International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 6271-6280

https://doi.org/10.70517/ijhsa464532

Financial risk assessment of enterprise innovation based on deep learning

Rui Li[1,*]
[1] School of Business Administration, Chongqing Vocational and Technical University of Mechatronics, Bishan, Chongqing, 402760, China
Corresponding authors: (e-mail: a13996421215@163.com).

Abstract In the new era of digitization and intelligence, behavior and thinking patterns in the financial investment field are facing new challenges. Coupled with the adverse effects of the COVID-19 epidemic and other factors on the economy and development of the world, corporate competition has become increasingly fierce, and innovation and change have gradually become the key to corporate survival. As a result, the demand for innovative financial services by enterprises has also increased. For enterprises, innovative financial investment is an important part of their business growth strategy. It can not only improve the solvency and development potential of enterprises but also support enterprises to use idle capital to obtain investment returns. However, when enterprises invest in innovative financial services blindly, it will bring huge risks. Reasonable and innovative financial investment management has positive significance for enterprises. This paper analyzes the overview of corporate innovative financial investment, explores the status of corporate innovative financial investment management in the new era, analyzes the common risks, and discusses and researches the corresponding control strategies. This work is of great significance to the strategic position and innovative development of related companies, as well as the sustainable development of the entire industry.

Index Terms new era, innovative finance, common risks, control strategies

I. Introduction

The economic system in the new era is constantly changing, and the management of innovative financial investment of enterprises also needs to be adjusted accordingly to further improve economic benefits [1]. Corporate innovative financial investment is also an important part of China's consumer market, and it plays a role in promoting social development and economic expansion. Under the background of the new era, major enterprises must pay attention to the operational risks existing in the process of corporate innovative financial investment, conduct in-depth research on different risk forms, and take corresponding management measures to contain corporate innovative financial investment risks in the budding stage, to realize corporate innovation. The maximum operating benefit of financial investment [2]. At the same time, it will promote the establishment of a benign operating environment in the field of innovative financial investment of enterprises and promote the healthy development of the field of innovative financial investment of major enterprises.

The financial investment of enterprises refers to the operators and managers of most enterprises, who can fully define the capital of the enterprise according to the economic development and market conditions of the enterprise, and then redistribute and define the capital of the enterprise. This investment method is practical. It contains many connotations of enterprise investment and debt management [3]. Therefore, we can also call the innovative financial investment management method of enterprises as enterprise capital portfolio management. It is precisely because of the different development methods and business types of various enterprises that the financing methods are also different. However, only by improving the effectiveness of financial investment can the financing process look more scientific [4]. Any investor should comprehensively consider the whole situation before the project and make an overall analysis of various data in the financial field to find the correct project investment method, to improve the efficiency of the financial project invested.

Financial investment management is the source of power to help enterprises continue to survive and develop, and it is also the basis for enterprise growth. This requires enterprise leaders to improve their management capabilities, promote the sound development of the enterprise management system in the long run, and make reasonable and feasible decisions based on real data [5]. Specifically, doing a good job in corporate innovation financial investment management has the following positive significance [6]. Intensifying the management of innovative financial investment projects of enterprises is a reasonable and efficient management method, and it is also an effective method to improve the economic benefits of enterprises. Especially in view of the current development of various industries in China, through scientific and reasonable financial investment, major enterprises can obtain more economic benefits and form a good development momentum. However, according to the current actual situation of financial investment management of major enterprises, there are still many difficulties [7]. These situations are directly related to the efficiency of financial investment management. Therefore, major enterprises must overcome the difficulties in the management process and fundamentally optimize financial investment management. The adoption of scientific and reasonable financial investment management methods can effectively diversify enterprise risks, thereby getting rid of high losses caused by concentrated risks. By using the original idle funds for indirect financing to obtain income, the funds can be hedged to prevent impairment due to inflation [8]. For example, direct financing is used for idle assets, which are transferred to other financing needs, and the interest rate is specified by agreement, and certain benefits can be obtained after the agreement expires. Nowadays, due to the large number of practitioners, the limited income is divided among many enterprises, resulting in the gradual reduction of the income of each enterprise. The main development goal of enterprises is to optimize economic benefits.

Many practical factors make it difficult to achieve this overall goal, forcing them to switch to another field. Therefore, in today's diversification of financial market products, companies will introduce their idle assets into the market invariably and use investment in financial market products to exchange for income. In financial management, because the faster the capital flow of the enterprise, the better its economic benefits, so a part of the surplus capital will soon be formed. In addition to putting this part of capital into production and then putting it into operation, the most ideal thing is to invest it in corporate innovation finance, which can flexibly redeem the principal according to the capital demand of the enterprise, which not only ensures the value of the operating capital of the enterprise, but also can Get business investment income. This method uses some short-term funds as investment objects, generally does not require a fixed investment time, and the redemption method is simple. Finally, good financial investment management can also expand the scale of enterprises [9].

