WEP Warwick Evans Publishing

International Journal of Global Economics and Management
ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijgem.v4n1.19
Journal homepage: https://wepub.org/index.php/IJGEM/index

Accuracy of Green Bond Issuance Predictor

Xiangyu Guo [1], Jinye Chen [2], Gexuan Ren [1, *]

[1] College of Mathematics and Physics, Xi'an Jiaotong-LiverpoolUniversity, Suzhou, China
[2] College of International Business Suzhou, Xi'an Jiaotong-LiverpoolUniversity, Suzhou, China
*Corresponding Author: rengexuan0723@126.com

ABSTRACT

Climate change is affecting the development of many industries in different aspects. These impacted enterprises transform into sustainable enterprises to avoid the risks, and by doing so they enter into the green bond market. The current literature provides effective reference indicators for participants in the green bond market. These indicators illustrate the funding size of the green bonds in different dimensions to the participants. As for the improvement of the policies about environmental protection there also emerge some new indicators such as ESG score. Besides, machine learning is an accurate and effective tool in many fields, and some researchers have established a model for predicting the issuance of green bonds but have not involved the new indicators in the past. In this paper, on the one hand, we discuss the new indicator, ESG scores, and how it affects the funding size of the green bonds, on the other hand, we add this new indicator and the old indicators into four machine learning models to compare the accuracy of predicting the issuance of green bonds of these four models. In these four models, the Random Forest Regressor and LGBM Regressor are the best models on average. The former has the best performance of accuracy but needs much more time than the latter. On the opposite, the latter is the most efficient model among all but is the second most accurate. Besides, other models have the best numerical measurements in different dimensions which means we could use different models depending on different situations. Choosing the proper model for the specific situation can optimize the benefit of the green bond market participant.

KEYWORDS

ESG investing; Green Bonds funding size; Machine learning

1. INTRODUCTION

With increasing the awareness of protecting the environment in every country green bonds burgeoning into fixed-income securities as a new concept of a financial instrument designed to support the financing of environmentally friendly projects. Issue green bonds are aiming to obtain further financing or refinancing capital for causes that contribute to the improvement of the natural environment (Anna, 2018). Green bonds have experienced significant growth in the global market in recent years. This growth not only reflects investors' focus on sustainable development but also demonstrates the important role of green finance in promoting environmentally friendly projects (CBI, 2021; ICMA, 2022). Finance is the core department of a company and some companies issue green bonds to achieve the company's green and sustainable development. In doing so, how many specific quantities of green bonds should be published becomes an emerging problem. According to Barua and Chiesa (2019), the supply of green bonds is supply-pushed not demand-pulled. A growing body of research suggests that green finance plays an important role in promoting sustainable economic development and combating climate change. For example, Kölbel et al. (2020) state that the expansion of the green bond market has not only improved the ability to finance environmental projects but also promoted environmental awareness among investors. Other studies have shown that the use of green financial instruments can effectively reduce the carbon footprint of firms and improve resource efficiency (Busch et al., 2016). Therefore, understanding the drivers of green bond issuance is crucial to promoting the development of green finance (Ehlers & Packer, 2017).

ESG is the abbreviation of the "Environment", "Social" and "Governance". It is a standard used to evaluate a company's performance in environmental protection, social responsibility, and corporate governance this three dimensions. ESG is not only a part of sustainable development for enterprises, but also an important indicator for investors to evaluate long-term risks and opportunities for companies. So far, there is plenty of research that has illustrated ESG has more or less impact ona company's performance. As Velte (2017) expounded in his paper: During the business year 2010-2014, an analysis covering 412 enterprises from three dimensions of ESG confirmed the three dimensions all had a positive impact on accounting-based financial performance. Furthermore, the score of governance performance shows the strongest impact among the three dimensions. The ESG system can protect enterprises by directing and managing the microscopic behavior of enterprises from adverse risks impact and reducing negative reactions in the environment and in the enterprise's social perspective (Wang& Wang, 2022). Combining current development trends and the consciousness of the environment sustainable development in society, some investors prefer to add ESG scores to their investment criteria to avoid emerging risks as much as possible to achieve maximum returns. Therefore, most enterprises also choose to transform into "green enterprises" or sustainable development enterprises due to this trend by issuing green bonds to raise funds which can used to support environmental protection and develop some sustainable development projects.

In Barua and Chiesa's (2019) research, they defined a broad set of impact factors of green bond publishment in three key categories: bond characteristics, issuer characteristics, and market characteristics. Better ESG practices will encourage companies to issue more green bonds (Wang& Wang, 2022). It follows that the issuance of green bonds is influenced by various factors, including ESG (Environmental, Social, and Governance) scores, bond characteristics, issuer characteristics, and market characteristics. Understanding how these factors affect the issuance of green bonds is crucial for investors, issuers, and policymakers. We use these four main factors or characteristics to model the funding size of green bonds through machine learning rather than traditional statistical methods in our research. Although traditional statistical methods have been used to analyze the impact of these factors, they are constrained when confronted with intricate relationships, machine learning provides a more precise and complex analytical tool that can reveal nonlinear relationships and interactions hidden in data (Baker et al., 2018). In this paper, we compare the accuracy of four different machine learning models in predicting the issuance of corporate green bonds. The models were established based on ESG scores, bond characteristics, issuer characteristics, and market characteristics, aiming to provide empirical data support and decision-making references for participants in the green bond market.

However, to fully understand the drivers of green bond issuance, the impact of ESG scores, bond characteristics, issuer characteristics, and market characteristics must be examined in depth. Existing literature suggests that these factors have a significant impact on the dynamics and complexity of the green bond market (Zerbib, 2019; Hachenberg & Schiereck, 2018; Larcker & Watts, 2022). With the development of big data and analytics, machine learning models have shown great potential in revealing these complex relationships (Baker et al., 2018).

Our research has practical reference significance for participants in the green bond market. First, we explored the nature of the impact of various factors and divided them into ESG scores, bond characteristics, issuer characteristics, and market characteristics then discussed these four characteristics' impact on the funding size of the green bond market separately and revealed different associations between green bond markets and these four characteristics. The second is to compare the machine learning techniques that can most accurately predict the issuance volume of China's green bond market. We have considered several algorithms, including Linear Regression, Random Forest Regressor, XGB Regressor, and LGBM Regressor.By comparing the prediction accuracy of different machine learning models, some participants in the green bond market can better utilize the appropriate model to make optimal decisions.

