

Research Article

Application of the Tax Policy in the Free-Trade Zone Based on Big Data and Internet of Things Technology

Hui Deng  and Dongwen Xie 

College of Accounting, Zhanjiang Science and Technology College, Zhanjiang 524094, Guangdong, China

Correspondence should be addressed to Dongwen Xie; 2003020127@st.btbu.edu.cn

Received 21 June 2021; Revised 24 July 2021; Accepted 13 September 2021; Published 1 October 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Hui Deng and Dongwen Xie. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of the Internet of Things and big data technologies and the rise of smart cities, more and more Internet of Things and big data technologies are applied in the economic field, and the construction of the Shanghai Pilot Free-Trade Zone will be important for my country's follow-up construction of free-trade zones and the whole country. Its economic development strategy has a significant impact, and it has also become a hot spot for domestic and foreign research. Among them, the use of big data and the Internet of Things application of the free-trade zone tax policy system for research cannot be ignored. This article puts forward the literature in the process of writing research method, quantitative research method, and case research method. We used the literature research method to study the theory of big data and tax informatization, used the quantitative research method to conduct data statistics on the tax system of the free-trade zone to understand the information construction of the tax authority, and used the case study method to realize the "data quality monitoring platform." Taxation informatization management in turn brings enlightenment and suggestions for big data to promote the taxation informatization construction of the free-trade zone. A total of 2,854 samples of tax data were extracted from the system of the State Taxation Bureau of a certain district using SQL statements, when the tax burden rate threshold value selected by the nodes of the tax burden discrimination decision tree for all industries is 2.23%, through experiments, various tax collection, and management. The work has achieved ideal results, providing a more scientific and persuasive argument for the proposal of countermeasure research.

1. Introduction

1.1. Background. With the development of economic globalization, economic cooperation between countries in the world has become more frequent, and economic relations have become closer and closer. The trend of trade liberalization has become the basic consensus of many scholars. Trade liberalization refers to the gradual reduction of restrictions on the import of foreign goods and services by a country, the process of providing preferential trade treatment for imported goods and services, and advocating market oriented, whether it is the previous General Agreement on Tariffs and Trade or the WTO. Organizations are all aimed at trade liberalization. China has become the new world production and international trade center, but under the influence of the domestic system and global market mechanism, the economic

structure and economic growth model have not been adjusted, unbalanced, and unsustainable [1]. The establishment of a free-trade zone in Shanghai through pioneering research has not only accumulated valuable experience for the transformation of many coastal areas but also has been very beneficial to domestic approval reforms and related investment system reforms. This is also a survey. It makes people realize the means and functions of big data and Internet of Things technology in government regulation. Big data and Internet of Things technology have become important tools for government managers to implement management and investigation, an important tool, an important basis for decision making. At the same time, it is also conducive to promoting the establishment and improvement of an open economic system on this basis and adapting to global economic and trade development trends [2].

1.2. Significance. In the context of big data and Internet of Things technology, the construction of tax informatization comprehensively uses modern scientific management concepts, coordinates information resources, integrates tax business application systems, and strengthens tax management system planning, construction, and management models. The construction of big data and Internet of Things technology tax information will realize a complete and standardized tax system, a mature and finalized tax system, a high-quality and convenient service system, a scientific and rigorous collection and management system, a stable and powerful information system, and an efficient and clean organization system, which is an important way for the development of modernization and gives full play to the role of taxation big data in serving national governance, promotes data standardization and quality management, gives full play to the advantages of big tax data, and strengthens the application of value-added data. While improving the efficiency of collection and administration and the level of taxation services, it will more profoundly reflect the economic situation, serve economic and social management and macro decision making, and provide strong support for improving China's governance capabilities. In the context of big data, it is of far-reaching significance to deepen the application research of taxation policy in China's free-trade zone [3, 4].

1.3. Related Work. Zhang Chuanyong's research found that as more and more Chinese and foreign banks enter the China (Shanghai) Pilot Free-Trade Zone (CSPFTZ), strengthening the risk management of offshore banks has become a top priority. At the international level, the supervision of offshore banking is mainly focused on tax evasion and money laundering activities. The existing regulatory framework has established standards to identify countries that provide money laundering facilities and establish penalties. In terms of strategy, it is recommended that relevant regulatory measures be improved in the short term. It is necessary to define the scope of offshore banking business and the regulatory measures for Chinese and foreign banks under an appropriate legal framework. Secondly, it is necessary to establish a risk exposure mechanism and risk assessment system for offshore banking business in the China Free-Trade Zone and implement risk management measures on the basis of on-site supervision [5]. But his research process did not consider comprehensive, so the experimental results are not very accurate. Research of Politou et al. shows that big data and machine learning algorithms have paved the way for the accumulation of large amounts of tax and financial data, which are used to provide consumers with novel financial services or to enhance authority through automatic consistency checks. In this regard, in the past few years, the international and EU policies for collecting and exchanging large amounts of personal tax and financial data to promote innovation and increase transparency in the financial and tax fields have greatly increased. However, the collection and use of a large amount of "big" tax and financial data

also raises concerns about privacy and data protection, especially when these data are fed into clever algorithms to build detailed personal information. This is especially true when automatic decision making has an abnormal impact on private life. In the end, these methods of profiling taxation and financial behavior provide fertile ground for distinguishing between individuals and groups [6]. But their research do not represent most of the research results, so there will be a certain error with the experimental results. The diversified development of e-commerce by Pjw and Kcl [7] makes its corresponding logistics management complicated. However, few studies have explored e-commerce logistics business models through big data analysis. Some investigated and explored e-commerce logistics business models from unstructured big data. Specifically, this work developed a mixed content analysis model to examine the basic knowledge of e-commerce logistics. The empirical results of this model combine the theories of Resource Dependence Theory (RDT) and Innovation Diffusion Theory (IDT) to generate logistics strategies. Although their experimental results are not reliable, one can still learn from their experimental process and results.

1.4. Innovation. The innovation of this article is (1) combining the background of the era of big data, analyzing the existing problems of the current tax policy, and continue to put forward specific countermeasures to improve laws and regulations and build an advanced tax collection and management system. So far, we have completed the macroeconomic research on deepening China's tax policy reform under the background of big data. (2) Aiming at the problems in the collection, analysis, and application of tax-related data in China's tax collection and management under the background of big data, and drawing on international experience, the application of big data technology in China's tax policy is simulated and applied.

2. Basic Theoretical Research on Big Data and Internet of Things Technology

2.1. Big Data Technology. The characteristics of big data are often summarized as four Vs, namely, Volume, Variety, Velocity, and Value. They have a point of view:

It is believed that "the connotation of 'big data'" far exceeds a large amount of data (TB) and the technology for processing large amounts of data or the so-called the 4 Vs cover all things that can be done on the basis of massive data [8]. This means that big data are an unprecedented way to help us obtain tangible products and services or intangibles of great value by analyzing huge amounts of data.

Knowledge ultimately forms the power of change[9]. The key to big data is not to define "what," but to "how to use." Only master the advanced big data technology, use the big data technology to analyze and mine the data, and obtain the undiscovered.