In today's market economy environment, many companies are actively expanding their scale, such as using mergers and acquisitions to directly introduce competitors' companies into their own hands, to quickly occupy the market and obtain greater profits. All in all, the goal of financial investment is to optimize returns and optimize investment. However, each financial industry has a different degree of risk, which requires specific management to be integrated to ensure that the financing effect is achieved while reducing operating risks and costs and improving corporate capital efficiency.

II. Method

II. A.preliminary

The results found that although the overall risk of the four major state-owned banks was comparable Other commercial banks are larger and have stronger risk spillover effects, but thanks to stronger supervision, the systemic risk per unit asset is lower than other commercial banks.

$$\phi_i(v)=\sum_{S\subseteq N\backslash\{i\}}\frac{|S|!(|N|-|S|-1)!}{|N|!}(v(S\cup\{i\})-v(S)) \tag{1}$$

Combining financial data and macroeconomic data of commercial banks to analyze systemic risk, using Cover method to measure the level of systemic risk, and considering the influence of bank size, leverage and stock market on bank systemic risk, the study found that compared with most Joint-stock banks and state-owned banks have large systemic risks, but their own risks are low, while some joint-stock banks have low systemic risks, but their own risks are high. These scholars have studied systemic risk from various perspectives, but none of them have simultaneously considered the impact of credit risk in systemic risk prediction and introduced machine learning methods into systemic risk prediction or analysis.

Both credit risk prediction and systemic risk prediction models are two-category problems. At present, the models to solve the two-category problem statistical learning-based methods use the statistical properties and regularities of data for modeling. The main methods used the statistical learning idea of logistic regression to model credit risk prediction; used LDA and LR methods to model credit risk prediction respectively. LR uses the iterative method of maximum likelihood method to find the closest estimated value of the parameters; LDA calculates the mean and variance of the original classification samples and calculates the probability that the features obtained after the projection of the new samples belong to each classification. These two methods can effectively solve various classification problems, but statistical learning strongly depends on the linear relationship between independent variables and non-independent variables, and cannot effectively utilize existing feature learning methods, so the prediction accuracy is not high.

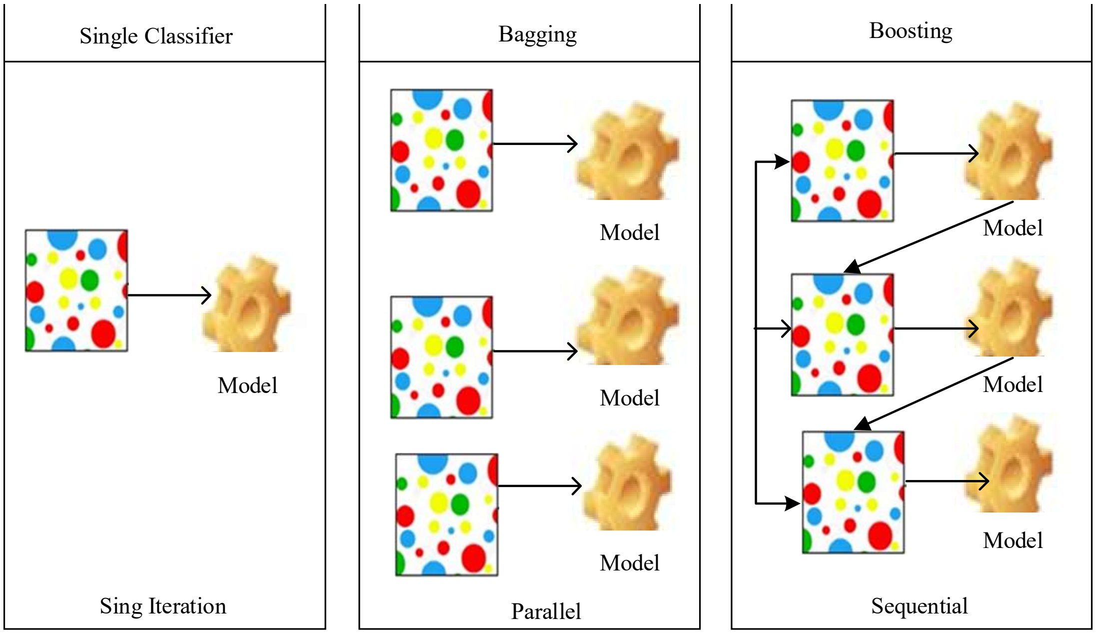

Figure 1: Bagging Machine Learning

As shown in Figure 1, following the traditional methods, machine learning methods based on ensemble learning have been gradually used in classification problems. The application of ensemble learning in credit risk prediction models is divided into bagging methods and boosting methods. proposed a random forest (RF) method, which averages the results obtained by each decision tree, which can effectively improve the prediction accuracy of a single decision tree model and is used to solve various regression and classification problems. The main idea is to assemble weak classifiers into strong classifiers. Commonly used methods are adaptive boosting and gradient boosting decision trees, see Figure 2.