The paper is structured as follows: Section 2 discusses the relevant literature and theoretical background; Section 3 outlines the materials and methodology used; Section 4 presents the result of our research and discussion on the practical significance of the research; And finally, section 5 provides policy implications and conclusions based on our research.

2. LITERATURE REVIEW

ESG scores, as a measure of a firm's environmental, social and governance performance, have become an important reference for investors to assess a firm's sustainability capabilities.Zerbib (2019) finds that firms with high ESG scores are able to realize lower issuance costs and higher market acceptance when issuing green bonds. This finding is in line with Hachenberg and Schiereck (2018) who, in their analysis of the German market, find that firms with high ESG scores are able to enjoy lower yields when issuing green bonds. This is supported by Flammer's (2021) study, which emphasizes the importance of ESG scores in investor decision-making. However, Baker et al. (2018) argue that despite the importance of ESG scores, the market's sensitivity to their impact may vary by region and industry. In the U.S. market, the response to ESG scores varies widely across sectors, with sectors such as technology and energy being more sensitive to ESG scores.Larcker and Watts (2022) use a machine learning model to find a significant nonlinear relationship between the impact of ESG scores on green bond issuance, suggesting that simple linear models may not be sufficient to capture this complexity and that more sophisticated models to understand the impact of ESG scores. Therefore, future research should introduce more sophisticated machine learning models that can handle large-scale data, capture complex nonlinear relationships, and provide more comprehensive analytical results, thus filling the gaps in existing research.

Characteristics such as maturity time, coupon rate, and credit rating are important factors affecting the issuance of green bonds.Tang and Zhang (2020) find through regression analysis that longer maturity time and higher credit ratings are positively correlated with the issuance of green bonds, which is in line with Flammer's (2021) finding in the international market. The empirical study by Hachenberg and This is supported by the empirical study of Schiereck (2018), who finds that green bonds with high credit ratings and long maturities are more attractive.Wang et al. (2019), through a study of machine-learning models in the Chinese market, find that higher rated bonds and longer maturities tend to attract lower risk premiums, which reduces the financing costs of issuers. Larger issue sizes also benefit from economies of scale, which further reduces the risk premium. These findings are supported by a multivariate statistical regression analysis that integrates macro and microeconomic factors to provide a comprehensive understanding of green bond pricing dynamics. It is important to note, however, that market responses to these characteristics can vary over time, are highly time-dependent, and may fluctuate with the economic cycle. Meanwhile, Larcker and Watts (2022) emphasize the importance of the impact of market conditions on these characteristics by using a machine learning model to reveal a non-linear relationship between the impact of different bond characteristics on issuance. Thus, while bond characteristics play an important role in increasing the market acceptance of green bonds, their impact may vary in different market environments and need to be analyzed comprehensively in the context of market conditions. However, while existing machine learning models by scholars provide important evidence for understanding the impact of bond characteristics on green bond issuance, these models ignore the complex interactions and non-linear relationships between characteristics. This paper aims to fill this research gap, hoping to enhance the timeliness and usefulness of the study by introducing more sophisticated machine learning models to more comprehensively analyze the complex relationships among these bond features and to process high-dimensional data to provide more fine-grained analysis results.

Issuer characteristics (including firm size, industry category, and financial health) also have a significant impact on green bond issuance.Flammer (2021) shows that large firms and firms with good financial health are more likely to issue green bonds. This is supported by Hachenberg and Schiereck (2018) who note that issuer characteristics play a key role in green bond pricing. Issuers with substantial assets and higher financial stability are perceived as less risky, which may lead to more favorable issuance conditions. In the Chinese market, Wang et al. (2019) used a machine learning model to analyze a large dataset, and their study suggests that state-owned enterprises (SOEs) and firms with higher returns on net worth are more likely to successfully issue green bonds, reflecting investor confidence in their credibility and long-term viability. This study identifies patterns and correlations that may be missed by traditional statistical methods. By considering multiple variables simultaneously, these models can more accurately predict issuance success. However, also in the Chinese market, Tang and Zhang (2020) found through machine learning analysis that the effects of firm size and industry category on green bond issuance exhibit a complex nonlinear relationship, suggesting that traditional statistical methods may not be able to adequately account for such effects.Baker et al. (2018) argued that small firms in the green bond market may face more issuance challenges, especially in the early stages.Larcker and Watts (2022) also point out that while large firms dominate the market, SMEs are gradually starting to gain market share and attract investors through innovative financing methods. Moreover, their study reveals a complex non-linear relationship between firm size and industry type on green bond issuance, further emphasizing the importance of segmentation and industry analysis.

Market characteristics such as macroeconomic environment, interest rate level, market demand, and policy support are also determinants of green bond issuance Larcker and Watts (2022) find that a low interest rate environment and high market demand significantly contribute to green bond issuance. This finding is consistent with Tang and Zhang's (2020) findings in the Chinese market, where Tang and Zhang (2020) point out that market demand is an important factor influencing the issuance of green bonds Hachenberg and Schiereck (2018) further point out that uncertainty and volatility in the market can also affect the decision to issue green bonds , especially during periods of high economic volatility. In the Chinese market, Wu et al. (2022) and others use PSM and DID models to accurately predict these effects by analyzing a large dataset that includes historical interest rate movements and bond issuance data, thus finding that lower interest rates tend to reduce the cost of capital for green bond issuers. In addition, Zerbib (2019) emphasizes the geographic and temporal variability of market characteristics, suggesting that the impact of market characteristics may vary by time and location, which complicates forecasting. In addition, Flammer (2021) further notes that the green bond market as a whole exhibits a high degree of resilience despite changing market conditions, which implies that investor confidence in green bonds is more solid. In addition, government policies and regulatory support are key to driving green bond issuance.Wu et al. (2022) use PSM and DID models to help assess the effectiveness of market policies by examining the impact of these policies on market behavior and issuance. The analysis includes an assessment of the impact of subsidies, tax incentives, and mandatory green certification requirements to provide a more detailed analysis of the impact of policy support on green bond issuance. However, while learning models proposed by existing scholars provide different evidence for the study, these traditional statistical models have limitations in dealing with high-dimensional and non-linear data. The aim of this paper is to explore the long-term impact on green bond issuance by introducing a novel learning model that can effectively integrate multiple market characteristics, deal with complex nonlinear relationships, and improve the robustness and reliability of predictions.

