



**Determinants of Government Support for Enterprises Based on Random Forest Analysis: A Case Study of Western Guangdong Province**

Li Sun<sup>1</sup>

<sup>1</sup>*School of Economics and Management, Guangdong University of Petrochemical Technology, Maoming 252000, China*

**ARTICLE INFO**

**Article History:**

Received: March 12, 2026  
Revised: May 04, 2026  
Accepted: May 28, 2026  
Available Online: June 05, 2026

**Keywords:**

Western Guangdong, Government Subsidies, Enterprise Innovation

**ABSTRACT**

*This paper studies the local enterprises that have received subsidies from Chinese local governments. Based on the 397 enterprises in the western part of Guangdong Province, data from three rounds of applications were collected between 2020 and 2024 for this analysis. The contents of the data set are as follows: type of industry, distribution area, revenue level, number of employees and innovation performance. According to the above indicators, a random forest model was employed in the study to examine the reasons for government subsidy distribution. The Model has an accuracy of 61.72%. Based on the above data, the conditions for support by local governments in China have moved away from mainly focusing on enterprise income to giving more weight to per capital innovation results. Characteristics of the industry and the level of enterprise income are also supporting factors for the evaluation.*



© 2026 The Authors, Published by AIRSD. This is an Open Access Article under the Creative Common Attribution Non-Commercial 4.0

Corresponding Author's Email: [sunli\\_gyig@gdupt.edu.cn](mailto:sunli_gyig@gdupt.edu.cn)

**Introduction**

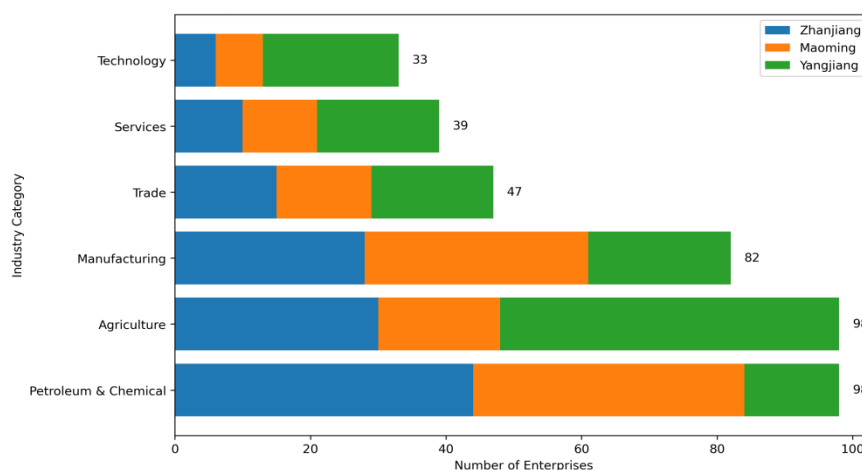
As of 2023, after the economic recession caused by the pandemic subsided in many countries, many developed economies have shown sluggish economic growth since then. A drop in the speed of economic growth and an increase in protectionism across the world have also contributed to this phenomenon. Many small and medium-sized enterprises are facing serious problems in production at present. They have faced increasing production costs and intensified market competition and tightened financing conditions recently, all are major problems for them now (Prasannath, Adhikari, &, 2024). The same as the general rise in staff wages. The local government plans to help the enterprises solve the problems mentioned above through many types of support policies at all levels to stimulate the market, stability fiscal revenue, and maintain employment in the area. The local government is trying to stimulate the economy to some extent.

Looking back on history, the traditional government support mechanism has mainly been based on administrative experience and policy direction, and as a result, has been inefficient

in terms of resource use (Aslam, &, 2023). Some of the limited funds will not be used to support more efficient enterprises; at the same time, too much government support may lead to over-reliance on public funds. Therefore, it has become necessary for local governments to identify enterprises that are genuinely capable of developing, have industrial spillover effects, and genuinely require financing. Strengthen the efficiency of enterprise identification to improve the operation of government services and promote the continuous development of the local economy.

In recent years, the western Guangdong area, which includes the cities of ZhanJiang, MaoMing and YangJiang, has made considerable progress in industrial upgrading, coordinated development of regions and construction of a modern industrial system. Guangzhou and ShenZhen, the first two cities of the Pearl River Delta, are significantly stronger in terms of industrial competitiveness than other areas in Western Guangdong; therefore, technological innovation at this level has not shown a large-scale increase and remains heavily dependent on policy support for economic development. The economy of this area is still relatively concentrated in traditional manufacturing, resource-based enterprises and agricultural products. Therefore, the local government will continue to introduce supporting policies for the industrial structure and employment situation.

