Analysis of the Influential Factors and Prediction of Corporate ESG Performance under Multi-source Data Fusion - Based on Frontier Machine Learning

Chenhong Zheng [a], Mengqian Zhang [b], Cong Zeng [c], Fangshun Xiao [d], Mengzhe Liu [e*]
[a]chenhongzheng_work@126.com, [b]mengqian_zhang20@126.com, [c]zengcong80s@126.com, [d]xfs_stri@126.com, [e*]liumengzhe20@126.com

Economic and Technological Research Institute, State Grid Fujian Electric Power Co., Ltd., Fuzhou, Fujian, China

Abstract—Environmental, Social and Governance (ESG) performance, which reflect the degree of corporate green development and sustainable development potential, is one of the key factors determining the long-term development of enterprises. Based on the integration of macro data and enterprise data, this paper firstly explores the factors influencing ESG performance from the perspective of both external environment and internal characteristics of enterprises using the Lasso method, and then screens them effectively. Then, based on the screening results of the influencing factors and the underlying data set, various machine learning methods such as BP, Random Forest, KNN and SVM models are used to predict the ESG performance of enterprises, and the prediction accuracy of each method is compared and analyzed. Finally, based on the prediction results and the screening results, the importance of each influencing factor on the ESG performance of enterprises is ranked comprehensively.

Keywords- ESG; Machine Learning; BP; Random Forest; SVM

1. Introduction

Due to the long-term pursuit of profit maximization as the goal of China's enterprises in the crude development model, environmental pollution problems, product quality failure, lack of corporate social responsibility and alienation phenomena are frequent. The 20th National Congress of the Communist Party of China clearly points out that we should focus on promoting high-quality development, adhere to the focus of economic development on the real economy, improve the quality of people's lives and promote green development. Nowadays, public awareness of ecological civilization is constantly strengthening, and the national high-quality development strategy also puts forward higher requirements for enterprises. How to coordinate the balance between profit growth, environmental protection and public interest is the key issue facing corporate governance under the contemporary new development concept. Moreover, in the era of big data and information development, the application of machine learning methods to corporate governance is an important tool for green transformation of enterprises.

ESG, as an important indicator of corporate environmental, social and governance performance, can effectively reflect the degree of corporate green development and sustainable development potential, and is also one of the key factors determining the long-term development of enterprises. Existing studies have focused on the economic benefits of ESG responsibility, and a large

BDEIM 2022, December 02-03, Zhengzhou, People's Republic of China

DOI 10.4108/eai.2-12-2022.2328731

number of studies have shown that corporate social responsibility can effectively increase corporate value [1-2]. It is mainly in the sense that corporate ESG responsibility can attract investors, help establish long-term trust relationships with investors, and alleviate financing constraints. In addition, in a market with high product homogeneity, consumers are more inclined to choose products with excellent ESG performance, and the fulfillment of ESG responsibility by companies can attract consumers through brand effect and reputation mechanism, gain competitive advantage, occupy market share, and increase product profits. With the rapid economic recovery in the post-epidemic era and the increasingly fierce competition faced by enterprises, it is important to motivate enterprises to fulfill their ESG responsibilities and improve ESG performance to protect people's welfare while developing the real economy and promoting high-quality development. Therefore, studying the impact factors of corporate ESG and pinpointing the driving forces of corporate ESG is beneficial for enterprises to make production decisions. Moreover, making effective predictions of corporate ESG performance based on these influencing factors can help to better understand the deviations between the current behavioral choices and results of enterprises, which is favorable for enterprises to efficiently adjust their existing production strategies and investment structures to quickly achieve green transformation and contribute to high-quality economic development.