3. METHODOLOGY

3.1. Data

In this study, we utilized data collected from 99 green-labeled bonds issued through the Wind data terminal over a one-year period in 2023. In order to investigate the accuracy of machine learning in predicting the issuance volume of green bonds, we defined a set of factors that may influence the issuance volume. Based on specific research on green bonds and financial reports, we selected potential variables from four categories: (1) corporate characteristics; (2) issuer characteristics; (3) market characteristics; and (4) ESG (Environmental, Social, and Governance), and expanded these four categories into 13 independent variables. Table 1 represents all 13 variables and their fundamental principles, which we use to depict the basic functional relationships as follows:

$$\text{Issue amount} = f\,(\text{Bond Characteristics, Issuer Characteristics, Market Characteristics}) \quad (1)$$

We expand function (1) and replace it with all variables:

$$\text{Issue amount} = f\,(\text{Term, Coupon rate, ROA, ROE, Debt\_to\_asset, Cash turn ratio, wgsd\_asset, Credit\_rating, Industry\_gics, ESG score, E\_score, S\_score, G\_score}) \quad (2)$$

Table 1. Factors/variables used

Determinant categories	Variables considered	Variable description	Rationale
Bond Characteristics	Term	The issuance term of a bond in years	The length of time between the date of issue and the maturity date of the bond
	Coupon rate	% rate of coupon interest per annum	Cost of financing
Issuer Characteristics	ROA(Return on Assets)	Return on Assets (Annual %)	Firm profitability
	ROE(Return on equity)	Return on equity (Annual %)	Firm profitability and asset utilization efficiency
	Debt_to_asset	% rate of total bond liabilities to total assets	Firm financial structure
	Cash turn ratio	% rate of cash flow to its average cash balance	Cash management ability and efficiency
	wgsd_asset(Total asset)	Value of total assets in CNY	Firm size
	Credit_rating	Numerical codes for rating either by agency	Creditworthiness and risk profile of the issuer
Market Characteristics	Industry_gics	Classification of industries	type of industry
ESG	ESG score	The comprehensive score of environment, society and governance	Corporate performance in sustainability and social responsibility
	E_score	Environment score	Performance on environmental issues
	S_score	Society score	Performance in social responsibility
	G_score	Governance score	The performance of internal governance of enterprises

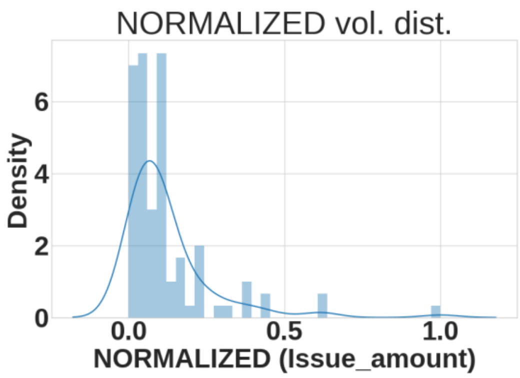

Figure 1. Distribution of normalized values ranging between 0 and 1

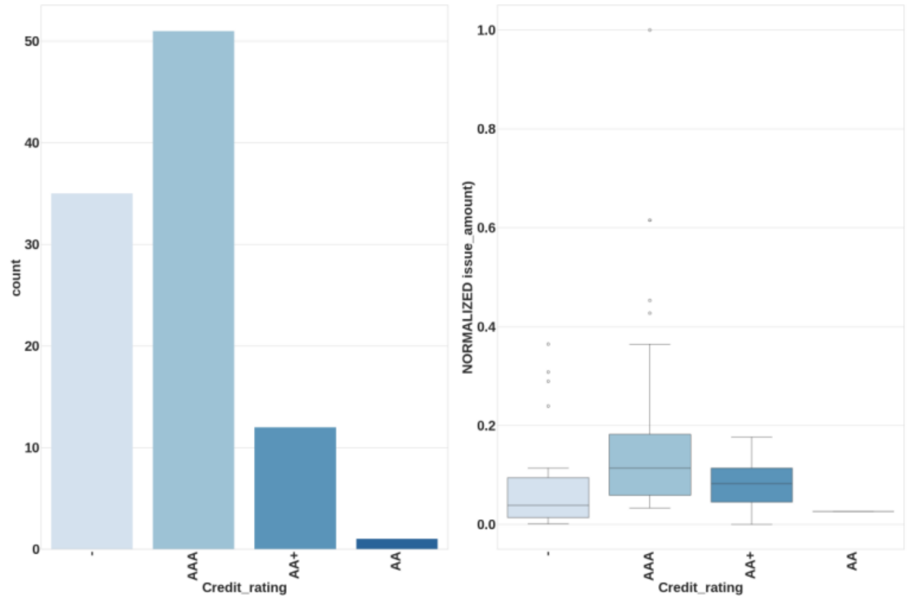

Figure 2. Quantity bar chart of credit ratio and box chart of normalized issue amount distribution

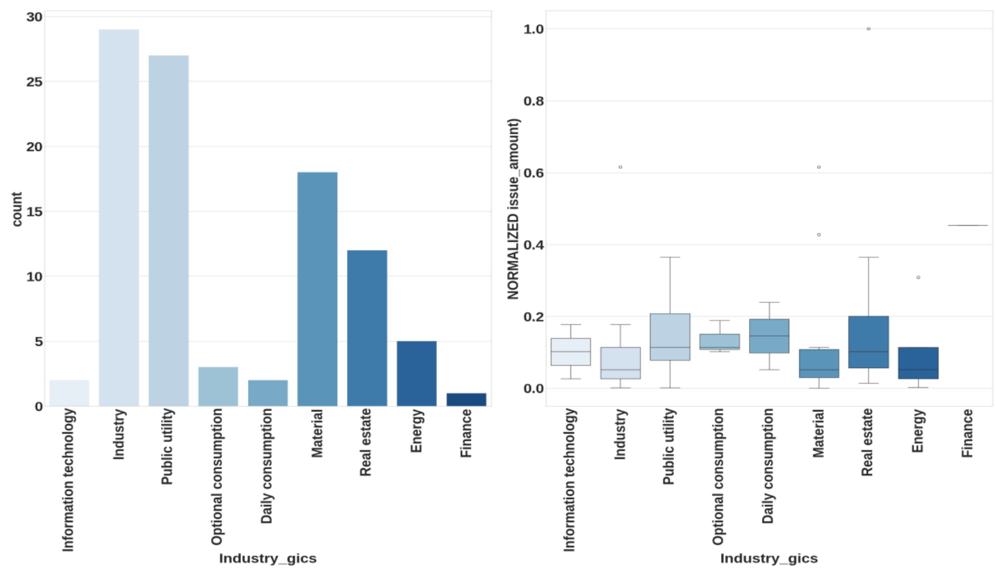

Figure 3. Quantity bar chart of industry_gics and box chart of normalized issue amount distribution