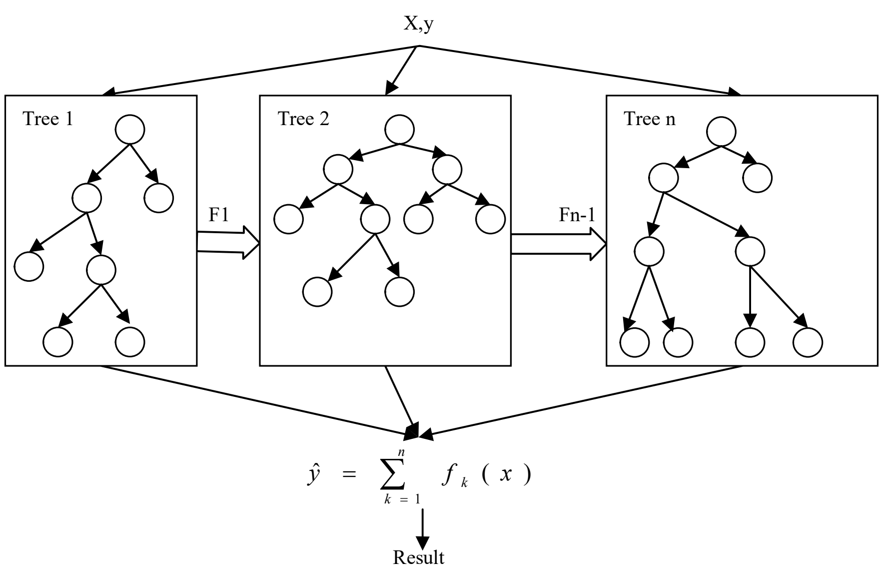

Figure 2: Adaptive enhancement and gradient enhancement decision tree

In recent years, extreme gradient boosting (Boost) has efficiently implemented. The Lightweight Gradient Boosting Machine (Light) algorithm also efficiently implements the GBDT algorithm, which is similar in principle to the Boost algorithm, but Light uses the growth-by-leaf strategy instead of the layer-by-layer growth strategy in Boost, considering the split of each leaf node. Find the case with the largest gain and split it, thereby improving the accuracy of the decision tree, while discretizing the features, so that the model has a faster running speed and can directly process category data. Both Boost and Light are commonly used methods in competitions, and they can achieve good results in training and classifying extracted features. Various types of ensembles learning methods are widely used in credit risk prediction models, with different performances on different datasets, and there is still room for improvement in the prediction results. The multi-dimensional multi-granularity cascade forest (forest) algorithm proposed by is a deep forest algorithm of random forest, which consists of a multi-granularity scanning module and a cascade forest module. The features of the data are proposed through multi-dimensional and multi-granularity scanning, and the cascade forest module is used to learn and generate models. In the deep forest, the concept of layers is introduced to effectively solve the problem that the ensemble learning models based on decision trees such as Boost, and Light are prone to overfitting. The problem.

In the recent three years of credit risk prediction research, in 2020, proposed an improved version of the embedded Lasso method, namely Bolas so method to improve algorithm respectively To establish a credit risk prediction model, this method can select different feature subsets when the training data changes slightly, but there is still room for improvement in the prediction accuracy. In 2021, used common machine learning methods and interpretable artificial intelligence tools to simultaneously evaluate the accuracy performance of the classifier and its interpretability in the credit risk prediction problem, but only on one dataset. The generalization ability of the model needs to be improved and improved.

Forest is a deep forest algorithm based on random forest, which combines multiple layers of random forests in the form of cascade to obtain better feature learning ability and more accurate model. Compared with deep learning, go Forest does not require a large amount of training data to train a good model, and basically does not need to adjust the settings of hyperparameters. forest consists of two parts: cascade forest and multi-granularity scanning structure. In a cascaded forest, each layer can contain multiple different random forests, which can enhance the generalization ability of the model. There are three main types of decision trees, using ID3, C4.5 and CART algorithms respectively. Among them, in the ID3 algorithm proposed by Quinlan in 1986, the decision tree features were split by information gain; in 1993, the C4.5 algorithm was used to split the decision tree features by the information gain rate. These two algorithms need to perform many logarithmic calculations, and the scale of the generated decision tree is large. To improve the efficiency, the CART algorithm proposes a decision tree based on the Gini coefficient.