However, only a few studies have explored the issue of corporate ESG influencing factors, mainly using traditional methods, based on a particular economic and environmental policy, or a firm's internal characteristics to examine the impact on corporate ESG performance [3-5]. The main factors studied include mandatory disclosure requirements for pollutant emissions [6-7], financial factors such as total assets, net profit margin, leverage ratio [8], and corporate governance systems [9]. These studies failed to encompass all possible external environmental and firm characteristics factors, and it is not possible to compare the contribution of each type of impact factors to corporate ESG performance. Currently, machine learning methods are generally recognized for their excellent performance in research, but are less often applied to the analysis and prediction of factors influencing corporate ESG performance. LASSO method is often used for factors screening, which was first proposed by the statistician [10]. The main idea is to construct a penalty function from least squares estimation, so that the coefficient estimates of some of the weak impact indicators are "compressed" to zero, thus achieving the purpose of refining the original set of indicators. And it is now widely used in studies of medicine [11], financial market behavior [12], and electricity spot prices [13]. In addition, machine learning methods such as BP neural network, random forest, KNN, and SVM are widely used for forecasting based on multiple influencing factors in different fields, with research focusing on daily average wind speed [14], short-term natural gas load [15], electricity market price [16], macroeconomic growth [17], sales volume [18], and cost [19].

Therefore, we select Chinese A-share listed companies from 2007 to 2020 as the research sample, based on the multivariate data fusion of macro data and micro data, combine big data with machine learning methods, and firstly, explore the impact factors of corporate ESG performance from both external environment and internal company characteristics using the Lasso method and effectively screens them; secondly, verify the validity of LASSO screening indicators using the econometric regression method; and then machine learning models such as BP, random forest, KNN and SVM are used to predict corporate ESG performance, and a comparative analysis of the prediction accuracy of each method is conducted; finally, the relative importance of the factors on ESG is synthesized based on the screening results and prediction results.

Compared with the established literature, this study has the following research features and meanings: (1) In terms of research content, we explore the influencing factors of corporate ESG under the general framework of the external environment and internal characteristics of enterprises, and further predict ESG performance, which enriches the research on corporate ESG. (2) In terms of research methodology, the traditional methods are innovated. We combine big data with machine learning methods and use machine learning to combine indicator screening with predictive models to study ESG issues of Chinese enterprises, which can not only fill the existing literature gaps, but also help enterprises to make better strategic adjustments. (3) In terms of application value, by analyzing the direction and degree of influence of various factors, we reveal the importance of external environment and internal characteristics in improving ESG performance of enterprises, which may provide information and reference for government regulation and policy making, the public behavior and corporate strategy choices.

2. Data and Methodology

2.1 Data

We use the data of Chinese A-share listed companies from 2007-2020 to analyze the impact factors and prediction of corporate ESG performance, with corporate ESG scores as the explanatory variables, which is from Bloomberg database, and 14 indicators of both corporate external environmental factors and internal characteristics are selected as explanatory variables. The external environmental factors are: (1) media attention (News), using the logarithm of total number of financial newspaper media reports and online media reports; (2) stock bar attention (Public), measured by the logarithm of the number of stock bar postings of listed companies; (3) economic policy uncertainty (EPU), using the economic policy uncertainty index constructed by Baker et al. (2016), where we employ the annual arithmetic mean to convert monthly economic policy uncertainty into annual economic policy uncertainty; (4) environmental regulation (Regulation), using the environmental regulation index measured by three waste emissions; (5) financing constraint (SA); (6) marketability index (Mrk), measured by the marketability index; the intra-firm characteristics are: (7) the proportion of independent directors (Ind); (8) the shareholding ratio of the company's first largest shareholder (Top); (9) asset-liability ratio (Lev); (10) rate of return on total assets (Roa); (11) age of company listing (Age); (12) nature of equity (SOE), 1 for state-owned enterprises and 0 for non-state-owned enterprises; (13) shareholding ratio of institutional investors (Inv); and (14) cashflow ratio (Cfr). Among them, the data of media attention and share bar attention are obtained from the CNRDS database, and the data of the remaining related variables are mainly obtained from the CSMAR database, the Statistical Yearbooks of Chinese provinces, and China Statistical Yearbook on Environment.