Before using machine learning algorithms for the measurements, we performed preliminary processing of the data. After excluding outliers and dealing with missing values, we performed feature engineering and normalized the issue count by converting each value of the issue count to a range of 0 to 1 and naming the normalized dependent variable as NORMALIZED(Issue amount). Fig. 1 represents the normalized density distribution of green bond issuance, from which it can be seen that the normalized issuance of green bonds is all positive and positively skewed in the distribution. Among the above characteristic variables, credit rating and Industry_gics are discrete variables due to the specificity of their measurement data, and their quantitative relationship and relationship with the normalized issuance volume can be seen in Fig. 2 and Fig. 3. In Fig. 2, credit ratings are divided into four classes -, AAA, AA+, AA, with the largest number of AAA ratings and the smallest number of AA ratings; in Fig. 3, Industry_gics are divided into information technology, industry, public utility, optional consumption, daily consumption, material, and other categories. In Fig. 3, Industry_gics is divided into nine industries: information technology, industry, public utility, optional consumption, daily consumption, material, real estate, energy, and finance, where industry and public utility have the largest number of data and finance has the smallest number. In the end, we selected eleven continuous variables to explore the correlations, and the statistical charts of the basic characteristics of all the remaining continuous variables are shown in Table 2. The heat map of correlation is shown in Fig.4.The correlation coefficients show that ESG score has a strong correlation with the number of issues.

Table 2. Descriptive statistical

	ESG_Score	E_score	S_score	G_score	ROE	ROA	Debt_to_asset	Cash_turn_ratio	wgsd_assets	Term	Time_to_maturity	Issue_amount	Coupon_rate
count	99	99	99	99	99	99	99	99	9.90E+01	99	99	9.90E+01	99
mean	6.799192	4.368384	5.179495	7.078182	3.138541	2.205454	64.05771	794.663752	2.69E+11	51.157275	2.255016	1.08E+09	3.753333
std	0.958605	2.311128	1.993336	0.992314	7.03309	2.237684	12.405854	714.26813	4.70E+11	31.074415	3.539375	1.19E+09	1.153977
min	4.78	0	0.77	4.97	-37.686	-1.825	24.2042	17.8216	9.62E+09	0.4098	0	9.00E+07	2.26
25%	6.05	2.57	3.78	6.245	1.3059	0.97875	56.3214	297.15605	4.03E+10	24.9098	0.716438	5.00E+08	3.005
50%	6.82	4.47	5.08	6.93	3.9869	1.931	63.5179	554.4414	7.64E+10	50.4098	1.421918	8.00E+08	3.4
75%	7.65	6.23	6.655	7.69	6.0991	3.1377	72.3785	1019.02715	2.83E+11	74.9098	2.142466	1.06E+09	4.22
max	8.71	8.47	9.47	9.79	21.7643	13.7624	91.5232	3344.6433	2.72E+12	132.4098	20.69863	8.07E+09	7.5

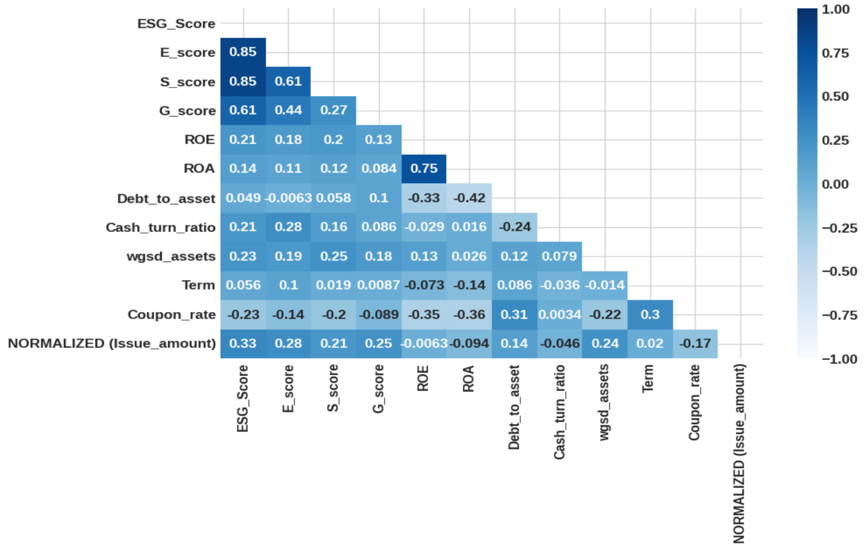

Figure 4. Correlated heat maps of variables

Before building the model, we employed the nested cross-validation method to ensure the stability of the model and select the optimal parameters. We performed one-hot encoding on the two categorical variables, Credit_rating and Industry_gics, to convert them into numerical variables and added them to the dataset, which was then split into multiple subsets for cross-validation testing. Ultimately, we obtained the parameter initialization from both internal and external cross-validation to use in model training and evaluation. In this study, we will separately use four machine learning algorithms, namely linear regression, Random Forest, Extreme gradient boosting (XGB) algorithm, and Light gradient boosting machine (LGBM), to test their accuracy in predicting the issuance volume.

3.2. Linear Regression

Linear regression is a classic machine learning algorithm that seeks to find the best linear relationship by fitting and training a simple and intuitive linear regression model. This algorithm is suitable for analyzing variables with simple linear relationships, and it offers strengths such as strong interpretability, simplicity in computation, and fast processing speed. Conversely, its limitations include the inability to handle non-linear relationships and susceptibility to the influence of outliers. The function of the linear regression model is as follows:

$$Y=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\ldots+\beta_n x_n. \tag{3}$$

Where Y represents the dependent variable (target variable), x1, x2, …, xn represent the independent variables (features), β0 is the intercept, and β1, β2, …, βn are the coefficients of the independent variables.

3.3. Random Forest

Random Forest is a commonly used ensemble learning algorithm, proposed by Breiman in 2001. It predicts samples by constructing multiple decision trees and ultimately derives predictions by averaging the results. The main advantage of Random Forest lies in its ability to resist overfitting. It achieves this by using the Bootstrap sampling method to obtain random feature subsets of the training data and then training multiple decision trees in different feature subspaces. After training, the predicted value of green bond issuance can be represented as:

$$\hat{y}=\frac{1}{N}\sum_{i=1}^{N} T_i(x) \quad (4)$$

Where T_i represents the i-th random tree learner, and x is the input feature vector.