**Figure 1: Western Guangdong Province Enterprise Category Classification Chart**



As shown in Figure 1, the industrial structure of enterprises in the western Guangdong area is as follows. The leading components of regional enterprises in this study were petrochemicals and agriculture and food processing; manufacturing and others followed this list. Enterprises in trade, service-oriented businesses and technology-based businesses account for a relatively small proportion. Zhanjiang and Maoming have relatively high concentrations of petroleum-related industries and manufacturing among the regions, and are therefore highly dependent on petrochemical production and heavy industries. Agricultural enterprises in Yangjiang have a higher proportion; thus, agriculture and other primary industries are still relatively large in the local economy. The above regional differences indicate that the enterprises in western Guangdong have complex structures and need more targeted, data-driven government support policies. Given the above circumstances, data-driven analysis methods can be employed to determine the features of enterprises and investigate the impact of government support on specific enterprise attributes.

Most of the previous studies on corporate support and financial assistance have used traditional econometric methods, case studies and survey data. Most of the previous research has studied the effect of financial subsidies on corporate performance and operating

efficiency. However, this study has found that corporate behaviour and government subsidy decisions are often associated with strong non-linearities, multi-dimensional heterogeneity and complex interactions among the variables. Therefore, based on this study, it has been found that the weaknesses of traditional linear analysis in dealing with high-dimensional feature spaces and complex decision-making problems have been exposed.

The Government support policy will consider several factors at the same time before deciding, such as a company's revenue, its industrial characteristics, job creation potential, innovative strength, etc. The above reasons are not in a one-to-one correspondence and are difficult to be described by a simple formula.

Continue to develop big data analysis and artificial intelligence technology, and set new research directions. Recently, many parts of the economy and government have been managed by artificial intelligence. Machine learning models are better suited for dealing with non-linear relationships and features in heterogeneous data compared to traditional econometric models. A new research direction has begun to collect new tools for studying public policy and efficiently allocating fiscal resources.

Based on the above context, this study selected enterprises in the western area of Guangdong Province, China, as the research subjects. The following are some representative indicators of the enterprise that have been selected for this study: revenue, number of employees, geographical location, industry type and innovation indicators. Using a random forest model to integrate enterprise application batch information and government support data in this study. The above model will be used to identify and analyze the general features of enterprises that have received support from the local government. This paper will explore the structure of funds in the local fiscal transfer system and offer some data to improve the efficiency and accuracy of government fiscal resource allocation. Offer some reference materials for other enterprises that need financial support according to government policies.

## **Literature Review**

Most studies to date on government support for enterprises and its impact have been concentrated on fiscal subsidies and other policy support measures. Most scholars agree that government support for financing will reduce the financing pressure, promote technological innovation, enhance the market competitiveness of small and medium-sized enterprises (Abu, N., da Silva, F. P., & Vieira, P. R , 2025). Generally speaking, the reduction in operating risk and enhancement of enterprise resilience achieved by fiscal subsidies in an economic downturn have seen some effect; however, due to rising fiscal pressure and a decline in public funds in recent years, the focus of attention has shifted away from the effectiveness of such subsidies and towards the efficiency and accuracy of the government's distribution system. Given the above reasons, a current problem in enterprise assistance research and public fiscal governance is how to identify enterprises with actual development potential, strong industrial driving forces, and significant financing needs. Government support policies will not be effective simply because they are fiscally generous; we also need to know which people are intended to be helped and have an orderly way of distribution.

Existing research has generally investigated several features of the enterprises that receive government subsidies to classify them, such as the size and operating capacity of the enterprises, their industry attributes, employment impact, and the economy of the region. Among the above factors, enterprise size is generally regarded as one that affects the decision of government support. Generally speaking, a larger company has stronger resource integration capabilities, thus better market competitiveness, and is more likely to be selected

for government subsidies and other policies. As a representative index of enterprise scale and employment absorption capacity, the number of employees has been widely used in research on enterprise management and fiscal policy.

However, some scholars do not think that the allocation of government support is limited to large enterprises. Therefore, the local government understands that various factors affect the overall effect of job creation, such as its development prospects, other regions it will influence, and economic benefits for that area; thus, the goals of their fiscal subsidies may vary. As a result, the relationship between enterprise size and the intensity of government assistance may be non-linear, and other than enterprise size, other operational capacities of enterprises have also drawn considerable scholarly attention. Based on the previous research, it has been found that there is a diminishing marginal utility of enterprise operating revenue with respect to policy subsidies; therefore, increasing fiscal support may gradually reduce the efficiency of resource allocation (Kim, Y., & Park, S, 2022). Other studies have also found that the ability to innovate, employment absorption capacity and sustainable development potential of enterprises are all reasons for inclusion in the list of fiscally supported enterprises (Xu, J., & Yeh, 2022).

The Characteristics of the industry and Regional features are also considered reasons for the extent of government support. Based on the previous studies, most of the fiscal support policies tend to be industrial policies, and there are considerable differences in the strength of subsidies, tax incentives and special funding for different sectors. Given the circumstances of industrial restructuring and high-quality development in the economy, the following types of strategic emerging industries, technology-oriented enterprises, and advanced manufacturing sectors are likely to be included in policy support.