2.2 Methodology

2.2.1 Lasso Regression.

In order to reduce the subjective influence in the influencing factors, this paper uses Lasso method for indicator screening and eliminates the intervention by introducing a penalty term in the equation. Considering that the study of this paper is based on panel data, the regression model is set as follows.

$$y_{it} = \beta_0 + \sum\nolimits_{\mathrm{i}=1}^{Q} \beta_n x_{\mathrm{int}} + \alpha_i + \mu_t + \varepsilon_{it} \qquad (1)$$

Where, $\mathrm{y_{it}}$ is the ESG performance of firm i in year t, x is the impact factor on corporate ESG, and β is the coefficient of each indicator. α_i is the firm fixed effect, μ_t is the time fixed effect, and ε_{it} is the random error term. For Equation (1), taking the annual average value of each indicator, Equation (2) is derived.

$$\overline{y_t} = \beta_0 + \sum\nolimits_{\mathrm{i}=1}^{Q} \beta_n \overline{x_{\mathrm{in}}} + \alpha_i + \mu_t + \overline{\varepsilon_\iota} \qquad (2)$$

Where$\overline{y_\iota} = \sum_{t=1}^{T} y_{it}/T\ , t = 1, \dots, T$, $\overline{\mathrm{x}_{\iota n}} = \sum_{t=1}^{T} x_{nit}/T\ , t = 1, \dots, T$, $\overline{\varepsilon_\iota} = \sum_{t=1}^{T} \varepsilon_{it}/T\ , t = 1, \dots, T$.

Therefore, we can get Equation (3).

$$y_{it} - \overline{y_\iota} = \sum\nolimits_{\mathrm{i}=1}^{Q} \beta_n (x_{int} - \overline{x_{\mathrm{in}}}) + \varepsilon_{it} - \overline{\varepsilon_\iota} \qquad (3)$$

Defining$\tilde{\mathrm{Y}} = [\mathrm{y}_{11}^* - \overline{\mathrm{y}_1^*}, \dots, \mathrm{y}_{\mathrm{nT}}^* - \overline{\mathrm{y}_\mathrm{n}^*}]^\mathrm{T}$, $\widetilde{\mathrm{x}_{int}} = \mathrm{x}_{int} - \overline{\mathrm{x}_{\iota n}}$, we can have the matrix as follows.

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{1,1,1} & \cdots & \tilde{x}_{1,p,1} \\ \vdots & \vdots & \vdots \\ \tilde{x}_{n,1,T} & \cdots & \tilde{x}_{n,p,T} \end{bmatrix}$$

Then, we can get the expression of LASSO.

$$\min_\beta \frac{1}{2} \left(Y - \tilde{X}\beta\right)^T \left(Y - \tilde{X}\beta\right) + \lambda \sum\nolimits_{n=1}^{N} |\beta_n| \qquad (4)$$

The penalty coefficient in Equation (4) is λ, which is determined by the cross-validation method. The penalty term is used to shrink the index coefficients, so that the influencing factors that are zero after shrinkage can be removed to reduce the subjective influence.

2.2.2 BP neural network model.

The goal of the BP (Back Propagation) neural network model is to find the minimum value of the squared network error. This method is able to dig deeper into the potential information in the data and identify the relationship between variables effectively. In the actual operation, the main steps are as follows.

a) Initialize and set up the neural network and set the relevant parameters.

b) Perform output calculations for the implicit layer and the output layer, respectively.

c) Calculate errors based on the predicted output and true output, and update the relevant variables.

d) Identify whether the iteration is completed or not, and if the iteration is completed, the computation is completed, and if the iteration is not completed, return to the second step for another computation.

2.2.3 Random Forest Model.

The random forest model relies on an integrated algorithm consisting of multiple decision trees. This method mainly samples the sample units and variables, and then forms a decision tree. The main idea is as follows, assuming that the sample set X has N sample units and M feature attributes.

a) Generate a decision tree by sampling Q samples in the sample set X based on the Bootstrap method.

b) Randomly select m features (m < M) for any node to ensure the same number of features for each node, and then generate all decision trees.

c) The prediction samples are analyzed using random forest trees, and the mean of the predicted values of each tree is used as the final predicted value.