3.4. Extreme Gradient Boosting (XGBoost)

Extreme gradient boosting (XGBoost) is an efficient machine learning algorithm proposed by Chen and Guestrin in 2016. It is based on decision trees and gradually improves the predictions of the target variable using gradient boosting while controlling model complexity through regularization techniques to enhance prediction accuracy. The core concept of Extreme gradient boosting (XGBoost) is to construct each tree in a depth-first manner, with each tree being trained on the residuals of the previous tree model. Additionally, XGBoost introduces a pruning technique to prevent overfitting by eliminating nodes with minimal contribution to the model. The prediction values are represented as follows:

$$\hat{y}_i=\sum_{k=1}^{K} f_k(x_i) \quad (5)$$

Where f_k represents the k-th tree model, x_i is the feature vector of sample i, and K is the total number of tree models.

3.5. Light Gradient Boosting Machine (LGBM)

LightGBM (Light Gradient Boosting Machine) is an efficient gradient boosting machine learning algorithm developed by Ke et al. in 2017. It is a decision tree learning model based on the Histogram algorithm, employing a Leaf-wise strategy to grow trees and selecting the best-split point at the current leaf node to minimize the loss at each node. LightGBM trains iteratively, optimizing the model in each iteration using the residuals from previous iterations. It utilizes pruning techniques to control model overfitting. LGBMRegressor algorithm performs exceptionally well in large-scale data and high-dimensional feature scenarios, exhibiting efficiency, accuracy, and robustness. It is a powerful tool for handling complex data problems. The LightGBM model can be represented as follows:

$$\hat{y}=\sum_{j=1}^{M} T_j(x) \quad (6)$$

Where $T_j$denotes the jth regression tree, x is the input feature vector, and M is the total number of trees. Each tree T_j is trained based on the residual from the previous iteration, continuously improving the model’s performance and prediction accuracy.

3.6. Performance of Models

We used three coefficients of measurement bias for the comparison of the accuracy of the predicted value of green bond issuance, namely the root means square error (RMSE), mean absolute error (MAE) and the coefficient of determination (R2). The functional expressions of these metrics are shown below:

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{Y_i}-Y_i)^2} \quad (7)$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i-Y_i\right) \quad (8)$$

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{Y}_i - \overline{Y}_i)^2}{\sum_{i=1}^{N}(Y_i - \overline{Y}_i)^2} \quad (9)$$

$\hat{Y}_i$, Y_i, and $\overline{Y}_i$ are expressed as the predicted, observed, and mean of the observations, respectively.

4. RESULTS AND DISCUSSION

4.1. Model Performance Evaluation and Comparison

Fig.s 5, 6, 7, and 8 show the degree of proximity of the predicted values to the true values of the four machine learning algorithms, respectively, and the accuracy of the model's predicted values is higher when the data points are closer to the regression line. As can be seen from the figure, the data points of the linear regression model and the XGB model are more dispersed, and the accuracy of RF and LGBM is relatively high.

