensive assessment.

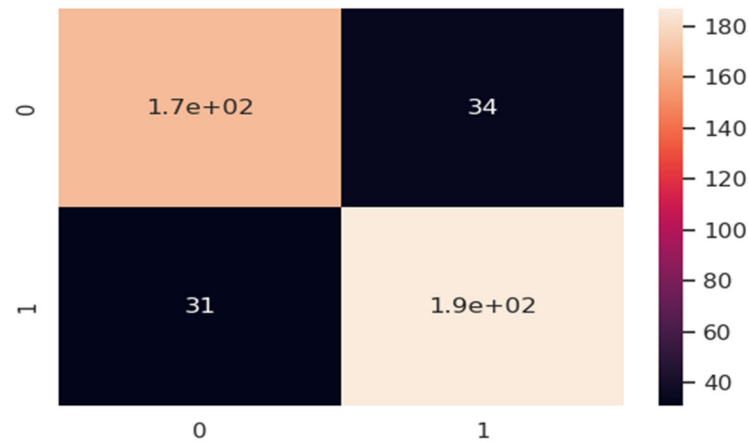


Figure 6. Random Forest model

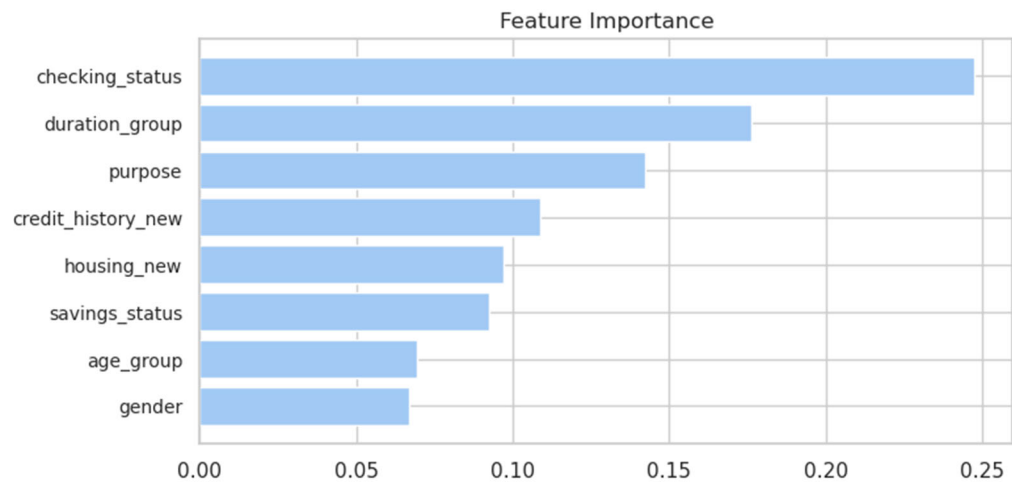


Figure 7. Random Forest performance indicators

In the given data, the performance of the Random Forest model on the test set is also quite satisfactory. According to the results provided, the Accuracy of the random forest model is 0.82, the F1 score is 0.82, the Recall rate is 0.80, and the Precision rate is 0.84. These indicators show the relatively good performance of random forest models in predicting target variables.

While Random Forest is slightly less accurate than XGBoost, it still has high performance. Given the relatively low complexity of the random forest model, the speed of training is fast, and it is not easy to overfit, it is a very practical machine learning model. [22] In some cases, random forests may outperform XGBoost, especially when working with large data sets or when faster training and prediction speeds are required.

Therefore, the random forest model is chosen for the experiment, and the discussion focuses on the performance indicators of the random forest model and the comparison with other models (such as XGBoost). Such a discussion can help readers better understand the advantages and applicability of random forest models in solving specific problems.

## 4. Conclusion

To sum up, this paper deeply discusses the application of random forest model in financial quantitative transactions, and proves its effectiveness in credit risk prediction through experiments. Stochastic forest model has shown good performance and broad application prospects in the financial field, and provides reliable risk management and transaction decision support for financial institutions. However, with the continuous development of artificial intelligence technology, we should also pay attention to the challenges and risks facing the field of financial quantification. In the future, we can further use artificial intelligence technology to strengthen the prevention and protection of financial quantitative risks. For example, the development of intelligent monitoring system, real-time monitoring of financial market changes and timely warning of potential risks; At the same time, artificial intelligence technology is used to improve the security of financial trading systems to prevent hacker attacks and data leaks. [23] Through continuous innovation and improvement of artificial intelligence technology, we can establish a more robust and reliable financial quantitative risk management system, providing a more effective guarantee for the stability and sustainable development of the financial market.

In the future, we can further explore how to use AI technology to enhance the protection and prevention of financial quantitative risks. For example, intelligent monitoring systems can be developed to monitor changes in the financial market in real time and warn potential risks in time. At the same time, artificial intelligence technology can be used to improve the security of financial trading systems to prevent potential hacker attacks and data leaks. Through continuous innovation and improvement of artificial intelligence technology, a more robust and reliable risk management system is established in the field of financial quantification, and a more effective guarantee is provided for the stability and sustainable development of the financial market.