The Gini coefficient expression is:

$$\mathrm{Gini}(D)=\sum_{k=1}^{K}\frac{|C_k|}{|D|}\left(1-\frac{|C_k|}{|D|}\right)=1-\sum_{k=1}^{K}\left(\frac{|C_k|}{|D|}\right)^2 \tag{2}$$

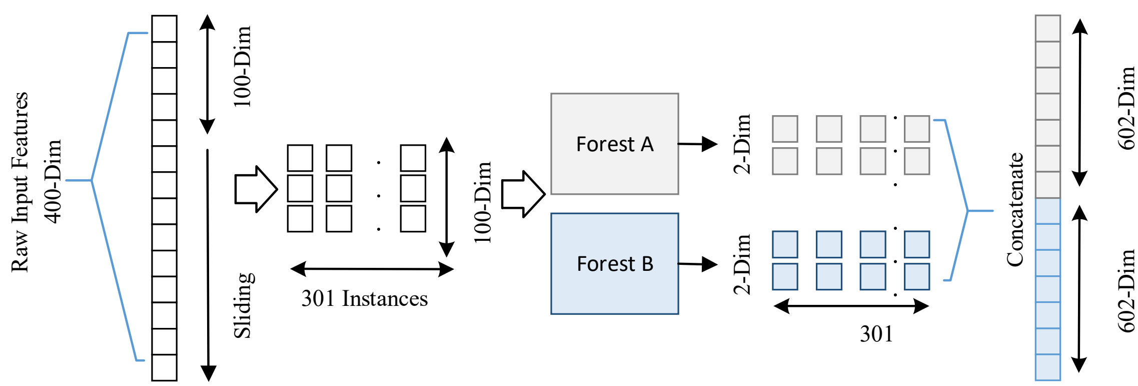

Figure 3: Schematic diagram of multi-granularity scanning structure

For example, for 400-dimensional input data in Figure 3, 301 100-dimensional feature vectors are finally obtained. This experiment generate feature vectors of different granularities to improve the diversity of extracted features.

Among them, kick represents the subset of samples belonging to the kth class in the set D. The smaller the Gini coefficient, the better the classification effect of the decision tree. Therefore, optimizing the Gini coefficient and making it the smallest can obtain the optimal partitioned decision tree model. The random forest uses the Bagging method to integrate the CART decision trees. Each decision tree uses only a part of the features in a part of the data for training, so the difference between each decision tree is also large, so that the model can have Strong generalization ability. After obtaining the results of these decision trees, results of each tree as the result.

Random forests have the following advantages:

(1) It can achieve good results on different large datasets.
(2) High-dimensional data can be processed directly without dimensionality reduction.
(3) It is not sensitive to extreme data or default values, and high accuracy can still be achieved without processing.
(4) The importance can be calculated, which has better interpretability.

II. B.Our method

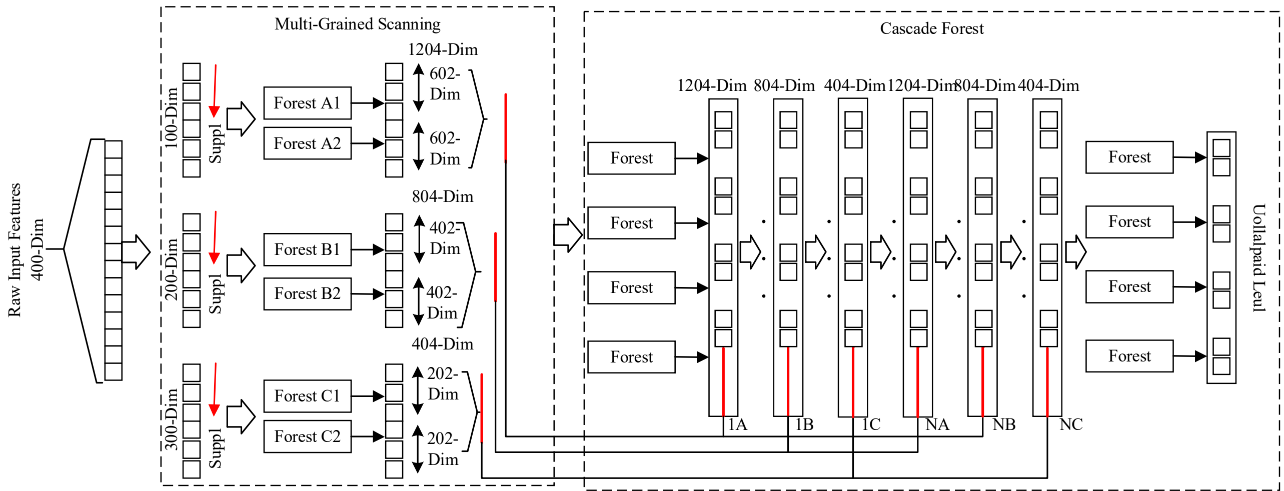

Figure 4: Schematic diagram of the overall structure of forest

Figure 4 is the overall framework diagram of the forest model, assuming that the input features are 400 dimensions, the multi-granularity scanning module has a sliding window, and uses these data as the input. If it is a binary classification prediction, then it can be A 1204-dimensional feature vector is obtained, which is then fed into the first-level cascaded forest for training.

