

Forecast of bond issuance based on ESG score

Mingxuan Zhao^{1,a}, Hao Wu^{1,b,*}, Le Liu^{2,c}, Xiping Mao^{3,d}

¹School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Suzhou, China

²School of Public Finance and Taxation, Capital University of Economics and Business, Beijing, China

³School of Business Administration, Capital University of Economics and Business, Beijing, China

^aMingxuan.Zhao22@student.xjtlu.edu.cn, ^bHao.Wu22@student.xjtlu.edu.cn, ^c2301113156@qq.com,

^dmxp123456_2022@qq.com

*Corresponding author

Abstract: With the rising focus on low-carbon initiatives, interest in sustainable products and services, including cap-and-trade policies, green bonds, and low-carbon stocks, has surged. This study comprehensively investigates how Environmental, Social, and Governance (ESG) scores influence bond issuance. Employing state-of-the-art research methods and techniques, we ensure data quality through preprocessing, including handling missing values and normalization. Our predictive model, powered by machine learning algorithms such as linear regression, KNNR, XGB, and LGBM, adeptly captures relationships and handles high-dimensional features. Feature engineering further enhances model performance. Rigorous cross-validation and evaluation metrics like RMSE, MAE, and R2 ensure objectivity. Our research offers valuable insights for investors, issuers, and regulators in sustainable finance decision-making.

Keywords: ESG score, bond issuance, linear regression, KNNR, XGB, LGBM, feature engineering, cross-validation, sustainable finance

1. Introduction

In the last few years, there has been a gradual increase in the financial industry's focus on environmental, social and governance (ESG) factors. ESG factors have attracted a lot of attention as investors, issuers and regulators realize the importance of integrating sustainability into the financial decision-making process ^[1]. The concept of ESG encompasses a range of environmental, social and governance criteria used to assess the sustainability and ethical impact of investments. In the bond market, ESG factors have become an important influence in bond issuance decisions. Increasingly, investors are seeking bonds that meet their ESG preferences, and issuers are recognizing the need to incorporate ESG factors into their financing strategies. A bond's ESG score becomes an important in assessing its attractiveness to investors and can significantly influence the demand for and issuance bonds.

Within the bond market, ESG considerations have emerged as influential factors in bond issuance decisions. Investors are increasingly seeking bonds that align with their ESG preferences, and issuers recognizing the need to incorporate ESG factors into their financing strategies. The ESG score of a has become an important determinant of its attractiveness to investors and can significantly impact demand and issuance volume.

Policy initiatives and regulatory measures have further elevated the importance of ESG in bond issuance. Governments and regulators have promoted the incorporation of ESG guidelines into markets to enhance transparency, risk management and sustainability. For example, some jurisdictions have introduced mandatory ESG reporting requirements that bond issuers must comply with, while others have encouraged the issuance of green bonds to raise funds for environmental projects ^[2].

The introduction of these policies and regulations has driven the widespread use of ESG factors in bond market. Bond issuers are increasingly considering ESG criteria in their decision-making process, and bond investors are paying more attention to the ESG performance of bonds. On the one hand, the policy push has enhanced sustainability and responsible investment in the bond market; on the other ESG reporting requirements for issuers have increased the transparency of disclosure and helped investors assess the risk and return of bonds.

A growing body of literature demonstrates the importance of ESG integration in bond markets.

studies have examined the relationship between ESG factors and the financial performance of bonds and have highlighted the positive impact of high ESG scores on bond pricing and investor demand ^[3].

These studies have concluded that bond issuers with high ESG scores tend to gain better pricing advantages and wider investor attention. High ESG scores can be viewed as a bond issuer's excellence in environmental, social, and governance performance, which is an important factor of attractiveness for investors.

In addition, the researchers highlight the role of ESG rating agencies in providing standardized assessments of ESG performance. These rating agencies help investors make informed investment decisions based on accurate ESG data by providing bond issuers with independent and reliable ESG ratings and reports ^[4].