2.2.4 KNN model.

The principle of KNN algorithm is to find the k nearest neighbors of the unknown sample. According to the principle that distance is inversely proportional to weight, different weighting values are assigned to the attributes of each nearest neighbor to obtain the attributes of the unknown sample. Based on this principle, the KNN model can identify the relationship between ESG and its influencing factors based on the distance information of the samples. The KNN model is basically set as follows.

$$\hat{Y}_t = 1/s \sum_{X_i \in N_s(X_t)} Y_i \tag{5}$$

In the space of influencing factors, the prediction value Y_i of ESG performance of the nearest k samples $X_i \in N_s(X_t)$ from the sample X_i is averaged as the final prediction value of the model. Although the KNN model does not directly estimate the relationship parameters between variables, the association between ESG and its influencing factors can be reflected by the distance information between samples. Therefore, the relationship between the two can also be effectively identified by the KNN method.

2.2.5 Super Vector Machine (SVM) Model.

The SVM model transforms the input space into a high-dimensional space by using a nonlinear transformation defined by the inner product function, and finds a nonlinear relationship between the output and input variables in the high-dimensional space to create a categorical hyperplane as a decision surface. In this paper, the SVM model takes the following functional form.

$$\hat{Y}_t = w^T \phi(X_t) + b \tag{6}$$

Where, $\phi(\cdot)$ is the basis function for the nonlinear change of ESG influencing factors, and w denotes the weights of the basis function. After the change of basis function, the SVM model can identify the nonlinear relationship between ESG and its influencing factors. It can effectively reduce the negative effects of variable multicollinearity as well as overfitting by optimizing the distance from each sample point to the support vector. Existing studies have shown that Gaussian kernel-based SVM model significantly outperform ordinary SVM with penalty terms in predicting firm performance. Therefore, in this paper, the SVM model based on a Gaussian kernel function is used to predict corporate ESG performance [20].

3. Empirical Results

3.1 The Results of LASSO

We use LASSO regression to screen multiple variables and calculate regression coefficients of each impact factor. And the results are shown in Table 1.

Table 1 Regression results of factors after LASSO screening

Variables	***Coefficient***	***Variables***	***Coefficient***
News	0.7903	Top	-0.6227
Public	0.9155	Lev	0.9412
EPU	-0.0298	Age	1.0348
SA	0.6771	SOE	1.0880
Mrk	0.5804	Inv	0.0310
Ind	0.0182	Cfr	2.6500
Regulation	0	Roa	0

As shown in Table 1, 12 eligible independent variables were obtained from 14 independent variables, namely: media attention, stock bar attention, economic policy uncertainty, financing constraint, marketability, percentage of independent directors, percentage of shares held by the company's largest shareholder, gearing ratio, age of the company's listing, nature of equity, percentage of shares held by institutional investors, and cash flow ratio. The other two variables are environmental regulation and rate of return on total assets, whose coefficient are zero after penalty compression, indicating that these two variables are not significant for firm ESG performance, suggesting that the relationship between these two indicators and ESG is not strong. For environmental regulation, environmental regulation policy only affects the environment, while the influence on social responsibility and corporate governance is limited, and the current implementation of environmental regulation policy in China is low. Therefore, the coefficient of environmental regulation will be compressed to 0 after penalty. Regarding the rate of return on total assets, this indicator reflects the level of comprehensive earnings of enterprises, which is mainly related to the assets and profits of enterprises, therefore, the coefficient of the rate of return on total assets is 0 after penalty. Therefore, only 12 significant variables are included in the final model. Further OLS regressions are performed on the variables that have been screened to test the parameters, and the results are shown in Table 2. The results show that after LASSO screening, the regression coefficients of all variables pass the significance test, indicating that the screening results of LASSO are valid and reasonable.

