Research on Risk Prediction Model of Green Financial Products Based on Machine Learning

Xu Zhang

Universiti Sains Malaysiae, Penang, 11600, Malaysia

Keywords: green finance, risk prediction, machine learning, model optimization, sustainability

Abstract: As an important means of promoting sustainable development, the effective management of green finance product risks is of great significance in promoting the coordinated development of the economy and the environment. Due to the high uncertainty and complexity of green financial products, traditional risk prediction methods have obvious limitations in dealing with these characteristics. With the rapid development of machine learning technology, its application in financial risk prediction shows strong potential and advantages. This paper focuses on the risk prediction problem of green financial products and constructs a systematic risk prediction model framework based on machine learning technology. Through data preprocessing, model selection and construction, as well as performance evaluation and optimization, this paper proposes a risk management scheme that can improve prediction accuracy and reliability. Through empirical analysis, the applicability and advantages of the proposed model in different scenarios are verified. The research in this paper not only provides a new technical path for the risk management of green financial products, but also provides a reference for the expansion of the application of machine learning in the field of sustainable finance.

1. Introduction

With global climate change and environmental problems becoming increasingly serious, green finance, as a key driver for sustainable development, has received widespread attention from Governments, businesses and academia[1]. Green finance achieves a balance between economic growth and environmental protection by providing financial support for sustainable development projects[2]. Risk management of green financial products, green financial products are mostly related to emerging technologies and long-term projects, resulting in high complexity and high uncertainty of their risks; traditional risk assessment and prediction methods show insufficient efficiency and lack of accuracy in the face of these complex characteristics; with the rapid expansion of the green financial market, the increase in the volume and dimensionality of the data has made it difficult for the traditional methods to deal with the large-scale, multi-dimensional dataset[3]. The increase in data volume and dimension makes it difficult for traditional methods to handle large-scale and multi-dimensional data sets.

Machine learning technology has gradually become an important tool in the field of financial risk prediction by virtue of its advantages in nonlinear modeling, high-dimensional data processing and automated learning[4]. Machine learning can not only efficiently mine potential laws from complex data, but also adapt to changes in the market environment through dynamic optimization, providing higher accuracy and reliability for risk prediction. At present, the research on the application of machine learning in risk prediction of green financial products is still in the primary stage, and the applicability, validity and interpretability of related models still need to be further explored.

The purpose of this paper is to study how to construct a risk prediction model for green financial products using machine learning technology, focusing on solving key issues such as data processing, model selection and optimization, and result interpretation[5]. Combined with the characteristics of green financial products, we construct a machine learning model framework suitable for this field to explore the applicability and advantages of different machine learning algorithms in risk prediction; validate the actual effect of the model through empirical analysis, and propose the direction of improvement and application prospects[6]. Through the above research, this paper expects to provide

technical support for the risk management of green financial products and contribute new research perspectives for the sound development of green finance.

2. Current Research Status of Risk Prediction for Green Financial Products

Green financial products mainly support projects related to sustainable development, clean energy, energy-saving and environmental protection technologies and ecological restoration[7]. These products are characterized by a long project cycle, high capital demand, low market adaptability and strong policy dependence, leading to a high degree of complexity and multi-dimensionality in their risk performance, including policy risk, market risk, technology risk and environmental risk[8]. Owing to the innovative attributes of the green finance sector, some products lack mature historical data support, further increasing the difficulty of risk management. Logistic Regression Model for Risk Prediction:

$$P(Y=1|X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n)}} \quad (1)$$

Traditional financial risk prediction methods mainly include multiple regression analysis, time series modeling and VaR (value at risk) methods[9]. These methods perform well in dealing with linear relationships in financial markets, but face many limitations in dealing with the complex nonlinear characteristics of green financial products[10]. Traditional methods are weak in handling high-dimensional data, making it difficult to effectively capture potential nonlinear relationships between risk factors. Green financial products often involve multiple domains, such as finance, environment and technology, and traditional methods are difficult to integrate multi-dimensional heterogeneous data, resulting in limited accuracy and efficiency of risk prediction, showed in Figure 1 :