Due to the differences in the level of regional economic development, fiscal capacity and industrial structure and employment pressure, the support plans of the local governments are not the same. Based on the above studies, the governments of the underdeveloped areas generally focus more on stable employment and the stability of the industrial structure, while those in the developed areas are more likely to priority technological innovation and industrial upgrading (Dvouletý, O., Srhoj, S., 2021). Therefore, by dividing the enterprise assistance by area, we can learn why these funds have been allocated locally and what reasons support that allocation.

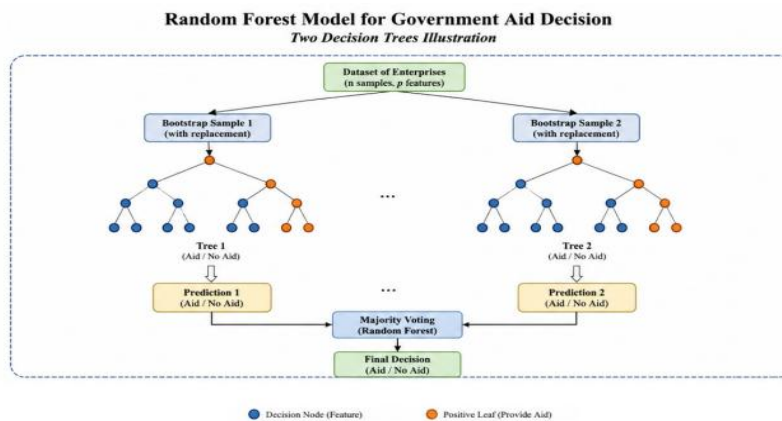
Although the previous studies have offered some basic information on enterprise subsidies and government assistance, most still employ traditional econometric models, questionnaire surveys or case studies. The above methods are generally unable to consider non-linear relations, multi-dimensional features and different characteristics of enterprises. In practice, the allocation of government funds has a variety of complex reasons and factors at the same time, such as a company's revenue, industrial characteristics, innovation ability, employment scale and regional economic conditions (Liu, Y., Li, Y., & Fan, P, 2021).

With the development of big data analysis and artificial intelligence technology in recent years, machine learning methods have also begun to be applied to the research of economic governance, public policy analysis, etc. Machine learning models can handle the non-linear and feature interactions in traditional econometrics more effectively, as well as high-dimensional heterogeneous data. Among the above methods, the random forest model is suitable for the problem of identifying feature importance, handling non-linear classification, and robustness, and is therefore well-suited to analyze complex enterprise assistance allocation mechanisms (Biau, G., & Scornet, E, 2016).

## Methodology

Random forest models have shown some benefits for adaptive decision-making and multi-dimensional feature identification, and these have been studied in recent years. In the allocation of government assistance, many factors are simultaneously involved in determining whether a company receives support, such as its scale and scope of operations, industrial structure, regional economic conditions, employment capacity and innovative capabilities. These factors are often highly non-linear and have varied associations, so traditional linear statistical models cannot accurately explore the allocation mechanisms of local government fiscal support; thus, a random forest model will be used in this paper to analyze enterprise characteristic data and determine the key factors influencing government assistance distribution. Random Forest is a model that builds multiple levels of decision trees to achieve both local feature classification and overall data set evaluation for more stable, high-accuracy enterprise identification. In this study, whether a company falls under the category of key government assistance recipients is a binary dependent variable; that is, "1" indicates that the company is a key assistance target and "0" indicates that it is not.

**Figure 2: Random Forest Model Structure For Government Support Figure**



Based on the research framework proposed in the introduction and literature review sections (Figure 2), several enterprise characteristic indicators have been selected for this study, such as industry type, number of employees, geographical location, enterprise revenue, and related derived indicators. The above indicators are added to the random forest model for an investigation of how enterprise operating conditions affect the need for government support. Random Forest models perform better than those for linear regression on high-dimensional and heterogeneous data; they can detect non-linear feature interactions and assess the impact of all factors in enterprise classification and fiscal resource allocation research thus proving to be suitable. Given that random forest models are relatively easy to interpret, this new way can identify the main reasons why companies choose to work with the government and provide data support for optimizing government fiscal resource allocation and improving the scientific accuracy of policy-making (Mienye, I. D., & Jere, N, 2024). Feature importance analysis and ensemble decision-making mechanisms can further uncover the hidden selection logic of local government enterprise assistance policies in the model.