According to the coefficients of effective variables, among the variables related to the external environment, economic policy uncertainty exhibits a significant negative effect on corporate ESG performance, indicating that under a high degree of uncertainty, the fulfillment of ESG responsibility by enterprises will generate a greater risk to their own financial liquidity and, therefore, is not conducive to enhancing ESG performance. Media attention, stock bar attention, financing constraints and marketization all have a driving effect on corporate ESG performance, indicating that the pressure of media opinion and public evaluation faced by companies, as well as the pressure of financing constraints, will motivate companies to fulfill their ESG responsibilities and improve their ESG performance, and local market development can also provide companies with a good competition mechanism, which can inject a source of ESG performance improvement for companies. Among the internal characteristics variables, the shareholding ratio of the largest shareholder has a significant inhibitory effect on ESG performance, which may be due to the externalities of ESG responsibility, so a higher concentration of equity will lead to decision makers focusing too much on their own interests, resulting in a lack of motivation to fulfill ESG responsibility, the percentage of independent directors, asset-liability ratio, age of listing, state-owned enterprises, institutional investors' shareholding ratio and cash flow ratio all show an enhancing effect on ESG performance, which is consistent with the general perception.

Table 2 Test results of impact factors

Variables	***Coefficient***	***Variables***	***Coefficient***
News	0.7926*** (0.0614)	Top	-0.8556* (0.4738)
Public	0.9219*** (0.0761)	Lev	0.7775** (0.3321)
EPU	-0.0318*** (0.0048)	Age	1.0695*** (0.0865)
SA	0.7941*** (0.2800)	SOE	1.0892*** (0.1394)
Mrk	0.5443*** (0.0312)	Inv	0.0320*** (0.0036)
Ind	0.0260** (0.0107)	Cfr	2.8735*** (0.8126)

Note: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.2 Comparison of Machine Learning Prediction Methods

Based on the above analysis, we select 12 indicators through screening for prediction of corporate ESG performance. To avoid the problem of missing values, we select the data of Chinese A-share listed enterprises from 2008 to 2020, with a total of 10,311 samples, of which 10,208 enterprises are selected as the training set, accounting for 99% of the total samples; the remaining 103 samples are selected as the prediction set, accounting for 1% of the total samples. We use Matlab software for BP neural network prediction analysis, and Python software for random forest, KNN and SVM prediction analysis, for which bootstrap randomness selection "random_state" are set to 2022. For BP neural network model, we select "newff" function, training algorithm is LM method, training times are 1000 times, the learning rate is 0.01, and the minimum error of the training target is 0.00001. For the random forest model, we set the amount of decision trees in the forest to 50000, the maximum depth of decision trees is 20, the number of features in finding the best segmentation is 0.99, and the number of basic classification samples "max_samples" is set to 0.99. For the KNN model, we set the number of nearest neighbors "n_neighbors" to 24. For the SVM model, the penalty factor $C = 5$ is chosen, and the parameters of the RBF function $\gamma = 0.3$. Combining the parameters of the above four models, the prediction results are shown in Figure 1- 4, and the comparison of the fitting degree and prediction error of the four models are shown in Table 3.