The objective of this study is to explore the relationship between ESG scores and bond issuance. Specifically, we aim to analyze how ESG scores influence the demand for bonds and whether they can serve as predictors of bond issuance volume. By examining the impact of ESG factors on bond issuance, this research seeks to contribute to the understanding of sustainable finance and provide valuable insights for investors, issuers, and policymakers. Understanding the role of ESG scores in bond issuance can inform investment strategies, facilitate the financing of sustainable projects, and support the integration of ESG considerations into financial decision-making processes.

2. Literature Review

In this section, we provide a comprehensive review of the literature relevant to our study on forecasting bond issuance based on ESG scores. Our literature review encompasses three main areas: the merits and limitations of ESG integration in bond markets, the relationship between ESG performance and bond issuance, and the impact of ESG factors on bond prices and returns.

Numerous studies have examined the advantages and challenges of ESG integration in bond markets. Bialkowski Jędrzej and Sławik Anna (2021) investigated the impact of ESG ratings on bond issuance and found that bonds with higher ESG ratings tend to attract greater investor interest and command lower yields ^[5]. Umar (2022) explored the influence of ESG ratings on bond prices and returns and revealed a positive correlation between higher ESG ratings and bond performance ^[6]. Gigante Gimede & Manglaviti Davide(2022) examined policy initiatives and regulations fostering ESG integration in bond markets, shedding light on the regulatory landscape and its impact on ESG integration ^[7]. Nagy Marek, Valaskova Katarina & Durana Pavol(2022).explored the role of ESG ratings agencies in bond markets, highlighting their influence in providing standardized ESG assessments for investors and issuers ^[8].

Some researchers have examined the relationship between ESG performance and bond issuance. Clément Alexandre, Robinot Élisabeth & Trespeuch Léo(2022) analyzed the connection between issuer ESG performance and bond issuance volume and observed that issuers with stronger ESG performance tend to have higher bond issuance levels ^[9]. International Monetary Fund(2022) conducted a comprehensive study on the differential impact of ESG performance on bond issuance volume across industries and regions, highlighting the varying effects based on specific contexts ^[10].

The significance of ESG factors in bond pricing and returns has garnered substantial attention. Several studies have demonstrated the relationship between ESG factors and bond performance. Zhong Shen, Hou Junzhu, Li Junwei & Gao Wei.(2022). examined the influence of ESG ratings on bond prices and returns and revealed a positive association between higher ESG ratings and bond performance ^[11]. Raza Hassan & Abbas Kumail.(2022). conducted empirical research on the impact of ESG factors on bond pricing, providing insights into the pricing dynamics influenced by ESG considerations ^[12]. Gao Jie, Li Jiahao & Luo Yuwei(2022)conducted a comprehensive analysis of the relationship between ESG performance and bond market outcomes, revealing the potential financial benefits associated with higher ESG scores ^[13]. Rohit Goel, Deepali Gautam & Mr. Fabio M Natalucci(2022) investigated the impact of ESG disclosure on bond pricing and found that greater transparency in ESG reporting positively affects bond prices ^[14]. Shushi Tomer (2022)explored the relationship between ESG factors and credit ratings, highlighting the influence of ESG considerations on creditworthiness ^[15].

In summary, the literature highlights the increasing relevance of ESG integration in bond markets, the positive relationship between ESG performance and bond issuance, and the impact of ESG factors on bond prices and returns. Based on these findings, our study aims to forecast bond issuance based on ESG scores. By employing appropriate models and methodologies, we seek to contribute to the existing

body of knowledge and provide valuable insights for investors, issuers, and policymakers in the sustainable finance domain.

3. Data

We obtained the bond market dataset from Reuters' Financial Data service, which offers comprehensive financial market data, including ESG scores, environmental, social, and governance pillar scores, asset returns, bond issuance ratio, and other indicators for nearly 500 bonds over the past five years. This timeframe allows for in-depth analysis and prediction of the relationship between ESG scores and bond issuance. Our data collection adheres to the terms of use and regulations set by the data provider.

During the data processing stage, we performed several steps. Firstly, we designed the data type for date-related information to facilitate time series analysis and calculations. Then, we checked for missing values and cleaned the data accordingly. To eliminate differences in bond sizes, we normalized the amount of bonds issued, representing it as a relative value between 0 and 1.