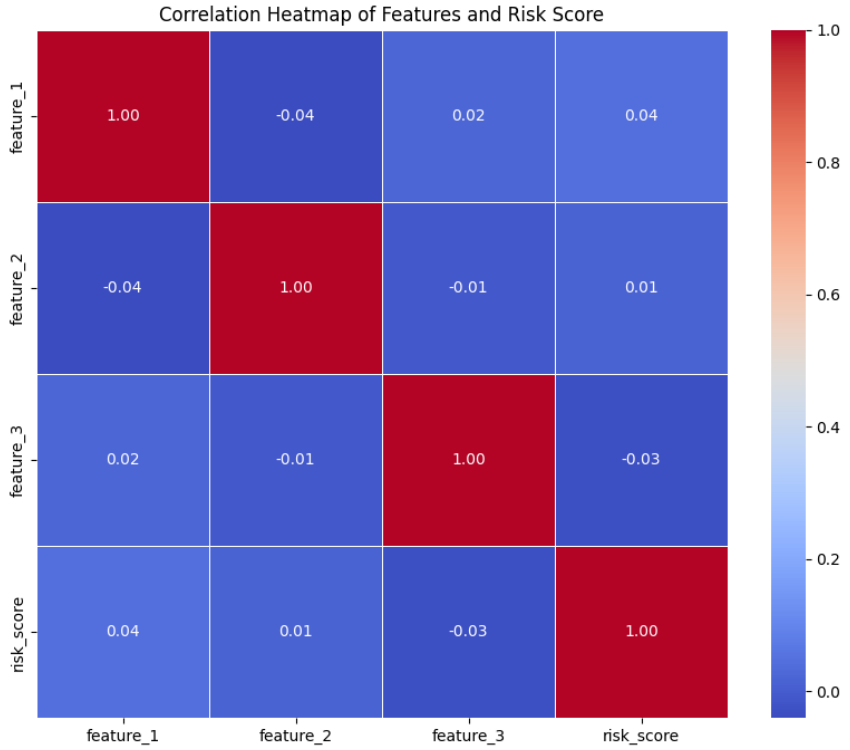

Figure 1 Pairplot of Features and Risk Score

With the rapid development of machine learning technology, its application in the financial field has achieved remarkable results, especially in risk prediction, which shows strong potential. Machine

learning algorithms Random Forest, Support Vector Machines and Neural Networks are able to deal with high-dimensional and diverse data, mine complex nonlinear relationships from them, and adapt to changes in the market environment through dynamic learning. These advantages make machine learning gradually become an important tool for risk prediction of green financial products, providing a brand new technical way to solve the limitations of traditional methods.

Research on risk prediction of green financial products has mainly focused on specific areas or single risk dimensions, and a systematic prediction framework has not yet been formed. The diversity and heterogeneity of green financial products put forward higher requirements for data acquisition and model design. How to improve the generalization ability, interpretability and operability of models is still a key challenge in research. The application of machine learning in risk prediction of green financial products is promising, but its effectiveness and practical applicability still need to be verified and improved through more empirical studies.

3. Machine Learning-based Risk Prediction Model Design for Green Financial Products

In order to effectively deal with the complexity and diversity of green financial product risks, machine learning-based methods are introduced to construct a systematic risk prediction model framework. Through data preprocessing and feature selection, model inputs are optimized to improve data quality and prediction efficiency; appropriate machine learning algorithms are selected, and models are designed based on the characteristics of green financial product risks to ensure accuracy and robustness in predictions. Model evaluation and performance optimization are conducted to validate the model's practicality and enhance its performance in real-world applications, thereby providing reliable technical support for green financial product risk management.

3.1. Data preprocessing and feature selection

Data preprocessing is the foundation of machine learning model construction, and the improvement of data quality directly determines the accuracy and stability of model prediction. In the risk prediction of green financial products, the data come from a variety of channels, including the financial market, policies and regulations, environmental monitoring, etc., which may have problems such as missing data, inconsistency or noise. It is necessary to ensure the completeness and consistency of the data through data cleaning techniques, missing value filling, outlier detection and processing, and noise data filtering.

Feature selection is an important part of optimizing model performance, aiming to extract the most relevant feature variables for risk prediction from multidimensional data. The feature variables of green financial products cover a wide range of aspects such as policy changes, market fluctuations, and technological innovations, etc. In order to reduce the impact of redundant data on the model, statistical methods such as correlation analysis and principal component analysis (PCA) can be used to screen key features. Algorithms based on feature importance ranking can also further enhance the scientificity and accuracy of feature selection. Random Forest Prediction Aggregation Formula:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(X) \tag{2}$$

Data normalization and standardization are common methods to deal with heterogeneous data, which can avoid the impact of eigenvalue magnitude differences on model training. In green financial risk prediction, different eigenvariables may have different magnitudes and ranges, and mapping the eigenvalues to uniform intervals through normalization or making them conform to the standard normal distribution through standardization can improve the convergence speed and prediction effect of the model. For time series data, it is also necessary to eliminate the effects of trend and seasonality through methods such as differencing or moving average.