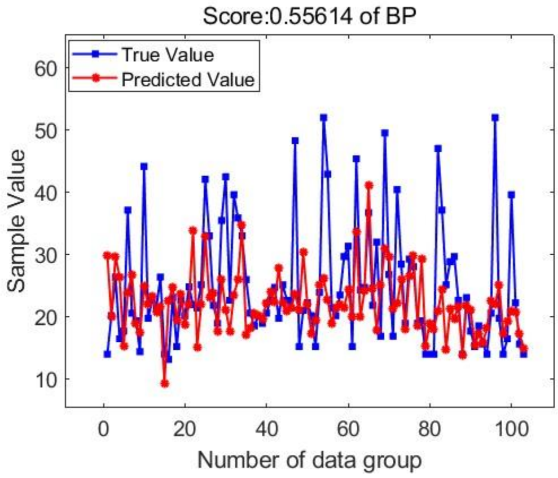

Figure 1. Prediction result of BP

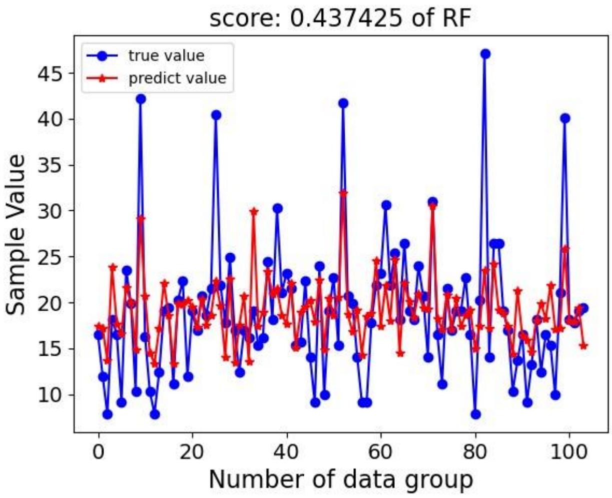

Figure 2. Prediction result of Random Forest

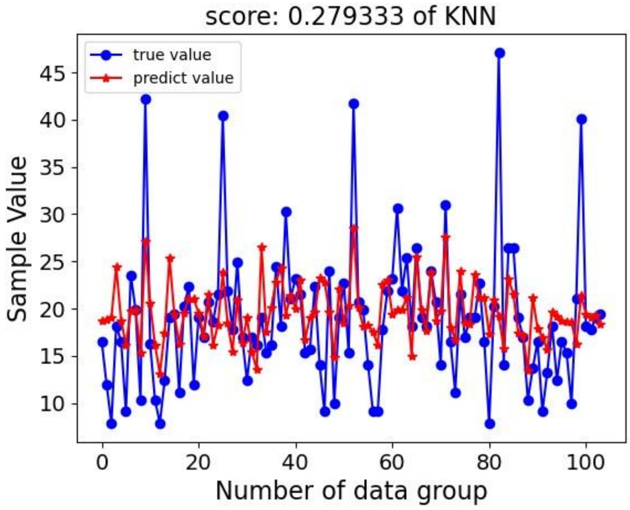

Figure 3. Prediction result of KNN

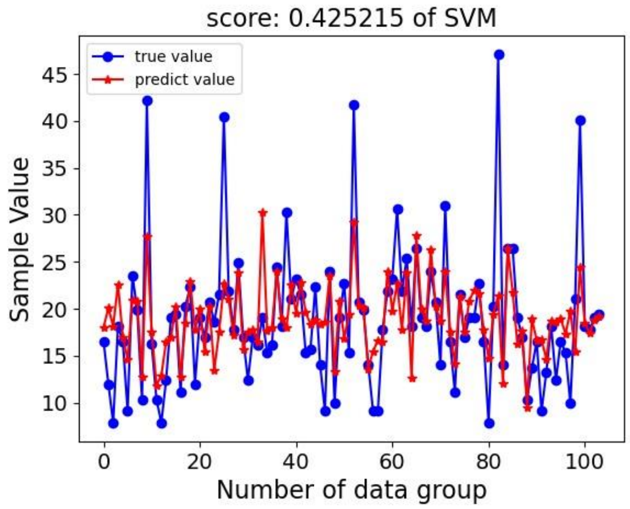

Figure 4. Prediction result of SVM

Table 3 Comparison of machine learning prediction error and goodness-of-fit

Method	***BP***	***Random Forest***	***KNN***	***SVM***
MAE	0.2121	3.9603	4.4201	3.6105
MSE	79.8822	30.1123	38.5742	30.7658
R2	0.5561	0.4347	0.2793	0.4252

From Figure 1 to Figure 4, it can be seen that all four machine learning models can well predict the corporate ESG performance, and the trend of the predicted and true values are basically the same. According to the comparison results of prediction error and goodness-of-fit of the four machine learning methods in Table 1, from the mean absolute error, the BP model has the smallest MAE of 0.2121. The KNN model has the largest MAE of 4.4201, followed by Random Forest and SVM models. From the mean square error, the BP model has the largest MSE of 79.8822. the Random Forest model has the smallest MSE of 30.1123. The SVM model has the second best MSE, which is only slightly higher than the Random Forest model, and the KNN model has a relatively large MSE of prediction. In terms of the goodness-of-fit R^2, the predicted value of the BP model has the highest goodness-of-fit to the true value of 0.5561, the random forest model is second to the BP model, followed by the SVM model, and the KNN model has the worst goodness-of-fit performance. Overall, the Random Forest model can effectively predict corporate ESG based on impact factors.