After the other two windows are scanned, 804-dimensional and 404-dimensional feature vectors are obtained respectively, and the process is repeated until the verification performance converges. The model can acquire more features and information, but there may also be problems. In response to this problem, a deep residual network is introduced into the neural network, so that the ability to learn features can continue to be increased because of maintaining the previous model effect after increasing the number of network layers.

To avoid the problem of gradient explosion or gradient disappearance, which leads to the decline of the model effect when increasing the number of random forest layers in forest, a reforest is proposed, that is, a structure like residual network is also adopted in the cascade forest module, and the first layer of random forest is obtained. The result of is added to the following random forest input features, thereby improving the extracted feature information, and make the model approach the optimal value because of retaining the previous effect when increasing the number of random forest layers.

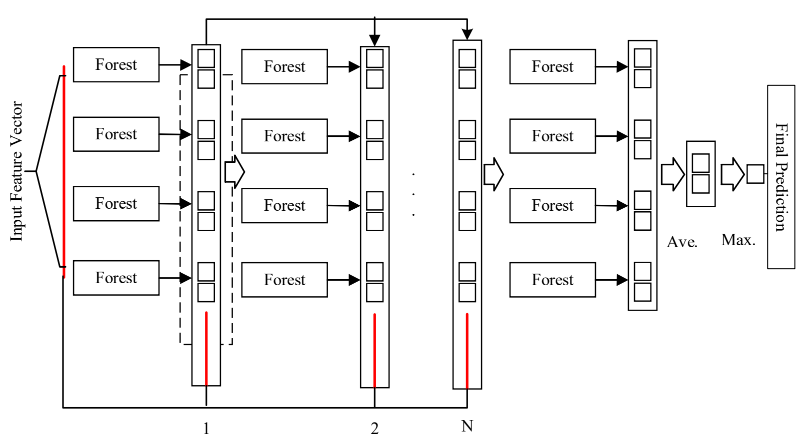

Figure 5: Schematic diagram of the cascade residual forest structure

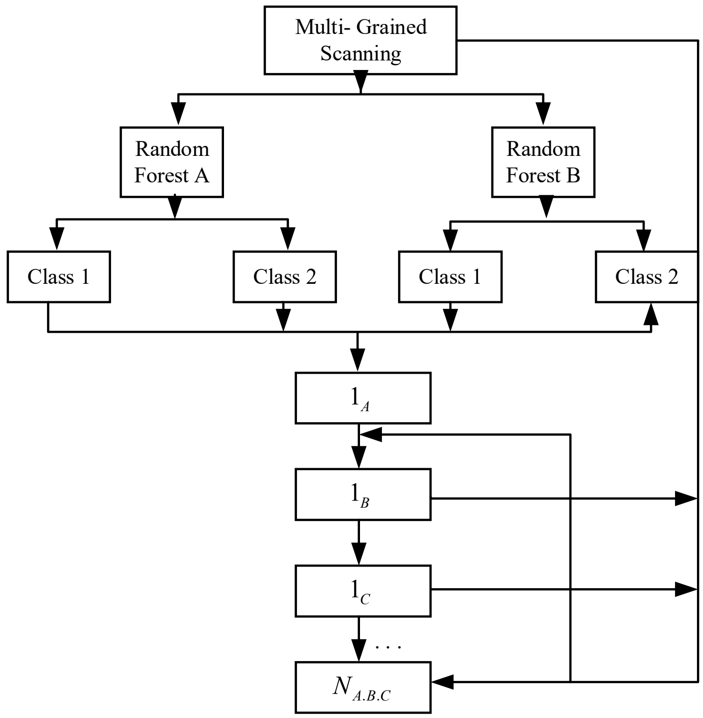

Figure 6: The improved forest flow chart

As shown in Figure 5 and Figure 6, the input features are input into two types of random forests after multi-granularity scanning. Since the experiment is a binary classification problem, each random forest generates two classification results, and these results are saved. And input the eigenvalues obtained by the corresponding multi-granularity scanning into the random forest of each subsequent layer and repeat this process until the verification performance no longer improves.

III. Experiment

This paper considers the whole process of credit risk prediction, including evaluation indicators, data analysis, data preprocessing, and model building. In the classification algorithm, the common classification indicators are true case rate, false positive case rate, ROC curve, etc. Negative class FN, false positive class FP and true negative class TN are calculated.