Additionally, we processed the issuance period of each bond and converted it into a numerical format for analysis and model building. Outliers were identified using Tukey's boxplot method based on the amount of bond issuance. Categorical variables were transformed into binary form using One-Hot Encoding for subsequent analysis and modeling. The binarized data is presented in Table 1:

Table 1: Bond rating after binarization

ISIN	ESG Score	...	Return On Asset	...	A1+	A3	AA-	BB	BB+	BB-	BBB	BBB+	Ba3	Caa3
TH0221032B04	73.4	...	2.2	...	0	0	0	0	0	0	1	0	0	0
FR0013048204	89.06	...	3	...	0	0	0	0	0	0	1	0	0	0
NZCEND0040L1	54.09	...	4.89	...	0	0	1	0	0	0	0	0	0	0
JP373480AKB4	64.28	...	6.97	...	0	0	0	0	0	0	1	0	0	0
SE0013882743	59	...	-0.3	...	0	0	0	0	0	1	0	0	0	0
...
JP356680AN38	34.07	...	2.07	...	0	0	0	0	0	0	1	0	0	0
JP356680AN53	34.07	...	2.07	...	0	0	0	0	0	0	1	0	0	0
JP358260BK34	56.51	...	3.36	...	0	0	1	0	0	0	0	0	0	0
FR0014003GT5	94.61	...	0.41	...	0	0	0	0	0	0	0	0	0	0
TW000B609072	72.75	...	21.33	...	0	0	0	0	0	0	1	0	0	0

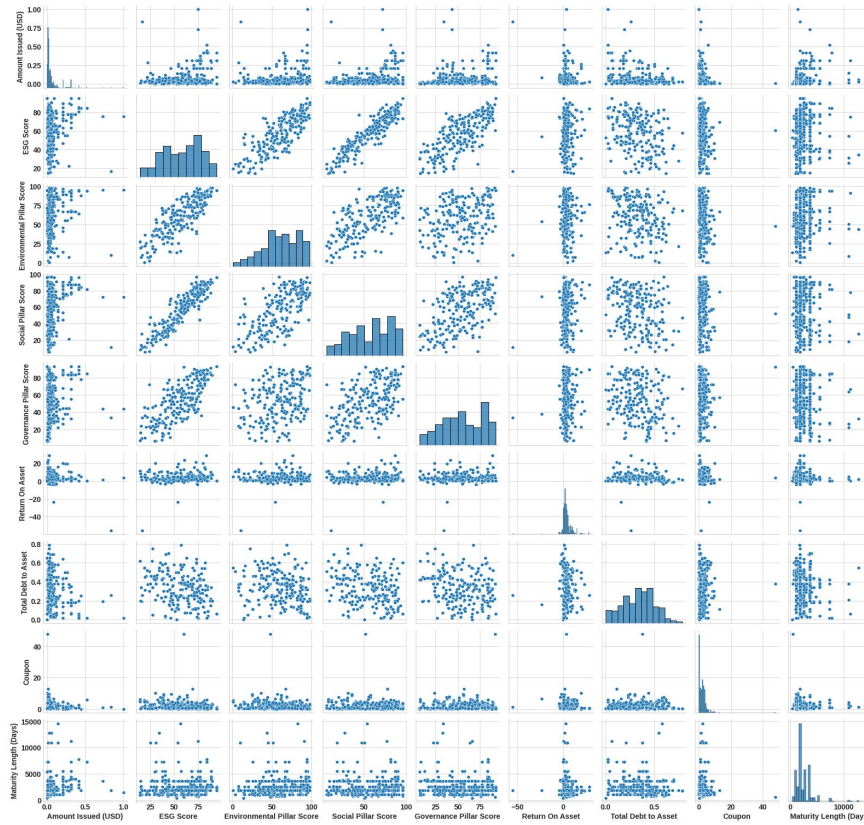


Figure 1: The scatter plot matrix between the continuous characteristic variables

We selected a set of characteristic variables as model inputs and reorganized their order. To observe the linear relationship between these variables more intuitively and identify outliers, we created a scatter plot matrix for the continuous characteristic variables. The distribution and their relationship to the normalized amount issued are depicted in Figure 1.