3.3 Importance of variables

The importance of variables obtained based on the random forest model is shown in Figure 5. It can be seen that the six most important variables are financing constraints, corporate listing ages, media attention, equity concentration, marketability and institutional investors' shareholding ratio in that order. While the three variables with relatively low importance are corporate nature, the percentage of independent directors and economic policy uncertainty, and the differences between the importance scores of these three variables and the rest of the variables are large. Among the three most important variables, two of them represent the external environment, namely financing constraints and media attention, indicating that the financing constraints faced by enterprises can influence their production decisions to a greater extent. Under the financing constraints, enterprises are more willing to fulfill their ESG responsibilities and achieve good ESG performance to gain investors' trust and attract investment. In addition, the pressure brought by media attention also motivates firms to actively fulfill their ESG responsibilities to obtain positive reports and public opinion in order to build up market reputation, which is conducive to long-term sustainable development. Among the top six most important variables, the variables of external environment and internal characteristics of enterprises account for half each, indicating that enterprises should not only consider their own characteristics but also pay attention to the influence of external environment when formulating ESG strategies. Government should also actively improve the financial market environment and media environment to effectively guide enterprises to fulfill their ESG responsibilities.

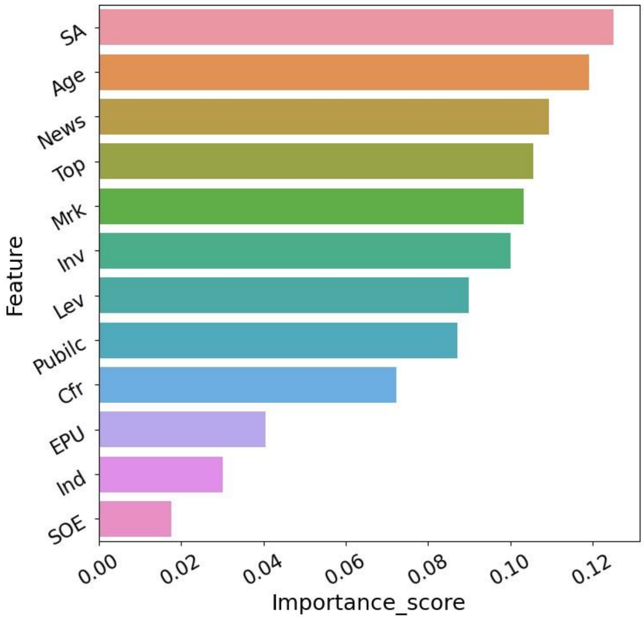

Figure 5. importance of variables based on Random Forest

4. Conclusion

In this paper, multivariate data such as macro data and enterprise data are integrated. On the basis of effective screening of factors affecting enterprise ESG performance by using Lasso method, we comprehensively use a variety of machine learning methods such as BP, Random Forest, KNN and SVM model to predict the performance of enterprise ESG. And then the prediction accuracy of various methods is compared and analyzed to determine the optimal model.

Based on the findings of this paper, we believe that the following aspects need to be done to better motivate enterprises to improve their ESG performance.

At the corporate level, corporate managers should improve their own social responsibility awareness. Listed companies should incorporate CSR performance into the performance appraisal framework of managers, so as to draw managers' attention to CSR and motivate them to better accumulate social capital for the company. At the level of the public and investors, the media and investment institutions should play an active role in promoting and monitoring corporate social and environmental activities. The media should pay more attention to the ESG performance of enterprises, not only to enhance the initiative of enterprises to fulfill their social and environmental responsibilities through reporting and publicity, but also to help enterprises with good ESG performance to build a responsible corporate image and attract the attention of institutional investors, so as to form a virtuous circle of mutual promotion. At the local policy level, the relevant departments should vigorously promote local market development and create a fair and stable policy environment to promote healthy competition among enterprises, so as to inject a source of power for ESG improvement.