Knowledge and information with great value can win in the era of big data; otherwise, it will be entangled and troubled by big data [10].

2.2. Internet of Things Architecture. The concept of the Internet of Things originated from network radio frequency identification technology (Radio Frequency Identification, RFID). This identification technology makes it possible for people's items in life to access the network and is the first to become popular in the logistics industry. Subsequently, the International Telecommunication Union (ITU) [11] formally proposed the definition of the Internet of Things (IoT) for the first time in 2005. The Internet of Things it proposed can realize diversified communications between any item at any time and any place. Dynamic network: with the development of more and more sensing technologies, such as infrared and 4G, the Internet of Things is defined as a network that connects all objects to the Internet through sensor technology for communication. The current Internet of Things architecture is generally considered to include three layers: perception layer, network layer and application layer, as shown in Figure 1.

The sensing layer is mainly composed of sensing devices, including various sensors such as RFID electronic tags and surveillance cameras. Its main function is to collect information and transmit control signals. The realization process of the perception layer is divided into two steps: the communication module needs to pass identity verification to connect to the local gateway: after the connection is successful, the perception layer can send data through the gateway, or receive control instructions or other information from the gateway, and execute the corresponding Operation.

The main function of the network layer is to transmit data: receive the data of the perception layer and send it to other networks; at the same time, send control instructions to the perception layer. On the one hand, the network layer converts the data obtained by the perception layer through format conversion to facilitate transmission in Internet, 2G/3G/4G [12] or other networks. On the other hand, it obtains data information from other networks and converts the format and then returns it to the perception layer. The network layer generally includes communication network, broadcasting network, GPRS, WiFi network, ZigBee, Ethernet, and so on. The application layer is the terminal that collects data. The application layer first collects sensor data and then stores, analyzes, and calculates them to provide users with specific services: in specific scenarios, send actual control instructions to objects in the perception layer to meet corresponding control requirements. In the final analysis, the application of the Internet of Things realizes a series of corresponding functions, such as intelligent environmental monitoring systems, intelligent transportation systems, and smart home systems.

The system architecture is adopted by this system. The data cloud platform designed in this article is a system framework that can support different types of sensor data management and storage. From an overall point of view, the design of the entire cloud platform includes two parts: client and server, and the server is divided into NodeJS multicore HTTP server and MongoDB database cluster.

As shown in Figure 2, the server receives the data collected from the sensor through the data interface and hands it to the NodeJS multicore HTML server for management.

NodeJS server mainly completes a series of basic operation functions for user information and sensor information and stores user data and sensor data in the MongoDB database cluster. The client accesses the server by calling the data interface and can operate the sensor or view the sensor data in the form of charts and other forms. The sensor data comes from the sensor device. Sensor devices are used as mobile sensing nodes. They can be independent collection units, or multiple sensor nodes can form a sensor network as a unit. As data uplink sensor devices, they collect their own sensor data or image information to complete the corresponding sensing information collection and then upload the collected data to the server; as data downlink sensor devices, they receive data from users through customers. The instructions are sent from the end to the server layer, the corresponding control sensors interpret the instructions and then make the corresponding device operations. Among them, the server-side implementation of the HTML server adopts a multicore architecture based on NodeJS, and a reverse proxy server is designed using NodeJS's own built-in modules [13]. Through this reverse proxy, requests from clients and sensors are distributed to a single-core NodeJS server and then handed over to each single-core NodeJS server to process related data services.

2.3. Data Mining Technology. The contradiction between people's desire for knowledge hidden behind huge amounts of data and outdated analysis tools has become increasingly prominent.

Against this background, data mining technology is booming. Data mining is "discovery-driven," which is an information technology to extract knowledge.

2.3.1. Data Mining Method. There are many data mining methods. This section integrates the four steps of big data processing and data mining tasks and focuses on data preprocessing.

Algorithms for management, classification, and regression and cluster analyses are as follows:

- (1) Data preprocessing is the first step of data analysis to ensure the quality of sample data
- (2) Classification and regression algorithms mainly include Bayesian, decision tree and table, neural network, regression analysis, and other algorithms, such as SVM support vector, K-nearest neighbor classification, and so on
- (3) The main algorithms of clustering analysis are K-means algorithm, improved K-means algorithm, multilevel clustering, EM maximum expectation algorithm, and DBScan density algorithm

2.3.2. Data Mining Algorithm Used in This Article. (1) *Decision Tree Algorithm.* Decision tree is a data mining method of classification and regression analysis. The ID3 algorithm (1979) proposed by Quinlan is the most famous algorithm among decision trees [14]. Later, the ID3 algorithm has been

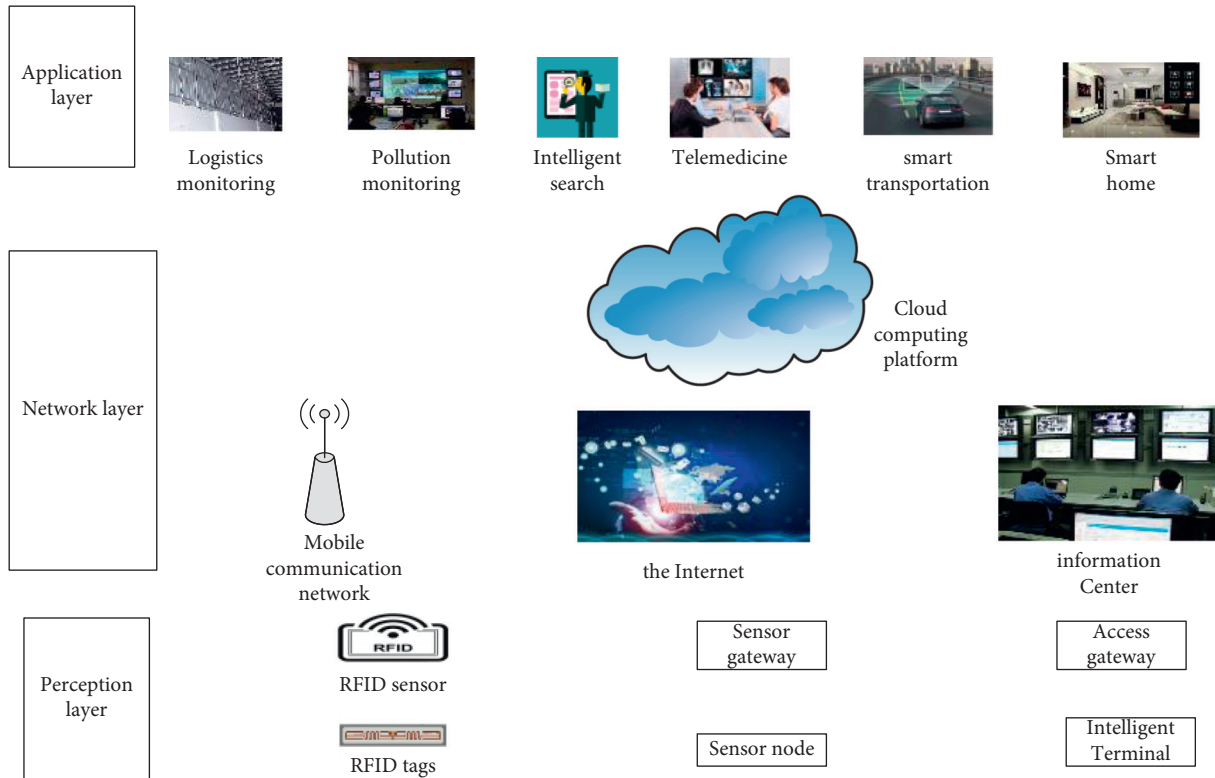


FIGURE 1: Schematic diagram of the Internet of Things architecture.