Feature engineering is not limited to the screening and processing of existing data, but also includes the construction of new features. In green financial product risk prediction, composite features can be generated through domain knowledge or interaction term design, and the product of policy uncertainty index and market volatility can be used to capture the nonlinear impact of policy

changes on market risk. This feature construction method can further enhance the model's ability to perceive complex risk relationships and provide a more solid foundation for accurate prediction.

3.2. Model Selection and Construction

The selection of appropriate machine learning algorithms is crucial in the design process of risk prediction models for green financial products. Different algorithms have different advantages and applicable scenarios. For the complex nonlinear and high-dimensional data characteristics of green financial products, common model choices include decision trees, support vector machines (SVM), random forests, gradient boosted trees (GBDT), and deep learning models. Decision trees and random forests can effectively extract important features when dealing with high-dimensional feature data, while deep learning models can capture complex nonlinear relationships through multi-layer networks and are suitable for dealing with large amounts of unstructured data,showed in Figure 2 :

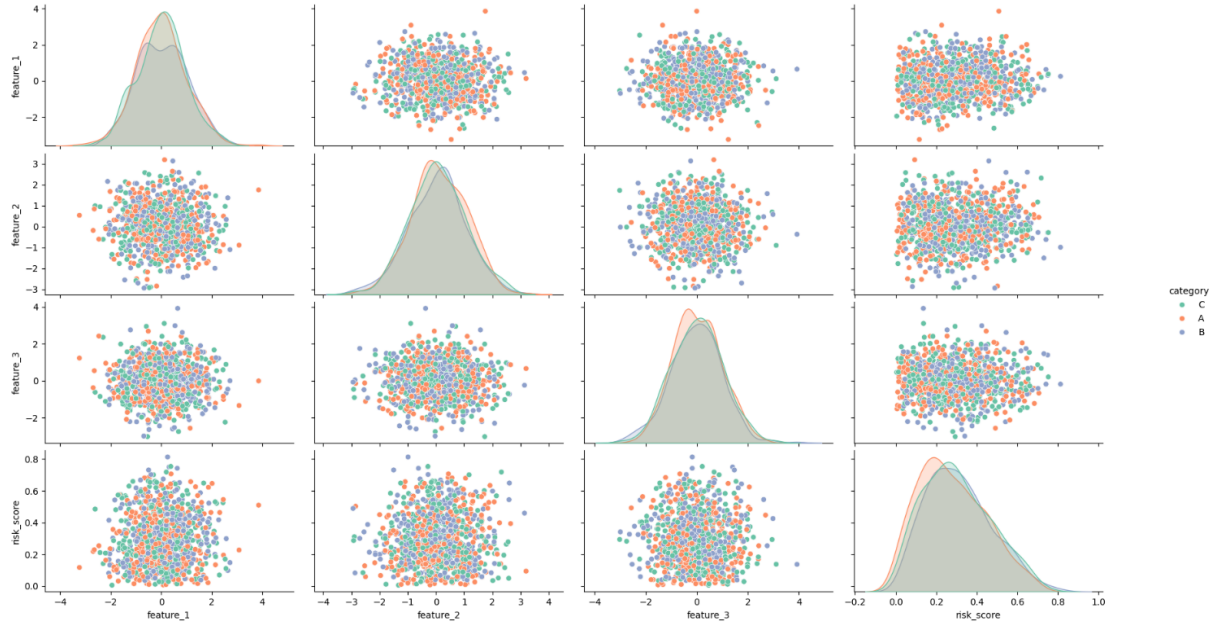

Figure 2 Correlation Heatmap of Features and Risk Score

Random forest and support vector machine are classical algorithms commonly used in risk prediction. Random forest has strong anti overfitting ability, and can effectively reduce the model error through the integration of multiple decision trees; support vector machine can find the optimal hyperplane in high-dimensional space, which is suitable for dealing with the risk classification problem with clear boundaries and fewer samples. For the risk prediction of green financial products, considering the possible noise and imbalance in the data, Random Forest is more suitable for its application and can provide more stable prediction results.