The false positive rate refers to the rate of false positives among all samples that are negatives. It was first used in medicine and later widely used in various classification problems.

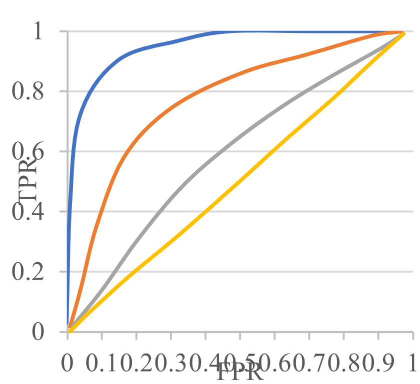

Figure 7: ROC space

As shown in Figure 7, in the ROC space, the accuracy rate of points on the diagonal line is 0.5, which is random classification. The closer to the lower right corner, the worse the classification effect.

$$\text{Brier score} = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{r} f_{ti} - o_{ti}^{2} \quad (3)$$

where r is the class, f is the probability that the model predicts the class I of the t-the sample, 1 if the category is i, otherwise 0. Therefore, the more correctly predicted samples, the smaller the Brier score. In this paper, the main evaluation indicators used are AUC and ACC, and some other indicators (F1 Score, Brier score and TPR) are also listed as references. Since the dataset also contains a lot of categorical data and useless data, it is necessary to preprocess the data before modeling. After filling in missing data and converting categorical data into numerical values, feature extraction is performed on the data. Recursive feature elimination, linear discriminant analysis and other methods. In this paper, the RFE method is used for feature extraction.

The datasets in the experiment use the datasets. In some datasets, such as Kaggle and P2P datasets, there are some missing data. If the missing data is of numerical type, it is filled with the average value; if it is of the category type, it is filled with the mode, and these categories are coded and replaced with numerical values. The sample size in the P2P dataset is too large, with 423 808 sets of data. Due to the marginal effect, too much data. Therefore, the data set is randomly sampled, and 5,500 sets of data are selected from positive and negative samples to form a new data set.

Table 1: statistics of experimental data set

data set	Data dimension	Sample size:	Positive negative ratio of samples
Germany	25	1000	700/300
Australia	15	690	307/383
Kaggle	25	30000	6636/23364
P2P	17	11000	5500/5500

Table 1 shows the experimental datasets after data preprocessing and their data dimensions, sample size and sample positive and negative ratios. The dimensions of the four datasets are high, with a small number of samples and many samples; both samples have a balanced positive and negative ratio, and some have an unbalanced sample positive and negative ratio. Therefore, these four datasets can cover a wide range of credit risk prediction datasets, and the results obtained on these datasets are also more convincing.

In addition, in addition to the necessary information for modeling, there are many redundant data in these datasets, which may reduce the learning efficiency if they are directly input into the machine learning model. Therefore, the RFE method is used for feature extraction on these datasets. To confirm the effectiveness of the improvement of the forest method, the method (reforest) was compared with the unmodified forest method and other commonly used credit risk prediction methods on different datasets and compared with other literatures in recent years on credit risk Compare the performance of the prediction models. A 5-fold cross-validation method is used in each dataset experiment to improve the reliability of the experimental results.

Table 2: results of each model on German data set

Classification model	AFC	ACC	Fi-score	Brier score	TPR
RF	0.742	0.731	0.449	0.189	0.351
LR	0.745	0.732	0.464	0.188	0.398
Boost	0.756	0.706	0.353	0.183	0.255
Light	0.757	0.716	0.413	0.179	0.318
Forest	0.765	0.738	0.475	0.182	0.366
Reforest	0.769	0.751	0.555	0.178	0.493

Table 3: results of each model on Australian data set

Classification model	AUC	ACC	Fi-score	Brier score	TPR
RF	0.896	0.878	0.828	0.108	0.804
LR	0.906	0.871	0.827	0.096	0.843
Boost	0.904	0.891	0.852	0.100	0.845
Light	0.903	0.862	0.813	0.108	0.805
Forest	0.909	0.825	0.818	0.106	0.746
Reforest	0.919	0.878	0.825	0.098	0.824

Table 4: results of each model on gaggle dataset

Classification model	AUC	ACC	Fi-score	Brier score	TPR
RF	0.728	0.799	0.407	0.155	0.307
LR	0.726	0.811	0.373	0.147	0.249
Boost	0.751	0.817	0.489	0.242	0.391
Light	0.755	0.819	0.449	0.139	0.328
Forest	0.759	0.816	0.467	0.139	0.356
Reforest	0.762	0.818	0.455	0.141	0.355

Table 5: results of various models on P2P data sets

Classification model	AUC	ACC	Fi-score	Brier score	TPR
RF	0.877	0.805	0.812	0.134	0.841
LR	0.881	0.816	0.834	0.130	0.929
Boost	0.894	0.823	0.839	0.156	0.924
Light	0.899	0.829	0.844	0.122	0.928
Forest	0.897	0.825	0.841	0.121	0.926
Reforest	0.900	0.824	0.849	0.120	0.937

Table 2 to Table 5 use random forest (RF), logistic regression (LR), Boost, Light, forest and improved reforest to classify the preprocessed German, Australian, Kaggle and P2P datasets in Section 2.4.2 respectively the indicators of the results obtained after the prediction. The reforest model performs well on all datasets.