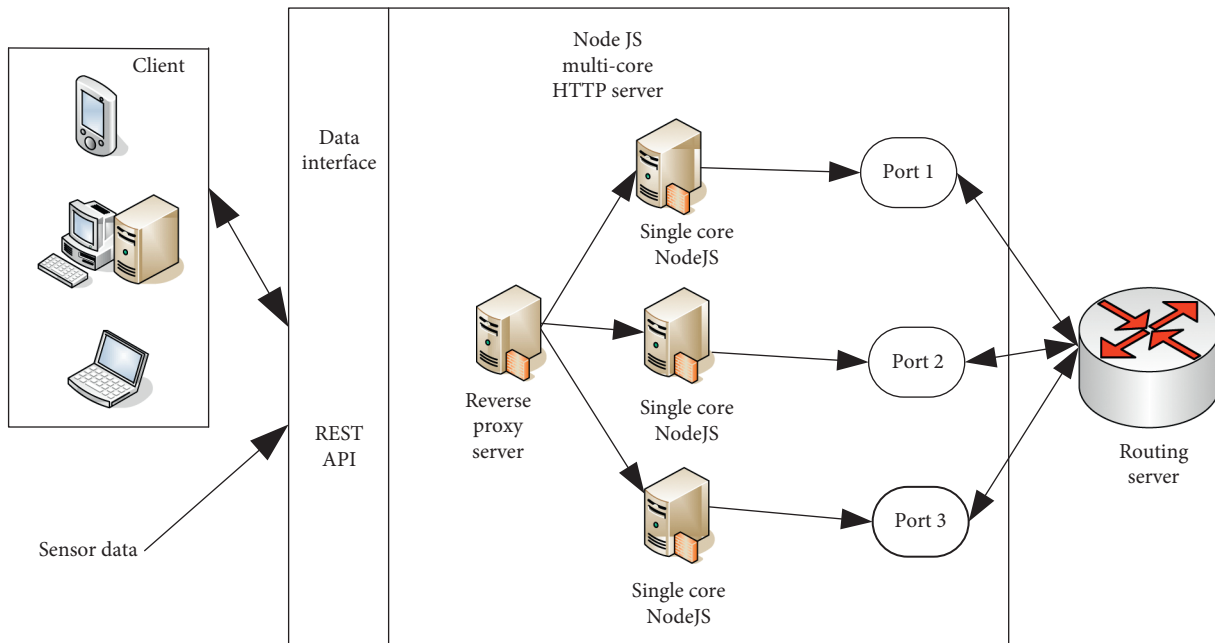


FIGURE 2: The overall system architecture design diagram.

continuously improved to form the C4.5 algorithm with a milestone in decision trees. C5.0 is essentially the same as C4.5. Because C5.0 is a commercial version, its execution efficiency and memory are higher than C4.5 [15]. However, on the one hand, because the classification value of

the model in the article is a numerical attribute, on the other hand, because the details of the C5.0 algorithm have not yet been disclosed, this article chooses the C4.5 algorithm.

The basic idea of the C4.5 algorithm is as follows.

First, all training samples must be established. The C4.5 algorithm selects an attribute from them. This attribute will be applied to each child node. The sample population is classified into a certain class according to the attribute, and the path of each leaf node represents a category. It can be seen that the selection of the node attribute value is a core step of the decision tree algorithm because this attribute value will affect the structure of the decision tree and affect the quality of the rule information found. So how to scientifically select node attribute values? The minimum information entropy contained in the nodes of the decision tree is the guiding principle for the attribute selection of the C4.5 algorithm [16].

The calculation of entropy is shown in the following formula:

$$\text{Ent}(U) = \sum (u_i) \log_2 p(u_i). \quad (1)$$

U_i represents various symbols of the information terminal, and $P(U_i)$ represents the occurrence probability of information U_i ($i = 1, 2, 3, \dots, r$). It can be seen from the formula that information entropy is the mathematical expectation of the amount of information, which represents the average uncertainty before the information source sends out the information [17, 18]. Formula (1) only expresses the entropy of a subset. If you repartition according to a certain attribute, you can get several subsets.

The weighted sum of the entropy of these subsets is shown in the following formula:

$$\text{Ent}U|V_i = - \sum u_i |v_i| \log_2 p u_i |v_i, \quad (2)$$

where V_j is a collection partitioned according to the attribute x . In informatics, the difference between the entropy of the set before the partition and the entropy after the partition is called the gain. The C4.5 algorithm selects the large gain as the decision node, and the calculation is shown in the following formula:

$$\text{Gains}(U, V) = \text{Ent}(U) - \text{Ent} \langle U|V \rangle. \quad (3)$$

Information gain reflects the degree to which information eliminates random uncertainty. According to algorithms and experience, in the case of the same number of samples, the larger the number of groups, the greater the information gain. To solve this problem, if necessary, the information gain rate can be selected as the standard. The mathematical definition of information gain rate is as follows:

$$\text{GainsR}(U, V) = \frac{\text{Gains}(U, V)}{\text{Ent}(V)}. \quad (4)$$

2.3.3. Clustering Algorithm. Clustering is widely used in business, especially to mine various customer information, so tax authorities can also use it to mine taxpayer information. Clustering is a method of summarizing data into several categories based on similarity without prior knowledge. The characteristics of classification are as follows: the data within the groups are similar, and the data between the groups are very different.

There are four types of cluster analysis methods: division method, hierarchical method, density-based method, and grid-based method.

The data mining of the tax system will have higher requirements for the algorithm to find clusters of arbitrary shapes, so the division clustering method is more suitable. The K -means algorithm also has outstanding advantages. It is good at processing large amounts of data and can sensitively reflect abnormal data [19]. This article mainly analyzes the K -means algorithm. The K -means algorithm follows the principle of minimizing clustering performance indicators. The typical clustering criterion is the minimum sum of squares of errors from each sample point in the set to the center, as shown in the following formula:

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2, \quad (5)$$

where E is the sum of squared errors of the sample population of the data set, P is the specific data object, and m_i is the average of the cluster C_i Value; the algorithm seeks a convergent minimum E value for the final result.