Through the above data processing steps, we loaded, cleaned and transformed the bond market dataset and were ready for the subsequent machine learning modeling work. These steps not only ensure the reliability and accuracy of the data, but also provide an effective data basis for subsequent research.

4. Model

4.1 Linear Regression

After going through the data processing stage, we will move to the model part and use the prepared data to build our prediction model. We first selected the linear regression model as the basis, and combined with the nested cross-validation method to optimize and evaluate the model, to preliminarily realize the use of the ESG score of bond issuing companies to predict their future bond issuance.

Linear regression model is a classical machine learning model used to establish linear relationships between variables. In our model, we include bond issuance as the dependent variable, and the ESG score of the bond issuing company and other relevant characteristics as the independent variables. The following is the basic expression of the linear regression model:

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \tag{1}$$

Where y is the bond issuance, $x_{i1}, x_{i2}, \dots, x_{in}$ is the ESG score and other relevant characteristics of the issuer, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ is the coefficient of the model.

The linear regression model has certain explanatory power and wide application. With the help of this model, we can gain insight into the relationship between ESG scores and bond issuance, providing predictability and interpretability to the bond market. In addition, linear regression models are computationally simple and suitable for small-scale data sets, making them an effective and reliable

choice.

To optimize and evaluate the performance of the model, we employed a nested cross-validation approach. In this method, the dataset was divided into an internal cross-validation set and an external cross-validation set. Internal cross-validation was used to select the best model hyperparameters, and external cross-validation was used to evaluate the performance of the model. The principle of nested cross validation is to integrate and compress the data set into a single sample of size n , and then divide the sample into n fold sets accordingly. One fold at a time is selected as the test set, and the remaining folds are selected as the training set^[16]. For each test set, we used internal cross validation to select the best model hyperparameters. In this way, we can make full use of the dataset for model selection and evaluation, checking the fit of the model and reducing the risk of overfitting. The algorithm and results are as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2)$$

$$RMSE = \sqrt{MSE} \quad (3)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (4)$$

Here is a brief explanation of the formula: \hat{y}_i is a generic estimate of y_i (Return value) the activity of the i th compound in the sample. \bar{y} for the sample mean of the y_i gives the estimate (sums are over the set of compounds used to assess the fit). SS_{res} : squared difference between the return value and the true value. SS_{tot} : variance of the true value of the data.

4.2 K-Nearest Neighbors Regression

Our second model is Nearest neighbor classification, which has another famous name---K-Nearest Neighbors Regression (KNNR). KNNR is an instance-based learning method for prediction in regression problems. It predicts according to the target value of the K-nearest patterns in data sample, and obtains the final prediction result by weighted average or simple average method. The mathematical formula can be expressed as:

$$\hat{y}_i = \frac{1}{k} \sum_{j=1}^k y_j \quad (5)$$

Where y_i is the target value of the i th neighbor.

Compared with linear regression models, the advantages of KNNR include simplicity, flexibility, and adaptability to nonlinear relationships. It can better identify nonlinear complex factors and some models that may be difficult to separate with simple rules or mathematical^[17]. Hence, it could achieve more accurate prediction of bond issuance. To evaluate and optimize the performance of the KNNR model, we evaluated the predictive performance of the model by calculating root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). By defining an objective function and an optimization algorithm to search for the best hyperparameter configuration, the distribution between the true and predicted values of the model is recorded. We will again use nested cross-validation. In the internal cross-validation, we use the Optuna library for hyperparameter optimization of the KNNR model, and the results show that the KNNR model performs well in predicting the bond issuance based on ESG score and remains close to the true value.

4.3 XGBoost Regressor

In this study, we adopt XGBoost (Extreme Gradient Boosting) as the third machine learning model to predict bond issuance combined with ESG scores. XGBoost is an ensemble learning algorithm based on gradient boosted decision tree, whose goal is to improve the performance and generalization ability of the overall model by integrating multiple weak learners.