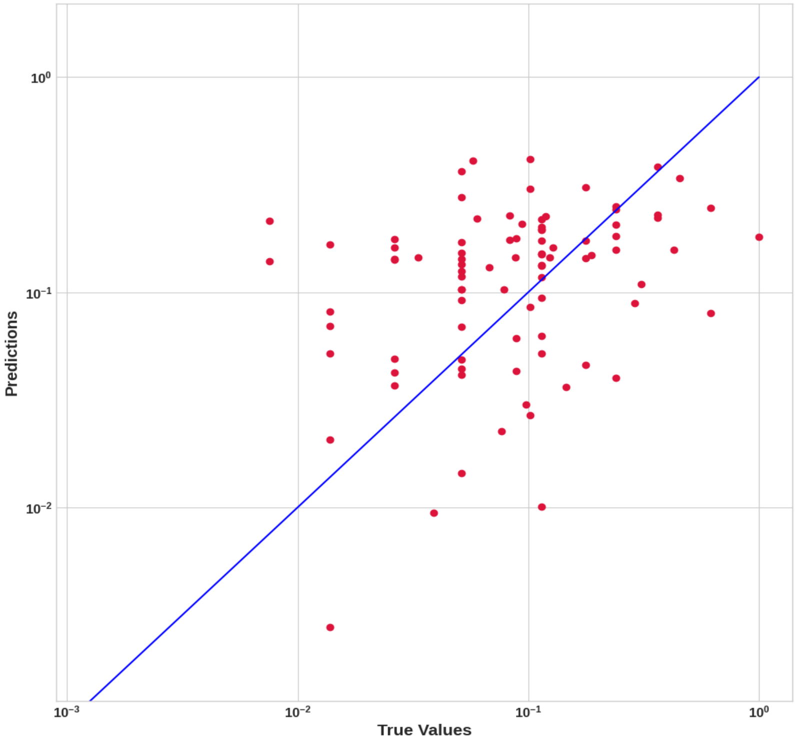

Figure 5. Scatter plot of linear regression model

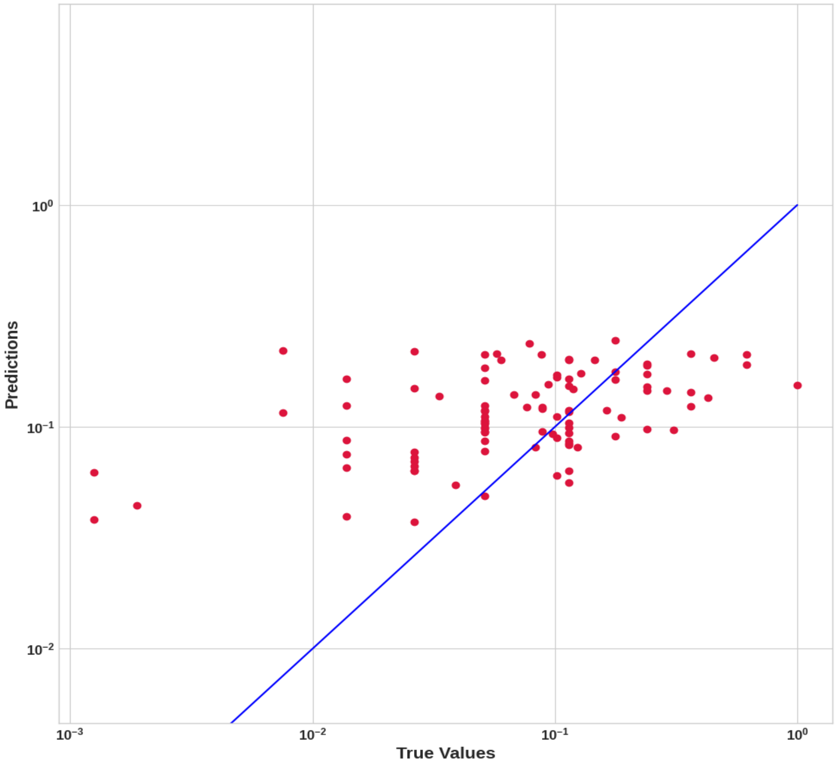

Figure 6. Random forest model scatter plot

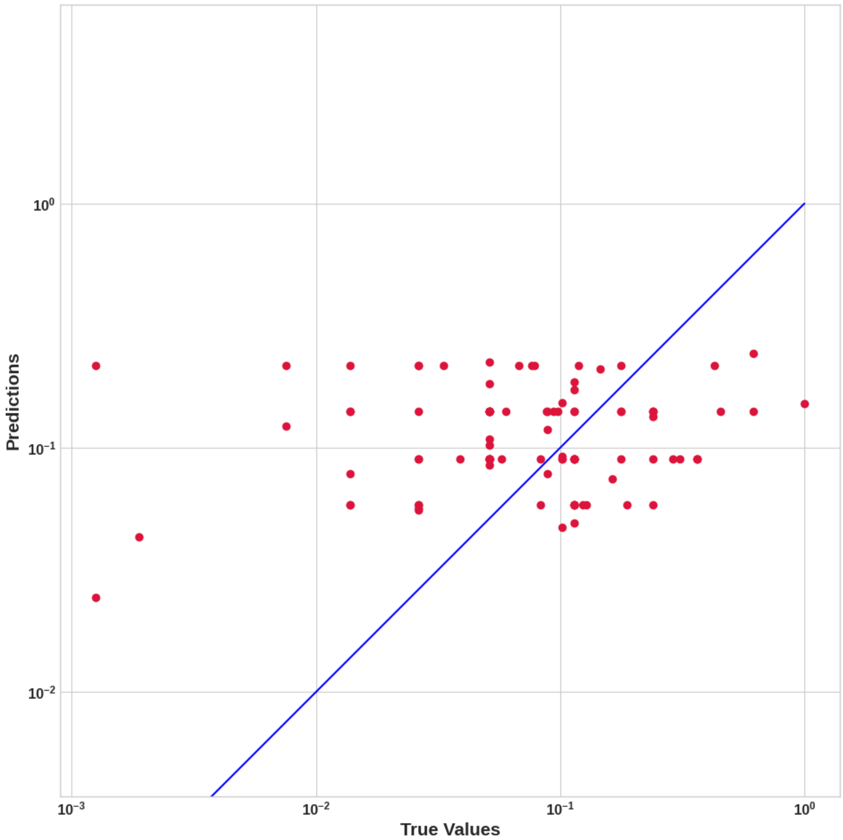

Figure 7. XGB model scatter plot

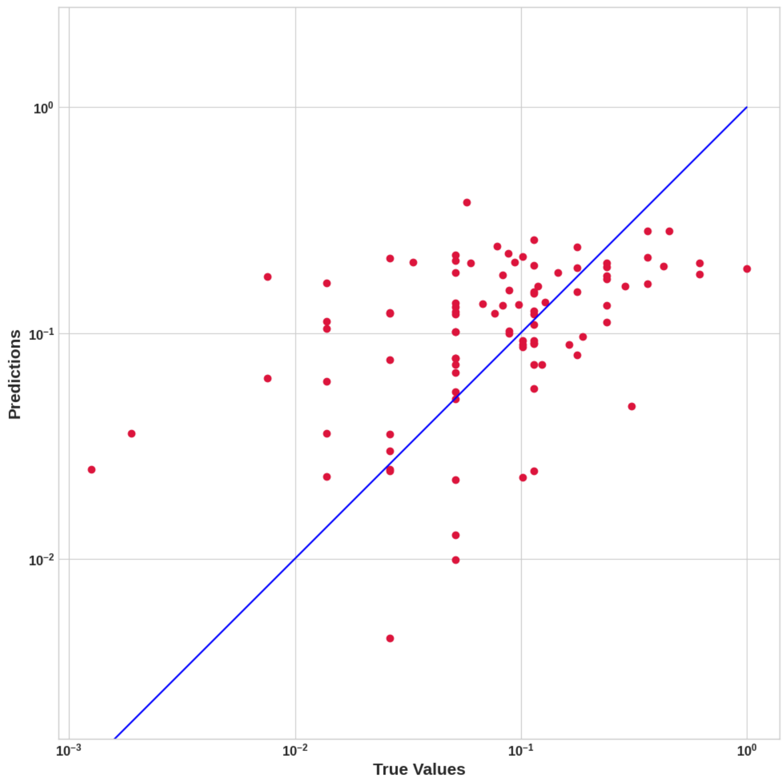

Figure 8. LGBM model scatter plot

Fig. 9 shows the boxplots of the predicted values of the four models. As seen in the figure, the LR, XGB, and LGBM models all have some degree of noise, indicating that they are more sensitive to outliers, and the models with larger differences in the figure indicate that the variance of the test results is larger.