2.3.4. Necessity and Feasibility of Applying Big Data in Tax Informatization. Taxation informatization is the deep integration of information technology and taxation business, the reorganization and reengineering of taxation management models, and the creation of a new modern taxation management model [20]. As a representative of advanced information technology, it is very necessary for big data technology to be integrated into the construction of tax informatization. First of all, there is a new understanding in concepts. People's emphasis on informatization has improved, and research on big data is also increasing. The application of big data depend not only in the academic community but also reached a consensus in the tax field. Secondly, there is a guarantee in the system [21]. Finally, there are successful applications in technology at home and abroad that can be used for reference. These provide the feasibility for the introduction of big data into tax information construction [22].

2.4. Construction of the Tax Calculation Model for Natural People. The natural person tax Z is equal to the explicit natural person tax X plus the hidden natural person tax Y . The calculation formula is given by

$$Z = X + Y = P + S + Y. \quad (6)$$

In formula (7), the data of personal income tax P can be obtained directly through query statistics. The key is to determine and calculate the social security fee S of natural persons and the implicit natural person tax Y [23].

Calculation of social security fee S for natural persons. Assuming that s_i ($i = 1, 2, 3$) represents the total income of the first type of social security fund and i represents the type of social security fund, r represents the proportion of each type of social security fund, then the formula for the total social security fee S for natural persons is

$$S = \sum_{i=1}^3 s_i r_i. \quad (7)$$

Calculation of hidden natural person tax Y . The subjects of national economic accounting include financial institutions, nonfinancial enterprises, households, and government [24]. Among them, the financial institution sector, nonfinancial corporate sector, and government sector correspond to the category of legal persons, and the household sector corresponds to the category of natural persons. That is, the preprofit tax in the net household sector production tax L is equal to the implicit natural person tax Y , then the household sector net production tax can be expressed as

$$n_h = L_h + m_h - k_h = Y + m_h - k_h. \quad (8)$$

After shifting formula (8), we can get

$$Y = n_h + k_h - m_h. \quad (9)$$

The difficulty in formula (9) lies in the calculation of the difference between household sector production subsidies and government fees ($k_h - m_h$). To calculate this difference, we first broaden our perspective to the difference between the production subsidies and government fees ($k - m$) of the four major sectors. Here, we introduce the concept of income tax w , which is the sum of the personal income tax and corporate income tax levied by the government on the four major departments. The sum of the income tax w paid by the four major departments and the net production tax n should be equal to the total government tax revenue T minus the difference between production subsidies and government fees ($k - m$), namely,

$$n + w = T - (k - m). \quad (10)$$

After shifting formula (10), we can get

$$k - m = T - (n + w). \quad (11)$$

Here, we assume that in this article, the proportion of the household sector's "production subsidies charges" in its net production tax is the same as the proportion of the four major departments' "production subsidies charges" in its net production tax, which is expressed by the formula:

$$\frac{(k_h - m_h)}{n_h} = \frac{(k - m)}{n}. \quad (12)$$

After shifting formula (12), we can get

$$(k_h - m_h) = n_h \times \frac{(k - m)}{n}. \quad (13)$$

Substituting formulae (11) and (13) into (9) in turn, the calculation formula for implicit natural person tax Y can be obtained:

$$\begin{aligned} Y &= n_h + k_h - m_h = n_h \times \left(1 + \frac{k - m}{n}\right) \\ &= n_h \times \left[1 + \frac{T - (n + w)}{n}\right] = n_h \times q, \end{aligned} \quad (14)$$

where the adjustment factor is

$$q = 1 + \frac{T - (n + w)}{n}. \quad (15)$$

3. Simulation Application of Big Data Technology in Tax Policy

Technology is a double-edged sword, just like the advent of the big data era, it not only brings us various challenges but also brings us ways to solve various problems. Therefore, we can learn Tai Chi's "strengthening" and use big data technology to solve the problems caused by big data. This chapter focuses on the shortcomings of tax collection and management technology in the context of big data, draws on international experience, and combines China's specific conditions to use big data technology to improve the efficiency of tax collection and management, promote the "contribution" to the realization of the "two highs and two lows" goal of tax collection and management, and carry out simulation analysis to provide more sufficient evidence for the countermeasure research below.

3.1. The Role of Data Mining Technology in the Collection of Tax-Related Information. The key to the collection of tax information is the collection of raw data. Under the background of big data, we must achieve comprehensive, active, and high-quality data acquisition in order to master the initiative. To achieve this, we must use big data technology, such as web crawler (bug) technology, for Data warehouse construction. Through the Internet tax-related information monitoring platform, we can achieve precise control of corporate tax-related information and establish a comprehensive database to implement scientific classification management. The main features of the platform are classification capture, real-time monitoring, and intelligent comparison. The platform mainly captures information such as the announcements of listed companies, the reduction of shares of the top ten shareholders of listed companies, and the lifting of the ban on restricted shares from the securities and financial websites: the dedicated computer of the risk control center captures, stores, and sorts relevant information in real time around the clock through the above. The data obtained in this way has a wide range of sources and diversified data. Before data analysis, relevant filtering, cleaning, extraction, standardization, etc. must be carried out. Here, data warehouse technology can be applied to realize the sharing of data from different sources to help the next step of data mining to obtain undiscovered knowledge.

3.2. Application of K-Means Clustering in Tax Source Management. With the continuous development of the economy and society, the ever-changing tax-related behaviors, and the ever-changing competitive environment, it is particularly important to manage in the face of taxpayers' complex and changeable production and operation, low tax compliance, and high tax risks, optimizing China's tax source classification management model. At present, the classification and management of tax sources in China

mainly has problems such as unscientific classification, unsmooth information, and extensive management methods. The introduction of big data technology into tax source management can establish a multichannel information collection mechanism and strengthen data analysis, helping to classify more scientifically. This article uses SPSS Clementine 20.0 version as the data mining platform, takes the *K*-means clustering algorithm as an example, and uses the financial data of Wind Information Finance Terminal (2015) for clustering empirical research.

3.3. C4.5 Decision Tree Application Analysis in Tax Assessment. The shortcomings of current tax assessment in the context of big data can be summarized as the low quality of tax-related information collection and the inadequate data analysis methods and technologies. We can only solve the problems caused by informatization through informatization technology. For this reason, through the learning of the decision tree algorithm in the previous article, this section applies data mining technology to construct a simple and convenient tax assessment model in response to the problems in the tax assessment proposed. It also helps tax assessment to get out of the predicament of being kidnapped by data. At the same time, using the tax data extracted from the system of the State Taxation Bureau of a certain district using SQL statements, the C4.5 algorithm decision tree is

used for application analysis, but based on the length of the article and the need for data confidentiality, only part of the experimental analysis results are shown in this section.

3.3.1. Calculation and Verification of the Tax Rate Threshold.

A total of 2,854 samples of tax data were extracted from the system of the State Taxation Bureau of a certain district using SQL statements. The tax rate calculation process of the furniture manufacturing industry with a large number of same industries and good sample quality is as follows.