The original model of XGBoost can be expressed as the following mathematical formula:

$$\hat{y}_i = w_0 + \sum_{j=1}^m w_j \cdot x_{ij} \quad (6)$$

Where, y_i is the predicted output of the i th sample, w_0 is the bias term of the model, w_j is the weight corresponding to the feature x_j , and x_{ij} is the j th feature of the i th sample.

XGBoost is an ensemble learning algorithm based on gradient boosted decision trees. Compared with linear regression and K-Nearest Neighbor regression (KNNR), XGBoost is able to handle complex nonlinear relationships, allow for interactions between features, and can better capture these complex relationships, thereby improving prediction accuracy. The most important factor behind XGBoost's success is its scalability across all scenarios, the weight of each feature can be adaptively learned to further optimize the model performance [18].

4.4 LightGBM Regressor

After evaluating the performance of previous models, we focused on the LightGBM regressor, which offers distinct advantages over XGBoost. LightGBM's capability to handle complex nonlinear relationships and feature interactions enables more accurate capturing of comprehensive influencing factors in the bond market. Its histogram-based decision tree construction strategy makes training on large-scale datasets faster, handling substantial data efficiently.

Despite these benefits, LightGBM regressors may suffer from overfitting, especially with limited data. To address this, we introduced the Gradient Boosting method based on differential privacy (GBDP), which enhances the model's generalization ability and robustness while protecting data privacy. GBDP adds differential privacy noise during the node splitting process in each decision tree, reducing over-reliance on sensitive data and mitigating overfitting risks.

For model training, a nested cross-validation method is employed to optimize hyperparameters such as learning rate, tree depth, and number of leaf nodes. This approach ensures the model's adaptability and generalization for complex bond market predictions, resulting in more efficient and concise models.

Through a series of experiments, we obtained prediction results for bond issuance using the LightGBM regressor. We evaluated the model's performance using mean output time, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R2). Introducing the GBDP method and enhancing the model structure slightly decreased prediction accuracy compared to previous models like XGBoost. However, this trade-off was offset by significant improvements in output time and information privacy.

5. Empirical results

In this study, we applied four different machine learning models for the bond issuance prediction task. These include linear regression, k-Nearest Neighbor regression (KNNR), XGBoost and LightGBM Regressor, and the optimized Gradient Boosting with Differential Privacy (GBDP) model. In this section, we'll compare the performance of these models on multiple dimensions for predicting bond issuance and draw some conclusions.

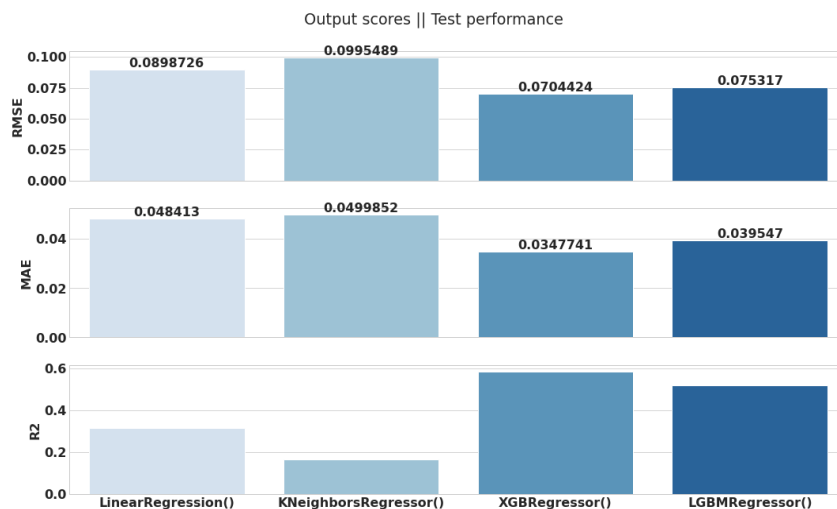


Figure 2: The performance comparison of different models on RMSE, MAE and R2

In Figure 2, the data frame displays the average performance evaluation metrics of various models on the test data, encompassing root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R2). These metrics gauge predictive accuracy and model fitting ability. Three subfigures illustrate the evaluation results for each metric, with the X-axis denoting the model name and the Y-axis indicating the corresponding numerical value. The first subplot shows the average RMSE, the second displays the average MAE, and the third presents the average R2. These visualizations facilitate straightforward comparison of different models' performance on these metrics.