Deep learning models, Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM), show strong advantages in processing time-series data and large-scale data. For green financial products, factors such as policy changes, market fluctuations and technological advances may change over time, and deep learning models can capture the deep laws of these time-series data through their powerful feature extraction and adaptive learning capabilities. Deep learning models are able to realize automatic feature learning through backpropagation and multilayer structure without explicit feature engineering, thus reducing the bias of human intervention.

Appropriate model tuning needs to be carried out in the model construction process according to the specific risk characteristics of green financial products, including the selection of appropriate loss functions, optimization algorithms and regularization strategies. The choice of loss function directly affects the target of model training and the measurement of error, while the optimization algorithms Adam, SGD, etc., affect the convergence speed and stability of the model. The regularization methods L2 regularization or Dropout can effectively prevent the model from overfitting, thus improving the model's generalization ability on unknown data. Through multiple rounds of training and parameter

tuning, an efficient model suitable for risk prediction of green financial products is finally constructed.

3.3. Model Evaluation and Performance Comparison

Model evaluation is a key part of the machine learning process, which ensures that the constructed risk prediction model can provide stable and reliable prediction results in practical applications. In green financial product risk prediction, due to the imbalance and complexity of data, a single evaluation metric cannot fully reflect the performance of the model. Commonly used evaluation metrics include accuracy, precision, recall, F1-score, AUC (area under the curve), and mean square error (MSE), etc. These metrics can measure the performance of the model in risk prediction from different dimensions. Accuracy measures the correctness of the overall prediction, while F1-score is able to combine precision and recall, which is especially important for unbalanced datasets.

In addition to traditional evaluation metrics, Cross-Validation is also a commonly used method for model evaluation, especially when the data set is limited. By dividing the data into multiple subsets, Cross-Validation can effectively avoid overfitting and assess the performance of the model on different data subsets, thus reflecting the generalization ability of the model more accurately. For risk prediction of green financial products, cross-validation helps to test the stability and accuracy of the model in the face of different types of risks, ensuring that the model maintains good prediction performance in diverse real-world scenarios.

In order to compare the performance of different models, comparative experiments on the same dataset are required. These experiments not only help to select the best model, but also reveal the strengths and weaknesses of different algorithms in risk prediction. While deep learning models are able to handle complex nonlinear data, they may require more computational resources and training time; whereas decision tree models perform better in terms of computational efficiency and interpretability and are more sensitive to noisy data. By comparing multiple models, it is possible to clarify which model is most suitable for risk prediction of green financial products and ensure that the prediction results strike the best balance between accuracy, stability and computational efficiency.

In practice, the performance of a model does not only depend on the algorithm itself, but is also closely related to the quality of the data, feature selection and the model tuning process. The improvement of model performance often requires continuous iteration and optimization. In green financial product risk prediction, in addition to conventional assessment methods, the model should be regularly adjusted and optimized according to the special characteristics of green financial products, policy changes and market fluctuations. Through long-term monitoring and tuning, it is ensured that the constructed risk prediction model always adapts to the dynamic changes in the green financial market, and has high practicality and reliability.

4. Model Optimization and Application Prospects

Model optimization is an important tool to improve the performance of machine learning models, especially in green financial product risk prediction, where model optimization is particularly critical due to the dynamics and complexity of the data. The hyperparameters of the model need to be systematically adjusted through tuning techniques, Grid Search and Random Search. These hyperparameters include learning rate, tree depth, regularization coefficients, etc. The purpose of tuning is to find the most suitable combination of parameters for the current dataset in order to improve the prediction accuracy of the model. Integrated learning methods (Random Forest, XGBoost) are also often used to enhance the generalization ability of the model, by combining the prediction results of multiple models to reduce the overfitting or misclassification that may occur in a single model.

Feature engineering is also an important part of model optimization. With the continuous development of the green financial products field, emerging risk factors will continue to emerge, the traditional feature set may not be able to fully capture these changes, and it is necessary to continuously update the feature set based on domain knowledge and data changes. Through the construction and optimization of new features, the interaction of policy, market, technology and other dimensions can be better captured to improve the adaptability and prediction accuracy of the model.

With the help of deep learning technology, automatic feature learning can reduce manual intervention and improve the efficiency and accuracy of feature selection, thus further enhancing the model performance.

In the application of green financial product risk prediction, the model not only needs to have high predictive ability, but also needs to have good interpretability, so as to facilitate financial decision makers to make reasonable judgment in risk management. Interpretability has gradually become the focus of research. For complex machine learning models, especially deep learning and integration models, methods such as feature importance assessment and locally interpretable models (LIME) can be used to help analyze the reasons for model predictions. These methods can effectively reveal the decision-making process of the model, enhance the transparency of the model, and increase the trust of decision makers in the model results, so that the model can be more widely supported and adopted in practical applications.