---

```

Algorithm 1: Random Forest Algorithm for Enterprise Assistance


---


Input: Training set  $D = \{(x_i, y_i)\}_{i=1}^n$ 
 $x$ : Enterprise feature vector
 $y$ : Assistance decision (batch, amount)
Output: Enterprise assistance decision  $\hat{y} = (\hat{y}_{batch}, \hat{y}_{amount})$ 
1 Initialization:  $M = \emptyset$ ;  $t = 1, 2, \dots, n\_estimators$ 
2 for  $t = 1$  to  $n\_estimators$  do
3   Draw a bootstrap sample  $D_t$  from  $D$ 
4   Train a decision tree  $T_t$  using  $D_t$ :
5     At each node:
6       Randomly select  $max\_features$  features from all features
7       Select the best split and split the node
8       Repeat until stopping criteria are met:
9          $max\_depth$  reached (if set)
10         $node\_samples < min\_samples\_split$ 
11         $leaf\_samples < min\_samples\_leaf$ 
12         $impurity\_decrease < min\_impurity\_decrease$  (if set)
13      The trained tree  $T_t$  outputs two targets:
14         $batch$  (classification)
15         $amount$  (regression)
16    Add  $T_t$  to random forest model  $M$ 
17 end for
18 Predicting phase: For a new enterprise  $x$ ,
19   Input  $x$ , into all trees  $T_1, T_2, \dots, T_n$ 
20   Each tree outputs  $(batch_t, amount_t)$ 
21   Final batch decision:  $\hat{y}_{batch} = majority\_vote(batch_1, batch_2, \dots, batch_n)$ 
22   Final amount decision:  $\hat{y}_{amount} = average(amount_1, amount_2, \dots, amount_n)$ 
23 return  $\hat{y} = (\hat{y}_{batch}, \hat{y}_{amount})$ 

```

---

**Algorithm 1 Random Forest Model for Enterprise Assistance**

Random Forests are in fact a way to reduce the variance of a single model and improve its generalization ability by training multiple decision trees on different subsets (Athey, S., Tibshirani, J, 2019). Random forests can reduce the problem of over-fitting that frequently occurs in traditional decision trees during model construction and are relatively robust to outliers; thus, they are more suitable for policy identification studies on complex enterprise sample data. The Algorithm 1 for random forests is as above.

To optimize the node-splitting process of the decision tree in this paper, a recursive feature partitioning method based on the Gini Index is employed to classify the enterprise samples. For each decision node, calculate the increase in Gini impurity due to splitting on various features and select the feature that reduces sample impurity most effectively as the optimal splitting criterion for that node. The node-splitting condition based on Gini impurity can be written as follows:

$$Gini(D) = 1 - \sum_{k=1}^K P_k^2 \tag{ 1 }$$

$P_k$  is the proportion of samples in category  $k$  of the data set  $D$ . A small Gini number is close to 1; it is generally better for classification. Random forests do not need to assume a linear relationship between variables and are not sensitive to multicollinearity, unlike traditional statistical methods that rely on significance tests and linear assumptions. Therefore, the model can recognize the latent structure of government-aided enterprises better based on this feature.

In addition to the Gini coefficient, information gain is also used to measure how well the feature split reduces the impurity of the data in a decision tree. Information gain is the reduction in entropy after dividing the data set and reflects how well a specific variable can explain the class of the samples. The expression for the information gain is as follows:

$$IG(D, a) = H(D) - \sum_{v=1}^V \frac{|D^v|}{|D|} H(D^v), \quad H(D) = - \sum_{k=1}^K p_k \log_2 p_k \tag{ 2 }$$

Given the allocation of enterprise assistance, factors such as the industry type of the enterprise, the number of employees, geographical location, enterprise revenue, and other indicators of innovation may together affect whether the government will provide support through various non-linear paths. Randomly select some features and recursively partition nodes to capture complex interactions among them effectively, thus improving the stability,

robustness and classification accuracy of enterprise assistance identification by a random forest model.

Train the model to train it, then use feature importance analysis to identify and rank the primary reasons for the allocation of government subsidies. Calculate the amount of reduction in node impurity at each split of the decision tree and sum them up to determine the total contribution to the classification accuracy. The formula of the feature importance score is as follows:

$$FI_j = \frac{1}{B} \sum_{b=1}^B \sum_{t \in T_b} \Delta i(s_t, j), \Delta i = i(t) - p_L i(t_L) - p_R i(t_R) \quad (3)$$

$FI_j$  is the importance score of feature J, and  $\Delta i$  is the impurity reduction caused by splitting node i. A relatively larger feature importance value indicates that the corresponding variable is a better indicator of whether a company will be given government financial support.

Through this way, the model can determine whether a company will be designated as a key government-aided enterprise and also uncover which characteristics of the company will be focused on by the local government for fiscal resource allocation. For instance, enterprises with relatively high financing constraints, a large proportion of R&D investment, and relatively low short-term profitability may be assigned higher importance weights in the random forest model. A machine-learning-based approach to policy judgment has been used instead of the old methods of experience accumulation to improve the science, accuracy and dynamism of government aid decisions (Rodriguez-Galiano, V.F., Sanchez-Castillo, 2015).