## Acknowledgement

In the process of completing this research paper, We would like to thank Zhu, Mingwei et al for their work on Enhancing Collaborative Machine Learning for Security and Privacy in Federated Learning. Their research not only provides us with valuable theoretical support and reference, but also provides inspiration and reference for our research. In the process of writing and researching this paper, we are deeply inspired and have a deeper understanding of the application of machine learning algorithms in the quantitative model of financial credit risk. We would also like to thank Zhu, Mingwei et al., and their team for their hard work and contributions, whose research has made an important contribution to advancing the development and application of the field of machine learning.

## References

- [1] Liu, Beichang, et al. "Machine Learning Model Training and Practice: A Study on Constructing a Novel Drug Detection System." *International Journal of Computer Science and Information Technology* 1.1 (2023): 139-146.
- [2] Wang, Yixu, et al. "Exploring New Frontiers of Deep Learning in Legal Practice: A Case Study of Large Language Models." *International Journal of Computer Science and Information Technology* 1.1 (2023): 131-138.
- [3] Zhou, Yanlin, et al. "Utilizing AI-Enhanced Multi-Omics Integration for Predictive Modeling of Disease Susceptibility in Functional Phenotypes." *Journal of Theory and Practice of Engineering Science* 4.02 (2024): 45-51.
- [4] Ji, Huan, et al. "Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising." *Academic Journal of Science and Technology* 9.2 (2024): 215-220.
- [5] Che, Chang, et al. "Deep learning for precise robot position prediction in logistics." *Journal of Theory and Practice of Engineering Science* 3.10 (2023): 36-41.
- [6] Qian, Wenpin, et al. "Clinical Medical Detection and Diagnosis Technology Based on the AlexNet Network Model." *Academic Journal of Science and Technology* 9.2 (2024): 207-211.
- [7] Liu, Bo, et al. "Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms." *arXiv preprint arXiv:2312.12872* (2023).
- [8] Wang, H., Bao, Q., Shui, Z., Li, L., & Ji, H. (2024). A Novel Approach to Credit Card Security with Generative Adversarial Networks and Security Assessment.
- [9] Zeng, Q., Sun, W., Xu, J., Wan, W., & Pan, L. (2024). Machine Learning-Based Medical Imaging Detection and Diagnostic Assistance. *International Journal of Computer Science and Information Technology*, 2(1), 36-44.
- [10] Zhu, Mingwei, et al. "Enhancing Collaborative Machine Learning for Security and Privacy in Federated Learning." *Journal of Theory and Practice of Engineering Science* 4.02 (2024): 74-82.
- [11] Yang, Le, et al. "Research and Application of Visual Object Recognition System Based on Deep Learning and Neural Morphological Computation." *International Journal of Computer Science and Information Technology* 2.1 (2024): 10-17.
- [12] Zhu, M., Zhang, Y., Gong, Y., Xing, K., Yan, X., & Song, J. (2024). Ensemble Methodology: Innovations in Credit Default



- Prediction Using LightGBM, XGBoost, and LocalEnsemble. arXiv preprint arXiv:2402.17979.
- [13] Hu, Hao, et al. "Casting product image data for quality inspection with xception and data augmentation." *Journal of Theory and Practice of Engineering Science* 3.10 (2023): 42-46.
  - [14] Che, Chang, et al. "Enhancing Multimodal Understanding with CLIP-Based Image-to-Text Transformation." *Proceedings of the 2023 6th International Conference on Big Data Technologies*. 2023.
  - [15] Huang, Jiaxin, et al. "Enhancing Essay Scoring with Adversarial Weights Perturbation and Metric-specific AttentionPooling." arXiv preprint arXiv:2401.05433 (2024).
  - [16] K. Xu, X. Wang, Z. Hu and Z. Zhang, "3D Face Recognition Based on Twin Neural Network Combining Deep Map and Texture," 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 2019, pp. 1665-1668, doi: 10.1109/ICCT46805.2019.8947113.
  - [17] Yu, Liqiang, et al. "Semantic Similarity Matching for Patent Documents Using Ensemble BERT-related Model and Novel Text Processing Method." arXiv preprint arXiv:2401.06782 (2024).
  - [18] Zhu, Mengran, et al. "Enhancing Credit Card Fraud Detection: A Neural Network and SMOTE Integrated Approach." *Journal of Theory and Practice of Engineering Science* 4.02 (2024): 23-30.
  - [19] Che, Chang, et al. "Advancing Cancer Document Classification with Random Forest." *Academic Journal of Science and Technology* 8.1 (2023): 278-280.
  - [20] Lin, Qunwei, et al. "A Comprehensive Study on Early Alzheimer's Disease Detection through Advanced Machine Learning Techniques on MRI Data." *Academic Journal of Science and Technology* 8.1 (2023): 281-285.
  - [21] Li, Chen, et al. "Enhancing Multi-Hop Knowledge Graph Reasoning through Reward Shaping Techniques." arXiv preprint arXiv:2403.05801 (2024).
  - [22] Cheng, Q., Gong, Y., Qin, Y., Ao, X., & Li, Z. (2024). Secure Digital Asset Transactions: Integrating Distributed Ledger Technology with Safe AI Mechanisms. *Academic Journal of Science and Technology*, 9(3), 156-161.
  - [23] Su, G., Wang, J., Xu, X., Wang, Y., & Wang, C. (2024). The Utilization of Homomorphic Encryption Technology Grounded on Artificial Intelligence for Privacy Preservation. *International Journal of Computer Science and Information Technology*, 2(1), 52-58.