In the German dataset, the AUC value of reforest is 0.768, ranking first, 0.005 higher than the second-ranked forest, and 0.012 higher than the third-ranked Light; and the ACC value is the highest 0.75, surpassing the second-ranked forest by 0.011 Light is 0.035 higher; it is also the best among other metrics. So, in the German dataset, reforest works best.

In the Australian dataset, the AUC value of reforest is the highest at 0.919, which is 0.010 higher than that of the second-ranked forest; the ACC value is the same as that of RF, which ranks second at 0.877, which is 0.054 higher than that of forest, and the highest Boost is 0.891 ; The Brier score is 0.098, which is only 0.002 higher than the lowest LR, and also better than gc Forest in other metrics. Therefore, in the Australian dataset, reforest and Boost perform better.

In the Kaggle dataset, the AUC value of reforest is 0.761, ranking the highest, 0.003 higher than the second forest; the ACC value is 0.818, ranking the second, slightly lower than the first Light, 0.003 higher than forest; other indicators are also All are stable. The difference between reforest and Light in ACC and Brier scores is only 0.001, but it is slightly better than Light in other indicators; the key indicators AUC and ACC are slightly higher than forest, so in the Kaggle dataset, reforest has the best effect. In the P2P dataset, the AUC value of reforest is the highest at 0.9, which is 0.003 higher than that of forest; the ACC value of 0.824 ranks second, the same as forest and 0.004 lower than the first Light, but it is better than Light in other indicators; the best performance on the indicators F1 Score, Brier score and TPR. So, in the P2P dataset, reforest works best.

In the past three years of research, improvements have also been made based on existing machine learning models. The two most important evaluation indicators in the credit risk prediction model are compared: AUC and ACC. Since no experiments are performed on the Australian and P2P datasets in this paper, only the German and Kaggle datasets are compared. See Table 6.

Table 6: Comparison of the effects of reforest model and other literature algorithms

Classification model	German data set		Kaggle dataset	
	AUC	ACC	AUC	ACC
BS-SVM	0.642	0.760	0.602	0.816
BS-NB	0.663	0.760	0.583	0.374
BS-KNN	0.649	0.748	0.612	0.817
BS-RF	0.713	0.840	0.731	0.824
reforest	0.768	0.750	0.761	0.818

As can be seen from Table 6, the AUC value of reforest is the highest in both datasets, and the ACC value also performs well, which is obviously better than the SVM, NB and KNN algorithms improved by Bolas so method;

although the ACC value is lower than BS -RF algorithm, but with higher AUC values. In the model, compared with the ACC index, AUC can make a comprehensive evaluation of a model, so the performance of reforest is the best.