Finally, in the prediction stage, the output of all decision trees is combined by an ensemble learning method, and the final classification result of enterprise assistance is obtained through majority voting. The function of the Random Forest model for prediction can be written as follows:

$$\hat{y} = \text{majority\_vote}(T_1(x), T_2(x), \dots, T_n(x)) \quad (4)$$

The random forest model is used to classify whether a company belongs to the category of key assistance recipients based on the voting results of all decision trees for a classification task. For the continuous prediction problem of assistance intensity or subsidy level, the final predicted value is obtained by averaging aggregation.

A Random Forest model has been used to combine predictions from multiple models, and it is more stable and generalization than single-model classification approaches; thus, it has reduced the impact of sample fluctuations and over-fitting risk. The model can also handle the problems of high dimensionality and heterogeneity in the data, thus adapting well to track the changes in government-aided enterprises under different economic conditions. Based on the above methodological framework, the following empirical analysis section will further examine the predictive performance, variable influence mechanism and policy identification ability of the proposed random forest model using enterprise sample data from Western Guangdong.

## **Empirical Analysis**

Gathering enterprise data from the RESSET database and question survey for empirical analysis in this study. The experimental environment was built with Python, and a Random Forest model was used in Scikit-learn. The data set was built by taking enterprises in the government assistance program in Western Guangdong between January 2020 and December 2024. A total of 397 enterprises were included in the sample, of which 306 received

government assistance in the three rounds of support, and 91 applied for assistance but were not selected.

The six types of enterprises included in the sample enterprises are Technology, Service, Trade, Manufacturing, Agriculture and Petroleum & Chemical. Given that the regional disparities in local fiscal support are relatively large, nine enterprise groups have been further divided into regional categories based on administrative and industrial spatial characteristics in this study. The three regions of the cities are central urban areas, industrial development areas and other peripheral areas. The central urban area is the main administrative and commercial area of a city, and the industrial development area is generally set aside for manufacturing and industrial production. The other areas are classified as non-central and non-industrial areas.

To expand the feature-representation capability of the random forest model, the enterprise attributes have been organized into the following five categories for data construction: operating income, number of employees, geographical distribution, industry class, and innovation achievement. Among these, regional location and industry category are taken as categorical variables, and operating income and innovation indicators are built as derived quantitative variables from enterprise operational data. The above multi-dimensional feature-construction framework can better capture all the various characteristics of enterprises and the detailed allocation logic for local government subsidies; before the model training process, all enterprise data were standardized and preprocessing to improve model stability and reduce the effect of outliers. Table 1 shows the descriptive statistics of all variables after data preprocessing.

**Table 1: Corporate Feature Information Description Table**

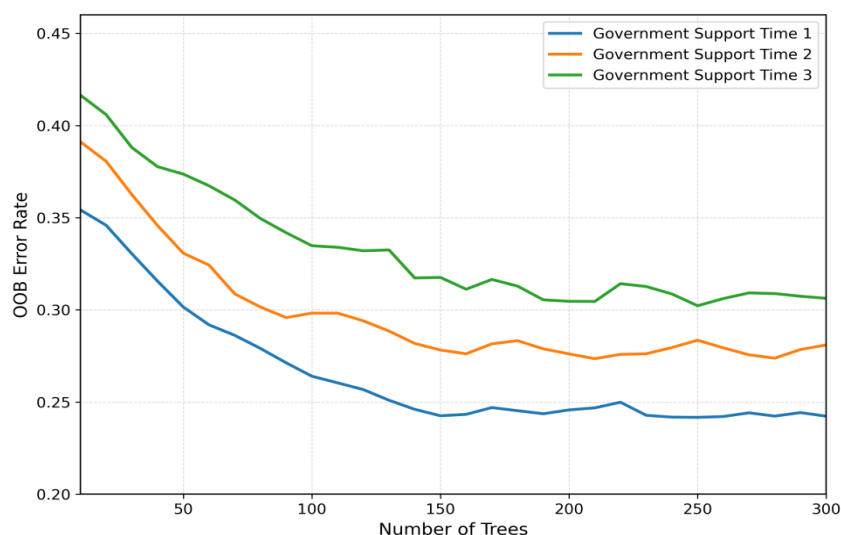
<b>Item</b>	<b>city</b>	<b>Maximum</b>	<b>Minimum</b>	<b>Average</b>	
Corporate Revenue (10K/CNY)	YangJiang	13252.21	22.77	341.57	
	MaoMing	32198.12	51.85	532.17	
	ZhanJiang	36541.32	90.21	413.23	
Number of Employees (Person)	YangJiang	117	1	6.73	
	MaoMing	187	1	11.25	
	ZhanJiang	207	3	10.32	
Corporate Area	YangJiang	9	7	/	
	MaoMing	6	4		
	ZhanJiang	3	1		
Industry of Corporate	YangJiang	5	1		
	MaoMing	5	1		
	ZhanJiang	5	1		
Innovative Achievements (Items)	YangJiang	105	0		5.15
	MaoMing	1514	0		89.17
	ZhanJiang	1054	1		62.32
Revenue Index	YangJiang	181.32	22.77		67.52
	MaoMing	287.17	51.85	111.30	
	ZhanJiang	244.32	30.07	85.53	
Innovation Index	YangJiang	2.31	0	1.23	
	MaoMing	8.10	0	2.27	
	ZhanJiang	5.09	1	1.68	