The number of samples in the furniture manufacturing industry is 113, and the effective data after preprocessing is 91. Classified according to the attribute of “whether it is an export company,” the sample numbers of export companies and nonexport companies are 35 and 56, respectively. The average tax rate of the sample population is 3.02%. The initial entropy classified by the attribute “whether it is an export company” (U) is

$$\text{ENT}(U) = \frac{35}{91} * \text{LOG}\left(\frac{35}{91,2}\right) - \frac{56}{91} * \text{LOG}\left(\frac{56}{91,2}\right) = 0.961\text{bite.} \quad (16)$$

Then, use the average tax rate to calculate the initial entropy after the threshold classification.

$$\text{ENT}\langle U|V \rangle = \frac{35}{91} * \left(\frac{28}{35} * \text{LOG}\left(\frac{28}{35,2}\right) + \frac{7}{35} * \text{LOG}\left(\frac{7}{35,2}\right) \right) - \frac{56}{91} * \left(\frac{37}{56} * \text{LOG}\left(\frac{37}{56,2}\right) + \frac{19}{56} * \text{LOG}\left(\frac{19}{56,2}\right) \right) = 0.8464\text{bite.} \quad (17)$$

After an iteration of a sample with a tax burden higher than 3.69%, the entropy of the two attributes has increased, and the gain. Both the value and the gain rate become smaller, and the loop terminates. In order to verify the scientific nature of the above classification method, we conduct calculations for all samples regardless of industry. Verification idea: If the above algorithm is correct, iterate over all industries, and the information entropy gain should be significantly reduced.

It can be seen from Table 1 that in the iterative process, the entropy classified by the tax burden rate attribute decreases obviously, and the gain value and gain rate are also decreasing. Therefore, it can be verified that the above algorithm is correct. Therefore, the tax burden judgment decision of all industries is that the tax rate threshold value selected by the tree node is 2.23%, which is also consistent with the average value of the warning value provided by the superior (Table 2).

3.3.2. Application Analysis of *K*-Means Clustering in Tax Source Management. Performing statistical calculations on the experimental part of the data can get the model running result 1. The initial clustering center result is shown in

Figure 3, and the number of accident clustering cases is shown in Table 3.

It can be seen from the running result 1 that according to the difference in attribute value, according to the set cluster number 3, the 188 overall samples are classified as follows: the number of samples in type 1 is 1, the number of samples in type 2 is 3, and the number of samples in type 3 is 182. The other two are missing invalid objects. The phenomenon that leads to the appearance of result 1 is determined by the *K*-means model itself because the *K*-means algorithm will have a great impact on clustering as long as there are outliers in the sample. The running results of this section are well supported. However, the limitations of the algorithm in practical applications are just conducive to quickly identifying abnormal data objects in the sample population and improving the efficiency of supervision. Of course, the existence of isolated points often makes it difficult for clustering to obtain ideal results. Therefore, it is worth considering and researching to eliminate extreme points as much as possible and reduce the impact on concentration. In practice, many iterations are often used to obtain a more ideal grouping result, and each iteration will remove outliers from the previous iteration from the sample. Therefore, in this section, the four extreme values of the running results

TABLE 1: Calculation of tax rate thresholds for all industries.

Number of iterations	Threshold (%)	Sample number of export companies		Sample number of nonexport companies		Entropy		Gain	
		Sf > threshold	Else	Sf > threshold	Else	Attributes	Sf	Value	Percentage
0	2.23	101	45	865	1352	0.3251	0.9562	0.2357	1.2563
1	4.63	34	63	235	235	0.2562	0.4253	0.5325	0.4528
2	6.53	13	22	100	153	0.9632	0.9632	0.17450	0.4542

TABLE 2: Model running result 1-the number of accident clustering cases.

Cluster	1	1000
	2	2930
	3	18200
Effective		189000
Invalidation		1960

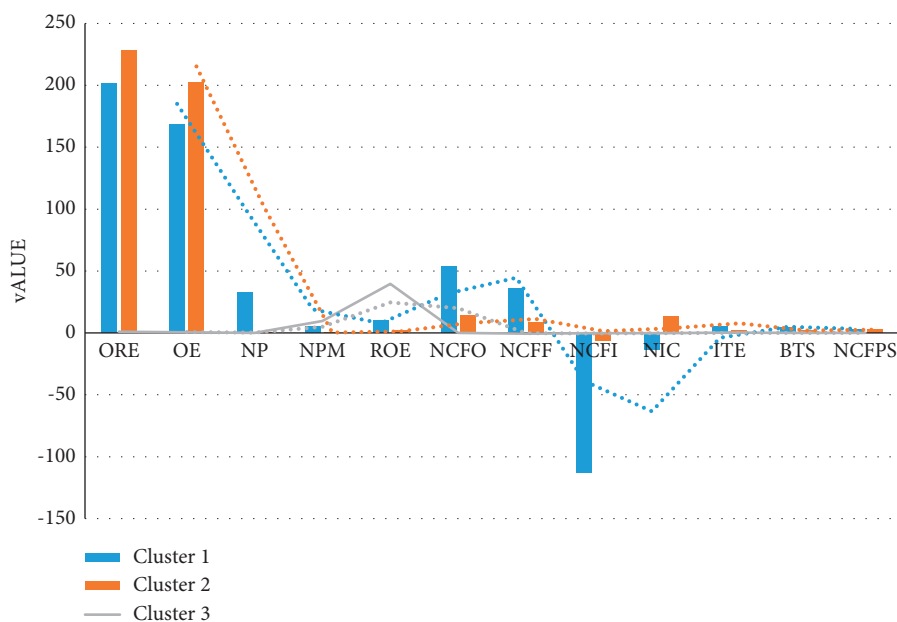


FIGURE 3: Model running result 1-initial clustering centers.

TABLE 3: Model running result 2-the number of clustering cases.

Cluster	1	1000
	2	160930
	3	21200
Effective		189000
Invalidation		2060

are deleted from the sample, and a new grouping is performed. Result 2 in Figure 4 and Table 3 is obtained.

Similar to running result 1, the running result is irrational, and there is also a situation where an object is classified into one category.

Through analysis, it can be seen that Jingwei Textile Machinery is separately grouped into one category mainly because of its operating income (ORE), net profit (NP), net sales margin (NPM), income tax (ITE), and business tax and surcharge (BTS). Both are the highest, and the net cash flow of various types is also the largest, and at the same time, the value of other objects is very different.