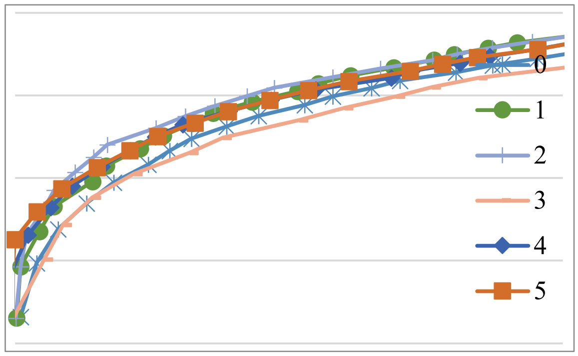

Figure 8: ROC curves of each model on the German dataset

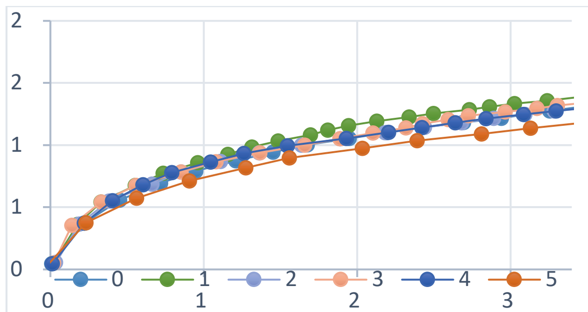

Figure 9: ROC curves of each model on the Australian dataset

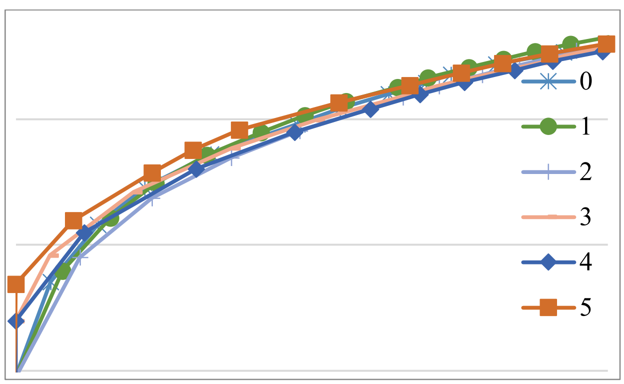

Figure 10: ROC curves of each model on the Kaggle dataset

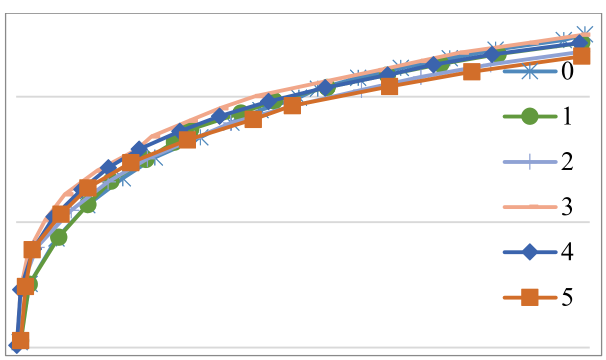

Figure 11: ROC curves of each model on the P2P dataset

Figure 8 to Figure 11 are the ROC curves of each model on the four data sets, and the AUC values corresponding to each model are listed below the curve. Combined with the evaluation indicators in Tables 2 to 5, the following points can be seen: (1) Among the commonly used credit risk prediction methods, Light and Boost are more effective than other methods. (2) The effect of forest is close to that of Light and Boost on each dataset. (3) The improved reforest is more stable and effective than the forest method. Except for the Australian dataset, which is comparable to the Boost result, it performs best in other datasets. Its AUC value is 1.13% higher than that of Light on average. 1.44% higher on average than Boost.

IV. Conclusion

In the process of innovative financial investment management, enterprises are bound to have different degrees of investment risks due to industry experience. Therefore, if an enterprise wants to pursue more economic benefits in its development and further increase its overall influence, it must establish reform measures related to modern innovative financial investment management to strengthen its ability to prevent and control risks. Based on the development prospects of enterprises, the government also needs to focus on the overall situation of enterprises' innovative financial investment. While building a talented team for enterprise innovative financial investment management, enterprises are required to do a good job in financial investment risk prevention and establish and improve the enterprise's innovative financial investment management system.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Funding Statement

There is no specific funding to support this research.

Reference

[1] Ali, S. , Bilal, H. , Ramsey, T. , & Yuefang, D. . (2022). From displacement to resettlement: the stakeholders strategies to mitigate impoverishment risks of the ghazi barotha hydropower project pakistan. Impact Assessment and Project Appraisal, 40(4), 305-319.
[2] Omalley, K. A. , Omonyêlé L. Adjognon, Davila, H. , Shin, M. H. , Sullivan, J. L. , & Engle, R. L. . (2022). Supporting veterans with dementia to remain in the community: strategies used in 12 veterans health administration programs. Home Health Care Services Quarterly, 41(2), 149-164.
[3] Lilja, M. , & Baaz, M. . (2022). New forms of power in a neoliberal era: artepolitics or the governing through non-governing. Journal of Political Power, 15(2), 189-201.
[4] Chen, L. . (2022). Christopher freeman: the soul mentor for guiding to write a new chapter in the new era. Innovation and Development, 12(1), 93-96.
[5] Ciara, B. . (2022). Immigrant irishwomen and maternity services in new york and boston, 1860–1911. Medical history(66-1).
[6] Shandu, S. , Gasura, E. , Mashingaidze, K. , & Derera, J. . (2022). Contribution of temperate germplasm to the performance of maize hybrids under stress and non-stress environments in south africa. South African Journal of Plant and Soil, 39(1), 66-76.
[7] Smith, E. M. . (2022). The global report on assistive technology: a new era in assistive technology. Assistive Technology, 34(3), 255-255.
[8] Cameron, J. . (2022). Soviet-american strategic arms limitation and the limits of co-operative competition. Diplomacy And Statecraft, 33(1), 111-132.
[9] Schmidt, A. T. . (2022). Evc and ata call for urgent reform to deliver new era of electric trucks in australia. Highway Engineering in Australia: Transport Infrastructure·Traffic Management·Intelligent Transport Systems(53-2).