Based on the enterprise feature data set introduced in the previous section, build a multi-tree random forest framework in this study. The proposed Random Forest model is built on

several levels of decision trees obtained by bootstrap sampling. Each decision tree independently performs enterprise classification based on randomly selected subsets of features during model construction, and then determines the final classification result by majority voting among all trees. Ensemble learning is used to reduce the problem of overfitting and improve the generalization ability of the model in a multi-class environment.

Given the relatively small enterprise sample size and the multi-dimensional characteristics of policy-related variables, we have further constrained the maximum tree depth, minimum node splitting conditions and feature sampling strategy in this study to improve the trade-off between classification accuracy and model stability. An out-of-bag validation method will also be used to test the model during training, and the results of policy identification will be more stable due to the reduction in sample randomness. The Structure of the final model includes enterprise revenue indices, employment characteristics, industrial sectors, regional factors and innovation-related variables to construct a full division system for addressing the interaction effects of local features in the decision-making process of government subsidy distribution.

**Figure 3: Learning Curve of Random Forest Model**

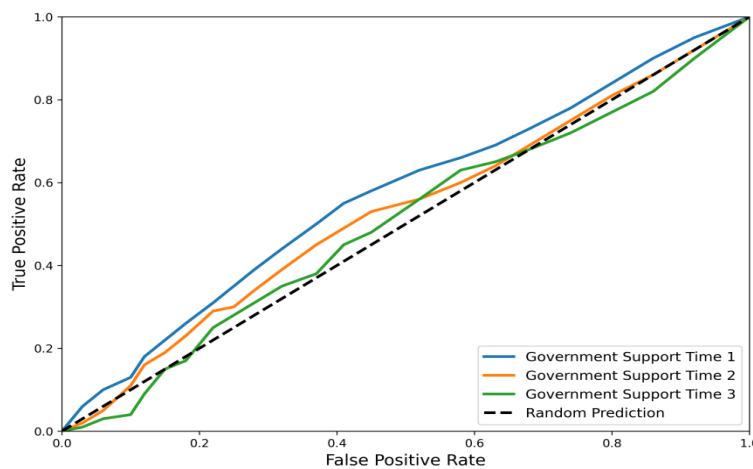


At the same time, to improve the model during training, more decision trees were added in series and the training process was repeated. Figure 3 shows the variation of the error for different groups of government assistance with an increase in the number of trees. As shown in Figure 3, the error rates of all classification groups decrease significantly in the first few rounds of training; therefore, the random forest model can quickly learn the non-linear relations among enterprise attributes and government aid allocation decisions. When the number of decision trees reaches about 150-250, the error curve generally stops decreasing and shows a relative stability, indicating that the model has reached a stable learning state and adding more trees will yield only minor improvements in classification accuracy.

Among the three groups of assistance classification, the first-stage government assistance category has the lowest error rate; that is to say, it shows better classification consistency and a clearer distinction of enterprise characteristics in the early stage of assistance selection. The third-stage assistance category has a relatively larger prediction error; therefore, it is expected that the government's assistance decisions will incorporate more diverse and complex characteristics of the enterprises with increasing time. Based on the results of the learning curve analysis, the proposed random forest framework shows good convergence and

classification stability when dealing with high-dimensional enterprise data, and thus offers a stable foundation for the following policy feature identification analysis.