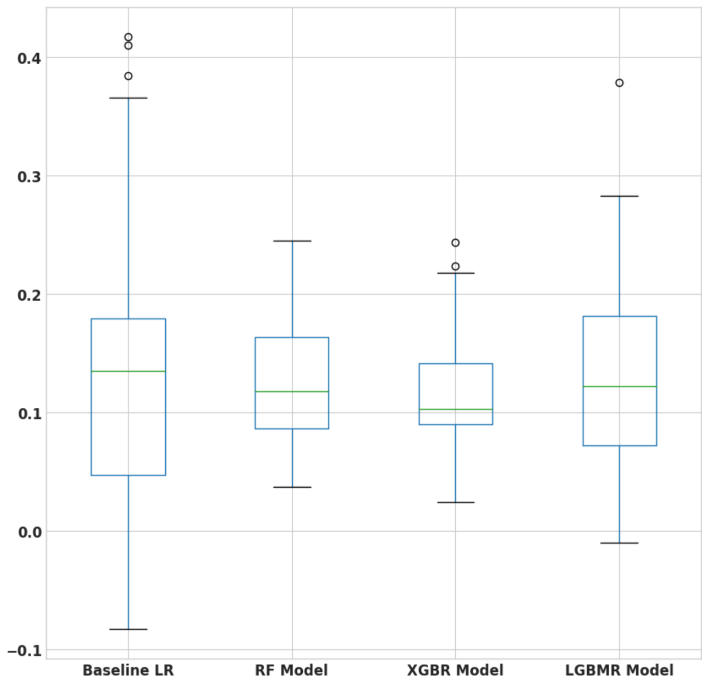

Figure 9. Box plot of predicted value distribution

Fig. 10 represents the final output scores for the three coefficients of the four models. Among them, the smaller the values of RMSE and MAE, the better the predictive ability of the model; the closer the value of R^2 is to 1, the better the model fits the data. The negative value of R^2 for the LR and

XGB models in the figure indicates that they fit the experimental data poorly and are not applicable to that experimental data. Among the four models, the models with better predictive ability are the RF model and the LGBM model, with the RF model having the smallest values of RMSE and MAE, followed by the LGBM model.

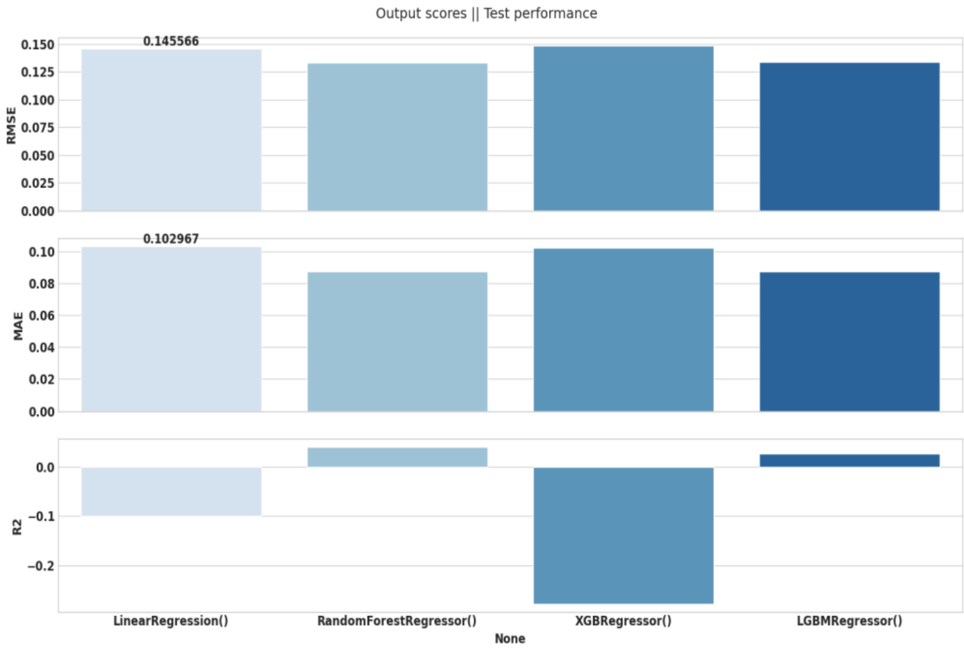

Figure 10. Output score of the performance test

Table 3 demonstrates the values of the three bias measure coefficients computed by the four machine learning algorithms mentioned above along with their average tuning times, where the average tuning times reflect the tuning efficiency and cost of the algorithmic models, and it is worth noting that the linear regression model in the table has an average tuning time that is much smaller than that of the remaining three algorithmic models, which coincides with the fast computation speed of linear regression that we have mentioned in Section 3.2 above . Among the two models with higher predictive ability, the average tuning time of the LGBM model is nearly double that of the RF model, indicating that the LGBM model is the most computationally efficient and the least costly under the requirement of accuracy, which implies that the LGBM model may be a more cost-effective method in future practical applications.

Table 3. Output score

	Baseline Linear Regression	RandomForest	eXtreme Gradient Boost	Light Gradient Boost Machine
Average tuning time	0.084946	83.503166	32.316981	45.927826
RMSE	0.145566	0.133147	0.148335	0.133655
MAE	0.102967	0.0871	0.10209	0.087206
R2	-0.099531	0.040451	-0.278338	0.025932

4.2. Discussion

In the above empirical results, we find that the Random Forest model is the model with the highest accuracy in predicting green bond issuance in this data, and it performs well in terms of prediction performance, with the RMSE value of 0.133147, the MAE value of 0.0871, and the R^2 of 0.040451 compared to the model established by Wang, Tang and Guo in predicting green bonds in 2022, with its RMSE at 0.517335 and MAE at 0.446490. CEEMDAN-LSTM model, whose RMSE is at 0.517335 and MAE is 0.446490, the prediction accuracy of this paper's model is more accurate, compared to the MLF-ANN model established by Cetin in 2022, whose RMSE is 1.7522, MAE is 1.2552, and R^2 is 0.9893, this paper's model has a smaller prediction bias than its prediction bias, but the fit is weaker, based on Cetin's (2022) study which takes the time series prediction, which shows that the data of the time series in these models are strong in explaining the issuance volume. While our current study has not yet dealt with time series, considering that it is applicable to our target variable of green bond issuance, this could be a possibility for our future related research direction.

5. CONCLUSION

The changes in climate and environment have made more and more people aware of the importance of sustainable development and have formulated relevant laws and regulations to promote environmental protection. As a result, the financial market has gradually derived the debt tool named green bonds. Although some research has already discovered and confirmed the factors of the green bonds' funding size, their researches have some limitations. Overall, existing research mostly relies on traditional statistical methods, which cannot fully capture the complex nonlinear relationships and dynamics between these factors. As for the limitations we talked about above we use machine learning to cover these limits and aim to compare the accuracy of the four machine learnings we choose to predict the funding size of green bonds.