With the rapid development of the green financial market, the application prospect of risk prediction of green financial products is very broad. Through the introduction of machine learning technology, the risk prediction of green financial products can not only improve the prediction accuracy, but also reflect the changes in the market and policies in real time, helping decision makers to develop more flexible risk management strategies. Green finance, as an important tool to support sustainable development, has an increasing status in the global financial system, and the application of prediction models will further promote the market-oriented development of green financial products. With the advancement of technology and the continuous accumulation of data, the risk prediction models of green financial products will play an increasingly important role in quantitative analysis, policy formulation, investment decision-making, etc., and help the construction of a global sustainable financial ecosystem.

5. Conclusion

The green financial product risk prediction model based on machine learning has significant application value and development potential. Through reasonable data preprocessing and feature selection, high-quality inputs can be provided for the model to enhance the prediction accuracy and stability of the model. In the process of model selection and construction, machine learning algorithms that are adaptable and capable of handling complex nonlinear data, random forest, support vector machine and deep learning models are used to provide strong technical support for risk prediction. The validity and reliability of the selected models in practical applications are ensured through multi-dimensional model evaluation and performance comparison, and the generalization ability of the models is optimized by means of cross-validation and other means.

The prospect of applying risk prediction models for green financial products still faces many challenges and opportunities. Model optimization not only requires precise hyper-parameter tuning and feature engineering, but also constantly adapts to the new changes and demands in the field of green finance. With the continuous development of the green financial market, the interpretability of the model will become a key factor for its wide application. By improving the transparency and explanatory ability of the model, it can enhance the trust of decision makers and provide more scientific and precise support for the risk management of green financial products. With the continuous development of big data technology and machine learning methods, green financial product risk prediction models will gradually become more intelligent and efficient. In the future, risk prediction in the field of green finance will not only be limited to traditional financial risks, but will also involve multi-dimensional factors such as environment, society and governance, promoting the sustainable development and marketization of green financial products.

References

[1] Hu Y, Liu H .Prediction of Passenger Flow on the Highway Based on the Least Square Suppoert Vector Machine[J].Transport, 2011, 26(2):197-203.DOI:10.3846/16484142.2011.593121.

[2] Lajevardi S A, Kargari M, Daneshpour M S A M .Hypertension Risk Prediction Based on SNPs by Machine Learning Models[J].Current Bioinformatics, 2023, 18(1):55-62.

[3] Lemming J .Financial risks for green electricity investors and producers in a tradable green certificate market[J].Energy Policy, 2003, 31(1):21-32.DOI:10.1016/S0301-4215(02)00114-3.

[4] Xie X, Huo J, Qi G, et al. Green Process Innovation and Financial Performance in Emerging Economies: Moderating Effects of Absorptive Capacity and Green Subsidies[J].IEEE Transactions on Engineering Management, 2016, 63(1):1-12.DOI:10.1109/TEM.2015.2507585.

[5] Liu R, Wang D, Zhang L, et al. Can green financial development promote regional ecological efficiency? A case study of China[J].Natural Hazards, 2019, 95(1-2):1-17.DOI:10.1007/s11069-018-3502-x.

[6] Murtaza M R, Hongzhong F, Hashmi S H, et al.The nexus of green energy, financial inclusion, government defense financing, and socioeconomic sustainability across diverse income-level groups: a global indication[J].Clean Technologies and Environmental Policy, 2024, 26(10):3391-3410. DOI:10.1007/s10098-024-02806-1.

[7] Olson E L .Perspective: The Green Innovation Value Chain: A Tool for Evaluating the Diffusion Prospects of Green Products [J].Journal of Product Innovation Management, 2013, 30(4):782-793. DOI:10.1111/jpim.12022.

[8] Xie G, Yue W, Liu W, et al. Risk based selection of cleaner products in a Green supply chain[J]. Pacific Journal of Optimization, 2012, 8(3):473-484.DOI:10.1112/S0025579311002087.

[9] Galicia, Isidro. Total Lean and Green Eco-Advantage "Integrated" Strategy.[J].Manufacturing Engineering, 2009.

[10] Bajaj K .Green Gadgets to Look out For[J].Technology review, 2009, 14(11):24-25.