**Figure 4 ROC curve of Random Forest Model**



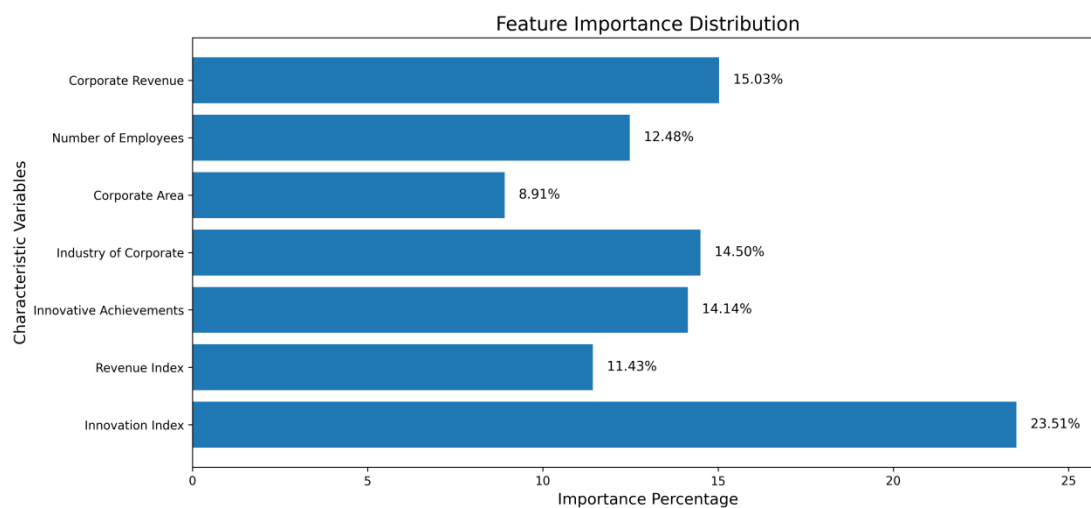
In addition, Receiver Operating Characteristic (ROC) curves were constructed for the various groups of government assistance to further examine the prediction performance of the proposed model. Figure 4 shows the ROC curves of the random forest classifier at the three stages of assistance. The empirical results show that the model has good classification performance at all levels of assistance, and thus it is thought that enterprise operating conditions contain relevant information for local government assistance decisions. The corresponding Area Under the Curve (AUC) values for the three stages of assistance are 61.72%, 58.26% and 54.31%, respectively. Among the above groups, the first-stage government assistance category has the best classification performance and an AUC greater than 0.60; therefore, this model is relatively stable in distinguishing enterprises selected during the early-stage support. On the other hand, the second-stage and third-stage assistance groups show relatively weak predictive performance and have gradually decreased AUC values towards the random classification benchmark.

A drop in AUC values at all levels of support suggests that the decision-making standards for early-stage government support are relatively loose, and only readily available enterprise data such as operating income, industrial structure, scale of employment, and innovation indicators are used. Given the above circumstances, the characteristics of the enterprises are relatively different in structure, and the random forest model can identify the policy allocation pattern more precisely. However, as the government's support programs have advanced, the enterprise selection system is now relatively complicated and includes more administrative considerations, regional coordination demands, fiscal balancing strategies, etc. Therefore, the division lines for assisted and non-assisted enterprises have become relatively vague, and the whole model's discriminative power has decreased.

Although the general ROC performance of the classifier is not very high, it is still feasible in the practical application of local fiscal affairs. Government support allocation is affected by all sorts of non-financial conditions at the level of society and policy, not just the economic indicators of the enterprise that are easy to count. Therefore, even with AUC values ranging from 54.31% to 61.72%, the proposed random forest framework is still relatively good at uncovering the hidden structural patterns in the decisions on government assistance distribution.

Figure 5 is the distribution of feature importance produced by a random forest model. Based on the above tests, the weight of different enterprise attributes in determining the amount of government financial support varies significantly, and among all of them, the weight of the Innovation Index is highest at approximately 23.51 per cent. Therefore, local governments in Western Guangdong have been increasing their support for enterprise innovation by focusing on those with stronger long-term development potential. At present, the innovative indicators are giving greater weight to the extent of government support for decision-making than before, under the former revenue-based assistance model. Enterprise operating revenue and industrial characteristics also have relatively high importance levels, at 15.03% and 14.50% of the total feature contribution, respectively. Based on the above analysis, local governments are still identifying essential beneficiaries for aid from the perspective of enterprise size and industrial strategy. The manufacturing, petrochemical and high-tech industries in the company will likely be selected for the policies because they have a good industrial spillovers and are important to the region's economy.

**Figure 5: Comparison chart of the importance of enterprise characteristics**



New technology and new jobs are also one of the reasons for the increase in cases. Innovation achievements account for 14.14% of the total feature importance, and the number of employees is at 12.48%. Therefore, the two aims of the local fiscal governance policy are to promote technological upgrading and stabilize employment at the same time. Regional variables have a relatively small number of importance scores, however. Therefore, although there are regional differences in the implementation of local fiscal policies, the primary reasons for a company receiving government support are still enterprise-specific operational characteristics.

Generally speaking, according to the feature importance analysis, the local government assistance policies in Western Guangdong are gradually shifting from the old model of scale-based subsidies to a multi-dimensional evaluation system that considers innovation ability, industrial contribution and long-term development potential. Based on the above empirical studies, machine learning can be employed to uncover the underlying decision-making logic of fiscal resource allocation by the government more effectively.

## Conclusion

A Random Forest-based enterprise identification framework was built in this paper to study the distribution method of government subsidy policies. Collect several enterprise attributes, such as operating revenue, number of employees, industrial structure, geographical location,

innovation achievements and derived combined indices, and then build a model to investigate the reasons behind government support allocation decisions. Based on the above research, the random forest model showed relatively stable prediction results in the problem of identifying company support, and its maximum AUC reached 61.72%. The above machine-learning framework is more likely to handle non-linear relationships and multi-dimensional feature interactions, as well as the heterogeneous structure of enterprise data, than the traditional statistical methods.