We first analyzed the impact of market characteristics, bond characteristics, issuer characteristics, and ESG scores on the issuance of green bonds through existing literature. We found that companies with high ESG scores achieved lower issuance costs and higher market acceptance when issuing green bonds. The characteristics of bonds can also significantly affect the issuance of green bonds in certain aspects. In terms of issuer characteristics, large companies, and financially sound companies are more inclined to issue green bonds, while small companies may face more challenges. Market characteristics also have a significant impact on the issuance of green bonds in certain aspects. After completing the discussion on the impact, further accurate prediction of the issuance volume of green bonds will provide better decision-making for participants in the green bond market. In doing so, we collect relevant data on the four characteristics of enterprises to establish four machine learning models. These four machine learning models are Linear Regression, Random Forest Regressor, XGB Regressor, and LGBM Regressor. We first train each model and then input data to use the four models to predict the funding size of green bonds. Comparing the predicted results with the actual data through average tuning time, the mean absolute error (MAE), the coefficient of determination (R^2), mean square error (MSE), and the root mean square error (RMSE) to measure the accuracy of the four models we talked about above.

Among these models, on average the most accurate prediction models are Random Forest Regressor and LGBM Regressor. These models have the best numerical value in different dimensions. For instance, the average tuning time of the LGBM Regressor is nearly twice as short as that of the Random Forest model but the RMSE and MAE of the Random Forest Regressor are the least rather than the LGBM Regressor. We find that the Random Forest Regressor has a better numerical value in RMSE, MAE, and R^2 than the CEEMDAN-LSTM Model which was established by Wang, Tang, and Guo (2022) and the MLF-ANN Model established by Cetin (2022), this means the Random Forest Regressor's prediction results have a small deviation from the actual values but the fitting degree is less than MLF-ANN Model.

According to the result discussion we discussed above, the model with the best performance of the accuracy is the Random Forest Regressor which means we could use this model to forecast the issuance volume of green bonds in most cases. As the definition of the average tuning time, we give above, the LGBM Regressor has less time to use which means based on the acquisition of level of accuracy, this model has the highest computational efficiency and the lowest cost, which means that in the future practical applications, the LGBM model may be a more cost-effective method.

Our research still has some limitations under specific circumstances. Firstly, our current research has not yet involved time series, it may not be possible to identify and predict long-term trends in data and help understand the long-term patterns of data changes. Secondly, we only selected 100 companies from China, which means that our algorithm prediction still cannot cover all the companies in China that have joined the green bond market. This means that enterprises or participants that were not selected in our data may have a little deviated from the optimum issuance size if they chose to use our research conclusions and insights. Secondly, different situations have different principles of model usage in the green bond market if participants make decisions based on our current research. It might be a little more complex to choose the appropriate model to fit some situations than one model applying to all situations.

REFERENCES

[1] Baker, M., Bergstresser, D., Serafeim, G., & Wurgler, J. (2018). Financing the Response to Climate Change: The Pricing and Ownership of US Green Bonds. National Bureau of Economic Research.

[2] Baker, M.P., Bergstresser, D.B., Serafeim, G. and Wurgler, J.A. (2018). Financing the Response to Climate Change: The Pricing and Ownership of U.S. Green Bonds. SSRN Electronic Journal, [online] 42. Doi: https://doi.org/10.2139/ssrn.3275327.

[3] Barua, S. and Chiesa, M. (2019). Sustainable financing practices through green bonds: What affects the funding size? Business Strategy and the Environment, 28(6), pp.1131–1147. doi: https://doi.org/10.1002/bse.2307

[4] Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

[5] Busch, T., Bauer, R., & Orlitzky, M. (2016). Sustainable development and financial markets: Old paths and new avenues. Business & Society, 55(3), 303-329.

[6] CBI (Climate Bonds Initiative) (2021). Green Bond Market Summary. [Online] Available at: www.climatebonds.net

[7] ÇETİN, D.T. (2022a) 'Green bonds in Climate Finance and forecasting of Corporate Green Bond Index value with Artificial Intelligence', Journal of Research in Business, 7(1), pp. 138–157. doi:10.54452/jrb.992368

[8] Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of

[9] Ehlers, T., & Packer, F. (2017). Green bond finance and certification. BIS Quarterly Review, September.

[10] Flammer, C. (2021). Corporate green bonds. Journal of Financial Economics, 142(2), 499-516.

[11] Hachenberg, B., & Schiereck, D. (2018). Are green bonds priced differently from conventional bonds? Journal of Asset Management, 19(6), 371-383.

[12] ICMA (International Capital Market Association) (2022). The Green Bond Principles. [Online] Available at: www.icmagroup.org

[13] Information Processing Systems, 2017-Decem (Nips), pp. 3147–3155.

[14] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017.

[15] Kölbel, J. F., Heeb, F., Paetzold, F., & Busch, T. (2020). Can Sustainable Investing Save the World? Reviewing the Mechanisms of Investor Impact. Organization & Environment, 33(4), 554-574.

[16] Larcker, D. F., & Watts, E. M. (2022). Where's the Greenium? Journal of Accounting and Economics, 73(2-3), 101415.

[17] Laskowska, A. (2017). The Green Bond As A Prospective Instrument Of The Global Debt Market. Copernican Journal of Finance & Accounting, [online] 6(4), pp.69–83. [Accessed 23 Jul. 2024] Available at: https://ideas.repec.org/a/cpn/umkcjf/v6y2017i4p69-83.html

[18] LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Mining. Association for Computing Machinery, New York, NY, USA, pp. 785–794.

[19] Tang, D. Y., & Zhang, Y. (2020). Do shareholders benefit from green bonds? Journal of Corporate Finance, 61, 101427.

[20] Cui, J. (2024). Does digital strategy, organizational agility, digital lead-ership promote DT? A study of digital strategy, organiza-tional agility, digital leadership affects corporate DT in Chinese technological firms. *Journal of Integrated Social Sciences and Humanities*.

[21] Cui, J., Chen, W., Wan, Q., Gan, Z., & Ning, Z. (2024). Design and Analysis of a Mobile Automation Testing Framework: Evidence and AI Enhancement from Chinese Internet Technological Companies: A Case Study. *Frontiers in Business, Economics and Management*, *14*(2), 163-170.

[22] Velte, P. (2017). Does ESG Performance Have an Impact on Financial performance? Evidence from Germany. Journal of Global Responsibility, 8(2), pp.169–178.

[23] Wang, J., Tang, J. and Guo, K. (2022) 'Green bond index prediction based on CEEMDAN-LSTM', Frontiers in Energy Research, 9. doi:10.3389/fenrg.2021.793413

[24] Wang, S. and Wang, D. (2022). Exploring the Relationship Between ESG Performance and Green Bond Issuance. Frontiers in Public Health, 10(2022). doi: https://doi.org/10.3389/fpubh.2022.897577.

[25] Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. Journal of Banking & Finance, 98, 39-60.