4. Preferential Tariff Policy of China-Z Country Free-Trade Agreement and Overall Trade Impact

4.1. Trade Impact of China-Z Country Free-Trade Agreement.

We can find that the total import and export volume in 2015 has declined compared with 2014, from 34.029 billion U.S. dollars in 2014 to 31.976 billion U.S. dollars in 2015. A closer analysis shows that the decline in the total import and export volume in 2015 is mainly caused by the decline in China's imports from country Z. China's imports from country Z dropped from 21.10 billion U.S. dollars in 2014 to 18.680 billion U.S. dollars in 2015, a decrease of 2.33 billion U.S. dollars. At the same time, the bilateral trade deficit narrowed. Combining Figure 5, we can see that the bilateral trade deficit between China and Chile has been shrinking since 2010, from USD 9.91 billion in 2010 to USD 5.384 billion in 2015, an average annual decrease of 11.49%. Although the bilateral trade deficit is shrinking, we can see that

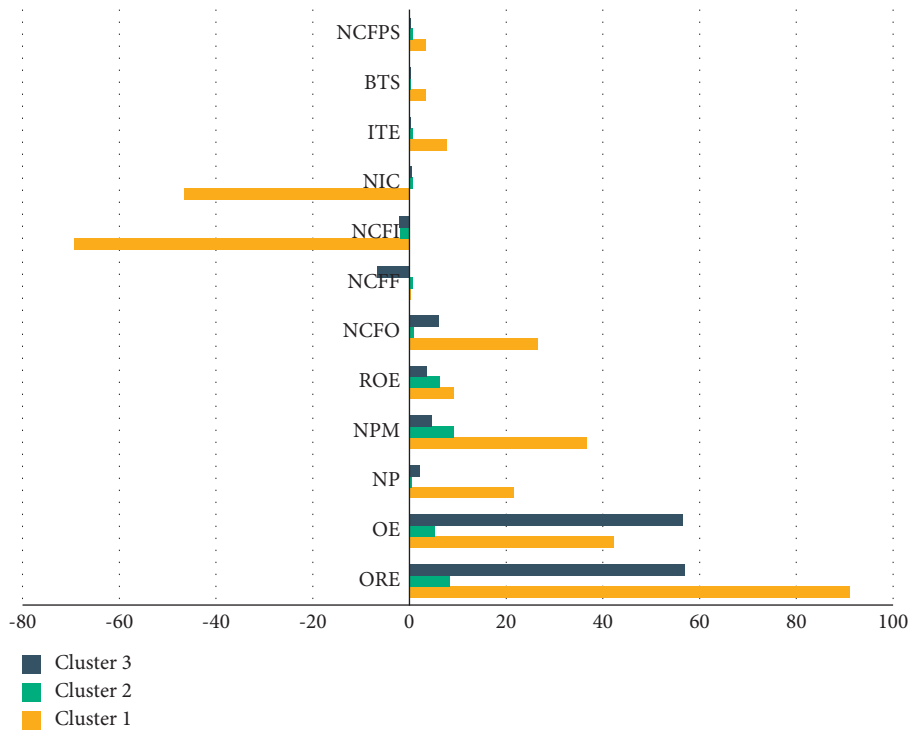


FIGURE 4: Model running result 2-final cluster centers.

China has always been on the side of the trade deficit in bilateral trade between China and Chile since 1999. See Figure 5 for details.

According to Figure 6, we can see that China’s imports from country Z did not decline during the financial crisis, but it rose from 17.935 billion US dollars in 2010 to 18.68 billion US dollars in 2015. We can see that country Z relies heavily on exports to China to revitalize its domestic economy. Since 2010, the growth rate of country Z’s imports to China has been higher than that of country Z’s exports to China. But we have seen China’s exports and imports to country Z have shown a downward trend since 2010. However, the decline in China’s exports to country Z has been relatively flat. At the same time, the gap between China’s imports and exports to country Z is also narrowing. Therefore, we judge that with the economic development of China and Chile and the changes in trade structure, the two countries will be in a state of trade balance.

From the perspective of the trading partners of both parties, China is the largest trading partner of country Z, and it is also the largest import market and export market of country Z. For China, country Z has huge trade potential. In 2015, country Z’s imports to China reached US\$12,948 million, accounting for 22.7% of its total imports. In 2015, country Z’s exports to China reached US\$16,374 million, accounting for 26.4% of its total imports. Table 4 shows the top ten import and export trading partners of country Z in 2015 and their share [25].

As shown in Table 4, the top three import markets of country Z in 2015 were China, the United States, and Brazil. In 2015, country Z imported US\$12,948 million from China, a decrease of about 1%. In 2015, country Z imported

US\$10,789 million from the United States, a decrease of about 20%. In 2015, country Z exported USD 4,897 million to Brazil, a decrease of approximately 14%. In 2015, the top three export markets of country Z were China, the United States, and Japan. In 2015, country Z exported USD 16,374 million to China, a decrease of about 10%. In 2015, country Z exported USD 8,287 million to the United States, a decrease of approximately 5 percentage points. In 2015, country Z exported USD 8,287 million to Japan, a decrease of about 30%.

4.2. Structure of Both Sides’ Import and Export Trade Products Is Highly Complementary. According to the monthly customs table, import and export commodities are classified by HS, and there are 22 categories of commodities. Each category of commodities is divided into chapters, a total of 98 chapters. The left half of Table 5 is the goods imported from China by country Z. It can be seen from the table that in 2015, the first product imported by country Z from China was electromechanical products, which accounted for 37.3%. The second is textiles and raw materials, which account for 15.5%. The third is base metals and products, accounting for 12.2%. The fourth is furniture. Toys and miscellaneous products accounted for 6.3%. The fifth is chemical products, accounting for 6%. The sixth is plastic rubber, accounting for 5.2%. The seventh is light industrial products such as footwear products, accounting for 4.91%. The eighth is sports equipment, accounting for 4.6%. From the commodities imported by country Z into China, we can see that among the commodities imported by country Z into China, mechanical and electrical products account for the largest

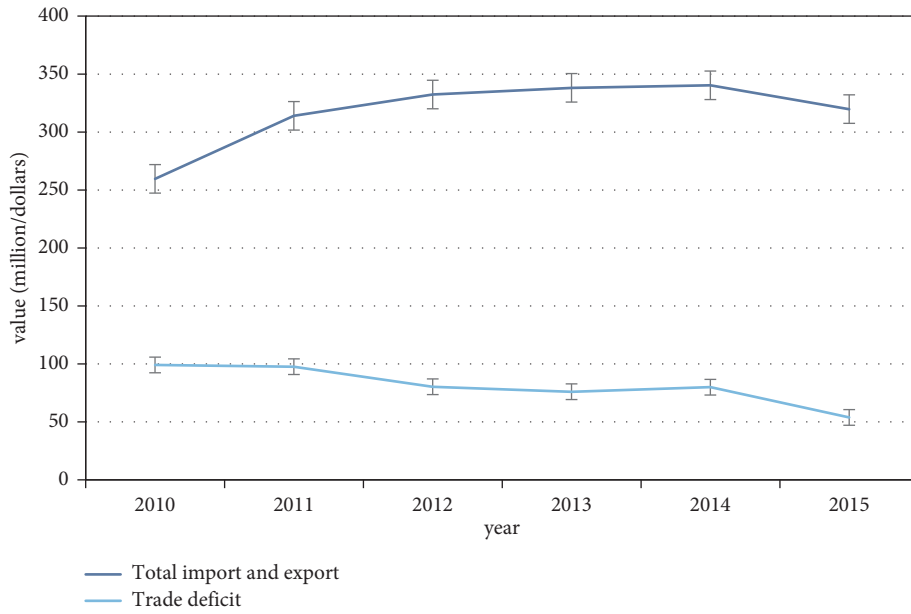


FIGURE 5: 2010–2015 China's trade with country Z.