Through comparisons among the four machine learning methods in bond issuance prediction, XGBoost and LightGBM exhibit superior predictive accuracy compared to Linear Regression and KNNR models. With lower prediction errors, XGBoost and LightGBM effectively capture the complex nonlinear relationships and feature interactions in the bond market, enhancing prediction accuracy.

Regarding the Coefficient of Determination (R2) index, XGBoost and LightGBM models excel, better explaining the variability of the target variable and showcasing stronger predictive abilities compared to Baseline Linear Regression and KNNR models. However, in terms of data processing time, XGBoost, optimized by the superposition of the random forest model, slightly lags behind the other three methods. Concerning data privacy and security, the LightGBM model experiences significant improvements through the superposition of the GBDP model.

In Figure 3, the data frame exhibits the standard deviation of predictions from different models on the test data. These standard deviations are indicative of the predictive stability and volatility of each model. Boxplots are employed to visualize the predictions, with the X-axis representing the model name and the Y-axis depicting the numerical value of the prediction. The boxes in the boxplot represent the interquartile range of prediction results, the whiskers portray the overall distribution range, and outliers are identified.

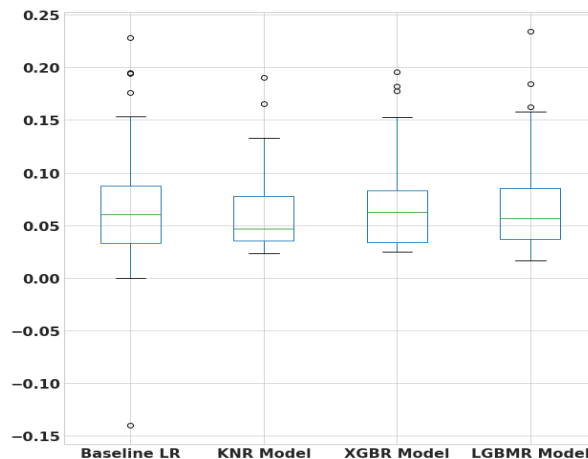


Figure 3: The boxplot of comparison of distribution and outliers of predictions from different models

From the boxplots, the following conclusions can be drawn about the four models:

The LightGBM model demonstrates stable and concentrated predictions around the median, signifying good data concentration and accuracy.

The XGBoost model also exhibits excellent data concentration and accuracy.

Conversely, the boxplots of the Linear Regression model and the KNNR model display a larger box length and a more divergent prediction distribution, suggesting lower data concentration and accuracy.

6. Conclusion

In conclusion, after comparing the performance of four machine learning models for bond issuance prediction, XGBoost and LightGBM stand out as preferred choices due to their superior prediction accuracy. However, considering specific requirements, XGBoost nested in random forest excels in accurately predicting low-dimensional data but may have limitations when dealing with high-dimensional data. Therefore, careful consideration is necessary when choosing the appropriate application scenario. On the other hand, GBDP nested LightGBM proves to be more suitable for handling large-scale data that requires privacy protection, effectively reducing potentially complex and

risky noise within the dataset.

For future research, further optimization of hyperparameter settings for XGBoost and LightGBM models could lead to improved prediction performance and generalization ability. This would enhance their applicability in diverse prediction tasks.