An ensemble of several decision trees is employed in random forest to reduce the risk of over-fitting and enhance the generalization power and stability of the model. According to the analysis of feature importance, the two reasons most likely to be associated with government support are innovation-related indicators and enterprise operating revenue, and their combined weights are 23.51% and 15.03%. Based on the above results, the policies of local government assistance in Western Guangdong are gradually moving away from the old mode of scale-based subsidy distribution to an all-weather assessment system that considers enterprise innovation capacity, operating stability and extended development prospects.

In addition, factors related to industrial characteristics and the scale of employment also have a relatively strong explanatory power in the classification framework; therefore, it can be concluded that local governments are still pursuing industrial policy orientation and employment stabilization objectives when distributing fiscal support. In short, the randomly forest framework proposed in this paper has strengthened the scientific basis and quantitative accuracy of enterprise assistance identification, provided empirical support for optimizing the efficiency of fiscal resource allocation, improved policy targeting, and strengthened the early warning system for enterprise risks under all kinds of economic changes.

There are still some deficiencies in the current study. Due to data limitations, the present model mainly uses observable enterprise operating indicators, and other significant factors such as financing ability, debt structure, management quality, ownership characteristics, and external market conditions have not been included in the analysis system to date. In the future, more dimensions of the enterprise's financial indicators can be added to the model, and dynamic panel data can be employed to improve both the explanatory power and the predictive ability of the government support identification model. At the same time, further research will be carried out on how changes in government support policies and enterprise conditions are affected by different macro-economic environments and stages of industrial transformation. Based on the above studies, managers of enterprises know better which features the government's support policies promote more strongly, such as investment in innovation and development, and continuous operation. At the same time, the local government will strengthen a mechanism of fiscal support through multiple targeted policies, such as financing subsidies, technological innovation incentives, talent support programs and differentiated industrial guidance policies, to help enterprises facing temporary operating difficulties that are still economically viable and have long-term development prospects.

### **Fund Information**

1. The funding of Science and Technology Planning Project of MaoMing city (Grant No. 2022055)
2. The funding of Fundamental Research Start-up Funds from Guangdong University of Petrochemical Technology ( Grant No.2019rc092)

### **References**

- [1] Abu, N., da Silva, F. P., & Vieira, P. R. (2025). Government support for SMEs in the

- FinTech era: Enhancing access to finance, survival, and performance. *Digital Business*, 5(1), 100099. <https://doi.org/10.1016/j.digbus.2024.100099>
- [2] Prasannath, V., Adhikari, R. P., Gronum, S., & Miles, M. P. (2024). Impact of government support policies on entrepreneurial orientation and SME performance. *International Entrepreneurship and Management Journal*, 20(3), 1533–1595. <https://doi.org/10.1007/s11365-024-00993-3>
- [3] Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- [4] Aslam, R., Awan, T. M., & Khan, M. A. (2023). Investigating the relationship between government support, financial literacy, and SME performance. *Journal of Business & Industrial Marketing*, 38(11), 2379–2395. <https://doi.org/10.1108/JBIM-01-2022-0024>
- [5] Kim, Y., & Park, S. (2022). Government support and innovation performance of SMEs in emerging industries. *Technological Forecasting and Social Change*, 182, 121846. <https://doi.org/10.1016/j.techfore.2022.121846>
- [6] Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148–1178. <https://doi.org/10.1214/18-AOS1709>
- [7] Mienye, I. D., & Jere, N. (2024). A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access*, 12, 102345 – 102378. <https://doi.org/10.1109/ACCESS.2024.3416838>
- [8] Liu, Y., Li, Y., & Fan, P. (2021). Regional inequality and coordinated development in China. *Habitat International*, 113, 102373. <https://doi.org/10.1016/j.habitatint.2021.102373>
- [9] Xu, J., & Yeh, A. G. O. (2022). City repositioning and competitiveness building in regional development: New development zones in China. *Cities*, 120, 103469. <https://doi.org/10.1016/j.cities.2021.103469>
- [10] Dvouletý, O., Srhoj, S., & Pantea, S. (2021). Public SME grants and firm performance in the European Union: A systematic review of empirical evidence. *Small Business Economics*, 57(1), 243–263. <https://doi.org/10.1007/s11187-019-00306-x>
- [11] Rodriguez-Galiano, V. F., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804–818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>
- [12] González-Tejero, C. B., & Molina, C. M. (2022). Training, corporate culture and organizational work models for the development of corporate entrepreneurship in SMEs. *Journal of Enterprising Communities: People and Places in the Global Economy*, 16(1), 168–188. <https://doi.org/10.1108/JEC-12-2021-0178>