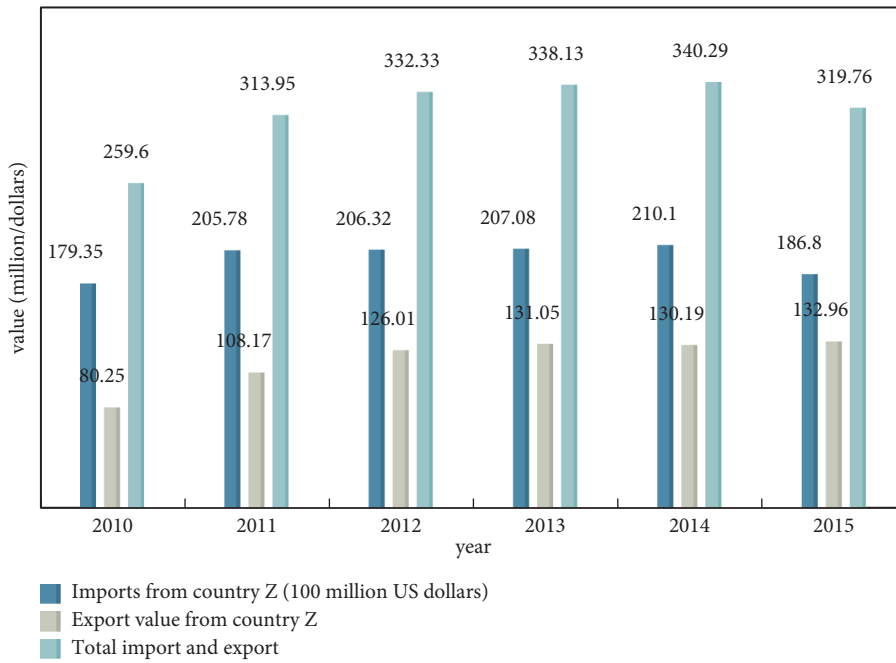


FIGURE 6: China's trade with country Z from 2010 to 2015.

TABLE 4: Top ten trading partners of country Z and their share in 2015.

Import source country/region	Amount	Percentage	Export source country/region	Amount	Percentage
China	12,948.00	22.7	China	16,374.00	26.4
United States	10,789.00	18.9	United States	8,287.00	13.4
Brazil	4,897.00	8.6	Japan	5,491.00	8.9
Argentina	2,669.00	4.7	South Korea	4,105.00	6.6
Germany	2,337.00	4.1	Brazil	3,131.00	5.1
Mexico	2,039.00	3.6	India	2,027.00	3.3
South Korea	1,882.00	3.3	Netherlands	1,600.00	2.6
Japan	1,642.00	2.9	Peru	1,548.00	2.5
Spain	1,495.00	2.6	Taiwan Province	1,432.00	2.3
Others	15,055.00	26.4	Others	16,734.00	27
Total value	56,964.00	100	Total value	62,042.00	100

TABLE 5: 2010–2015 China’s imports from country Z accounted for the proportion of total imports.

Years	China’s imports from country Z	China’s total imports	China’s imports from country Z accounted for the proportion of total imports (%)
2010	179.35	13948.3	1.29
2011	205.78	17434.6	1.18
2012	206.32	18178.3	1.13
2013	207.08	19502.9	1.06
2014	210.1	19602.9	1.07
2015	186.8	16820.7	1.11

proportion. Excluding mechanical and electrical products, there is little difference in the proportion of imports of other commodities.

The main commodities imported by country Z from China in 2015 were mechanical and electrical products, finished textile products and primary textile raw materials, base metals and products, furniture, toys, miscellaneous products, and chemical products. In 2015, the country’s imports of various commodities from China were 4.831 billion U.S. dollars, 1.584 billion U.S. dollars, 820 million U.S. dollars, and 771 million U.S. dollars, accounting for 37%, 16%, 12%, and 6% of its total exports, respectively. In total, it accounts for 77% of country Z’s total imports from China, as shown in Figure 7.

As shown in Figure 8, from the categories of goods exported from country Z to China, we can see that base metals and products, mineral products, paper, plant products, food, beverages, and tobacco are the main products exported from country Z to China. Exports in 2015 were 7993 million U.S. dollars, 5572 million U.S. dollars, 1,056 million U.S. dollars, 645 million U.S. dollars, and 343 million U.S. dollars, accounting for 49%, 34.0%, 6%, 4%, and 2%, respectively.

Although from the perspective of their export commodities, both have base metal products. However, after careful analysis, we found that the base metal products of the two are quite different. The base metals China exports to country Z are mainly iron ore and iron products. However, what country Z exports to China is copper and copper products. Therefore, we can see from the China-Chile bilateral export product structure that the product structure of the two countries is very different and highly complementary.

4.3. Trade Intensity Index between China and Country Z Is High. In order to further quantify the strong trade complementarity between China and Chile and the closeness of trade ties, the trade creation effect and trade transfer effect of China-Chile regional trade cooperation are analyzed. This article will calculate the trade intensity index between China and country Z.

The trade intensity index was proposed by Brown in 1949. Later, Kojima (1964) and others supplemented the statistical and economic significance of the trade intensity index on the basis of Brown in 1964. The trade intensity index is mainly to measure the degree of interdependence between the two countries in terms of trade. It is a comprehensive indicator. The trade intensity index is a

proportional relationship between the ratio of one country’s exports to another country to the country’s total exports and the ratio of another country’s total imports to the world’s total imports. Usually, the trade intensity index is compared with 1. When the trade intensity index is greater than 1, it proves that the two sides of the trade have more trade exchanges. On the contrary, it proves that the trade volume of the two parties is small, and the number of trades is small.

From 2010 to 2015, China’s import trade intensity index from country Z was greater than 1 and reached 2.82 in 2015, indicating that China and Chile have a high degree of trade integration and trade exchanges. China’s import trade intensity index from country Z is high. It also shows that there is a large trade deficit in the trade between the two. In 2012, China’s import trade intensity index from country Z has been rising from the previous level, reaching 2.82 in 2015. It shows that the formal implementation of the China-Z country trade agreement has made China-Z country trade ties closer, but it also shows that China’s imports of country Z products have a certain trade creation effect. It can be seen from Tables 5 and 6.

It can be seen from Figure 9 and Table 7 that since the signing of the China-Z country free-trade agreement in October 2010, China’s export trade intensity index from country Z has been on an upward trend, rising from 1.3 in 2011 to 1.49 in 2015, indicating that China Products exported to country Z have a certain trade creation effect.

4.4. Overall Value of Benefited Goods Is on the Rise. From 2010 to 2015, the bilateral trade volume between China and country Z increased from 16.46 billion U.S. dollars in 2010 to 29.32 billion U.S. dollars in 2015, of which exports increased from 9.95 billion U.S. dollars to 16.37 billion U.S. dollars in 2014, and imports increased from 48. US\$800 million to US\$129.5 in 2014. We can see that since 2010, China’s benefit to country Z has shown an upward trend. From 381 million US dollars in 2010 to 1.627 billion US dollars, an average annual growth rate of 51.42%. It can be seen that as the implementation time of the China-Z country free-trade agreement has been lengthened, country Z’s utilization of the China-Z country free-trade agreement has continued to increase, as shown in Figure 10.