References

- [1] Sinha Ray Rupamanjari & Goel Sandeep. (2023). *Impact of ESG score on financial performance of Indian firms: static and dynamic panel regression analyses*. *Applied Economics*(15). doi:10. 1080/00036846. 2022. 2101611.
- [2] Kalia Deepali & Aggarwal Divya. (2023). *Examining impact of ESG score on financial performance of healthcare companies*. *Journal of Global Responsibility*(1). doi:10. 1108/JGR-05-2022-0045.
- [3] Tahmid Tahani, Hoque Muhammad Nazmul, Said Jamaliah, Saona Paolo & Azad Md. Abul Kalam. (2022). *Does ESG initiatives yield greater firm value and performance? New evidence from European firms*. *Cogent Business & Management*(1). doi:10. 1080/23311975. 2022. 2144098.
- [4] Bolognesi Enrica & Burchi Alberto. (2023). *The impact of the ESG disclosure on sell-side analysts' target prices: The new era post Paris agreements*. *Research in International Business and Finance*. doi:10. 1016/J. RIBAF. 2022. 101827.
- [5] Bialkowski Jędrzej & Sławik Anna. (2022). *Does Companies' ESG Performance Make a Difference for New Zealand's Stock Market Investors during the COVID-19 Pandemic?*. *Sustainability*(23). doi: 10. 3390/SU142315841.
- [6] Umar Muhammad, Mirza Nawazish, Rizvi Syed Kumail Abbas & Naqvi Bushra. (2022). *ESG scores and target price accuracy: Evidence from sell-side recommendations in BRICS*. *International Review of Financial Analysis*. doi:10. 1016/J. IRFA. 2022. 102389.
- [7] Gigante Gimede & Manglaviti Davide. (2022). *The ESG effect on the cost of debt financing: A sharp RD analysis*. *International Review of Financial Analysis*. doi:10. 1016/J. IRFA. 2022. 102382.
- [8] Nagy Marek, Valaskova Katarina & Durana Pavol. (2022). *The Effect of CSR Policy on Earnings Management Behavior: Evidence from Visegrad Publicly Listed Enterprises*. *Risks*(11). doi: 10. 3390/RISKS10110203.
- [9] Clément Alexandre, Robinot Élisabeth & Trespeuch Léo. (2022). *Improving ESG Scores with Sustainability Concepts*. *Sustainability*(20). doi: 10. 3390/SU142013154.
- [10] International Monetary Fund. Monetary and Capital Markets Department. (2022). *Global Financial Stability Report, October 2022: Navigating the High-Inflation Environment*. *Global Financial Stability Report*(002). doi: 10. 5089/9798400219672. 082.
- [11] Zhong Shen, Hou Junzhu, Li Junwei & Gao Wei. (2022). *Exploring the relationship of ESG score and firm value using fsQCA method: Cases of the Chinese manufacturing enterprises*. *Frontiers in Psychology*. doi:10. 3389/FPSYG. 2022. 1019469.
- [12] Raza Hassan, Khan Muhammad Anees, Mazliham M. S. , Alam Muhammad Mansoor, Aman Nida & Abbas Kumail. (2022). *Applying artificial intelligence techniques for predicting the environment, social, and governance (ESG) pillar score based on balance sheet and income statement data: A case of non-financial companies of USA, UK, and Germany*. *Frontiers in Environmental Science*. doi: 10. 3389/FENVS. 2022. 975487.
- [13] Gao Jie, Li Jiahao & Luo Yuwei. (2022). *Does the enterprise's choice of intangible capital help to promote its ESG score?* *Frontiers in Environmental Science*. doi:10. 3389/FENVS. 2022. 978955.
- [14] Rohit Goel, Deepali Gautam & Mr. Fabio M Natalucci. (2022). *Sustainable Finance in Emerging Markets: Evolution, Challenges, and Policy Priorities*. *IMF Working Papers* (182). doi: 10. 5089/9798400218101. 001.
- [15] Shushi Tomer. (2022). *The optimal solution of ESG portfolio selection models that are based on the average ESG score*. *Operations Research Letters* (5). doi:10. 1016/J. ORL. 2022. 07. 008.
- [16] Douglas M. Hawkins, Subhash C. Basak, & Denise Mills. (2003) *Assessing Model Fit by Cross-Validation*. *J. Chem. Inf. Comput. Sci.* 2003, 43, (579-586) DOI: 10. 1021/ ci025626i
- [17] J. Kacprzyk & L. C. Jain (2013) *Dimensionality Reduction with Unsupervised Nearest Neighbors*. *Intelligent Systems Reference Library, Volume 51*. DOI 10. 1007/978-3-642-38652-7
- [18] Tianqi Chen & Carlos Guestrin (2016). *XGBoost: A Scalable Tree Boosting System*. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining August 2016 785–794*, DOI:10. 1145/2939672. 2939785