The research work in this paper can be expanded in the following aspects in the future. First, the scope of ESG influencing factors can be further expanded and the classification of influencing factors can be refined. Due to data limitations, it is currently impossible to quantify variables such as regional institutional culture and internal corporate culture, which prevent a more comprehensive understanding of corporate ESG influencing factors. Second, the types and ranges of adjustment parameters of machine learning algorithms should be expanded to better fit the data characteristics related to ESG performance of enterprises. Third, the machine learning algorithm should be improved and enriched to enhance the prediction effect of ESG performance.

References

[1] In, S. Y., Rook, D., & Monk, A. (2019). Integrating alternative data (also known as ESG data) in investment decision making. Global Economic Review, 48(3), 237-260.

[2] Awaysheh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. Strategic Management Journal, 41(6), 965-987.

[3] Dong, Q., Wen, S., & Liu, X. (2020). Credit allocation, pollution, and sustainable growth: theory and evidence from China. Emerging Markets Finance and Trade, 56(12), 2793-2811.

[4] Chen, Y. P. V., Zhuo, Z., Huang, Z., & Li, W. (2022). Environmental regulation and ESG of SMEs in China: Porter hypothesis re-tested. Science of the Total Environment, 850, 157967.

[5] Husted, B. W., & de Sousa-Filho, J. M. (2019). Board structure and environmental, social, and governance disclosure in Latin America. Journal of Business Research, 102, 220-227.

[6] Grewal, J. (2021). TEMPORARY REMOVAL: Real effects of disclosure regulation on voluntary disclosers.

[7] Downar, B., Ernstberger, J., Reichelstein, S., Schwenen, S., & Zaklan, A. (2021). The impact of carbon disclosure mandates on emissions and financial operating performance. Review of Accounting Studies, 26(3), 1137-1175.

[8] Alam, M. M., Tahir, Y. M., Saif-Alyousfi, A. Y., Ali, W. B., Muda, R., & Nordin, S. (2022). Financial factors influencing environmental, social and governance ratings of public listed companies in Bursa Malaysia. Cogent Business & Management, 9(1), 2118207.

[9] Aich, S., Thakur, A., Nanda, D., Tripathy, S., & Kim, H. C. (2021). Factors Affecting ESG towards Impact on Investment: A Structural Approach. Sustainability, 13(19), 10868.

[10] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

[11] Kohannim, O., Hibar, D. P., Stein, J. L., et al. (2012). Discovery and replication of gene influences on brain structure using LASSO regression. Frontiers in neuroscience, 6, 115.

[12] Roy, S. S., Mittal, D., Basu, A., & Abraham, A. (2015). Stock market forecasting using LASSO linear regression model. In Afro-European Conference for Industrial Advancement (pp. 371-381). Springer, Cham.

[13] Ziel, F. (2016). Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure. IEEE Transactions on Power Systems, 31(6), 4977-4987.

[14] Guo, Z. H., Wu, J., Lu, H. Y., & Wang, J. Z. (2011). A case study on a hybrid wind speed forecasting method using BP neural network. Knowledge-based systems, 24(7), 1048-1056.

[15] Yu, F., & Xu, X. (2014). A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. Applied Energy, 134, 102-113.

[16] Mei, J., He, D., Harley, R., Habetler, T., & Qu, G. (2014, July). A random forest method for real-time price forecasting in New York electricity market. In 2014 IEEE PES General Meeting| Conference & Exposition (pp. 1-5). IEEE.
[17] Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. Computational Economics, 57(1), 247-265.
[18] Kohli, S., Godwin, G. T., & Urolagin, S. (2021). Sales prediction using linear and KNN regression. In Advances in machine learning and computational intelligence (pp. 321-329). Springer, Singapore.
[19] Zhou, J., Huang, S., Wang, M., & Qiu, Y. (2021). Performance evaluation of hybrid GA–SVM and GWO–SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation. Engineering with Computers, 1-19.
[20] Jones, S., Johnstone, D., & Wilson, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. Journal of Banking & Finance, 56, 72-85.