4.5. Tax Reduction and Exemption Increase Year by Year. After the implementation of the China-Z State Free-Trade Agreement, excluding the impact of the financial crisis, tariff reductions and exemptions have been on the rise since

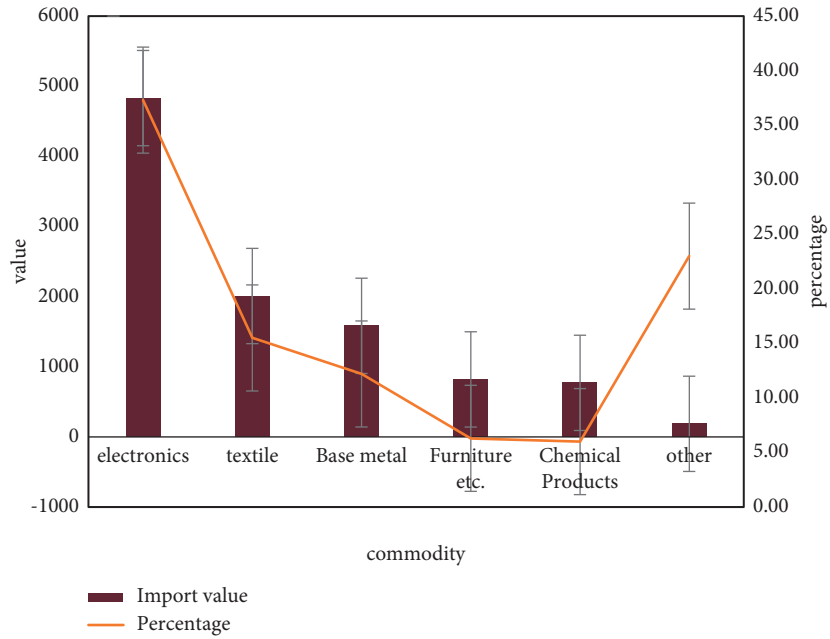


FIGURE 7: The top five commodities imported by country Z from China in 2015.

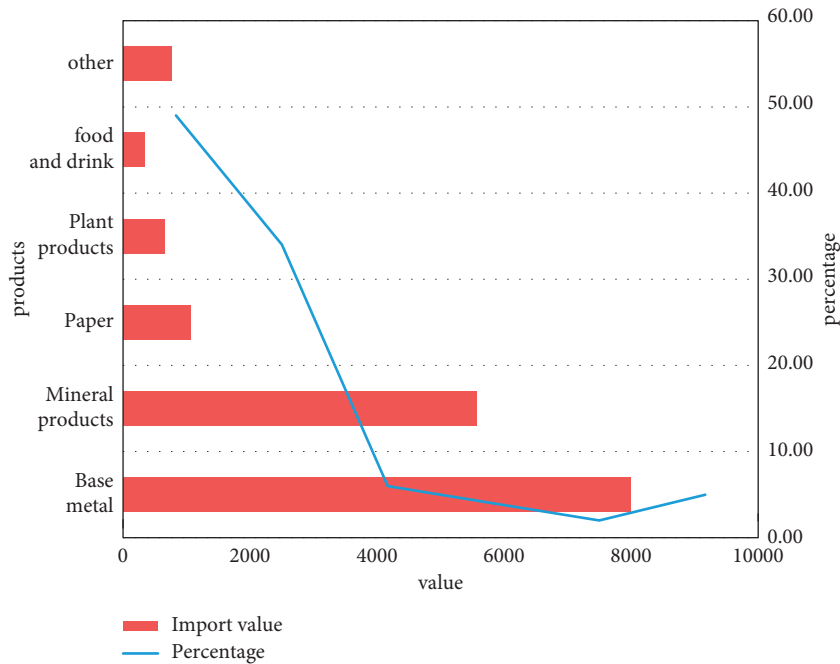


FIGURE 8: The top five commodities exported from country Z to China in 2015.

TABLE 6: 2010–2015 China’s import trade intensity index from country Z.

Years	Total exports from country Z	Total world export	The proportion of country Z’s total exports to the world’s total exports (%)	China import trade intensity index
2010	711.06	150152	0.47	2.72
2011	814.38	179494	0.45	2.60
2012	779.65	179038	0.44	2.61
2013	766.84	185414	0.41	2.57
2014	766.39	184601	0.42	2.58
2015	633.6	160649	0.39	2.82

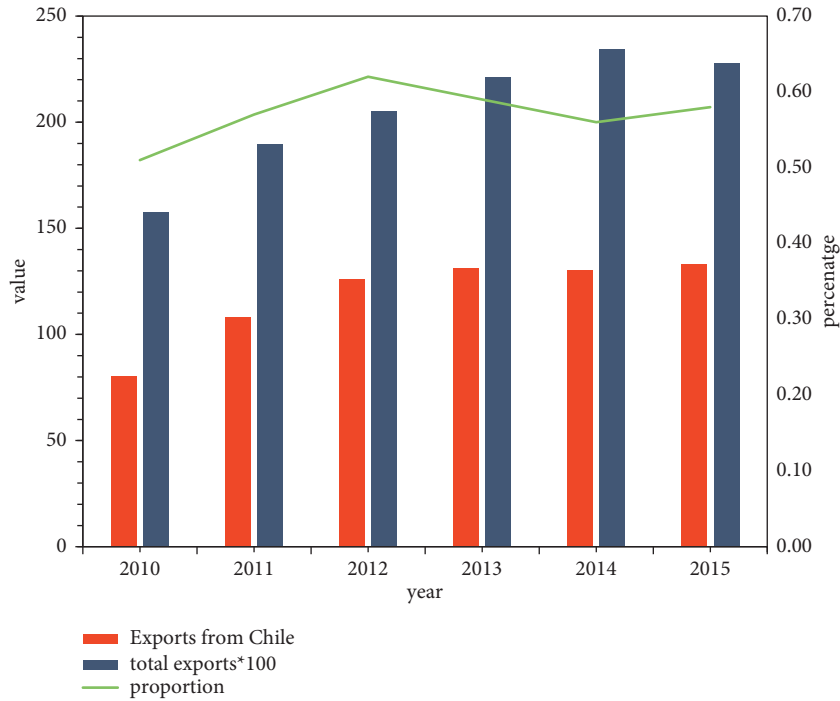


FIGURE 9: 2010–2015 China’s exports from country Z accounted for the proportion of total exports.

TABLE 7: 2010–2015 China’s export trade intensity index from country Z.

Years	Total imports from country Z	Total world imports	Proportion of total world imports (%)	China export trade intensity index
2010	592.07	151456	0.39	1.3
2011	746.94	180203	0.41	1.37
2012	800.67	180510	0.44	1.39
2013	791.73	184246	0.43	1.38
2014	723.44	184294	0.39	1.42
2015	630.38	160646	0.39	1.49

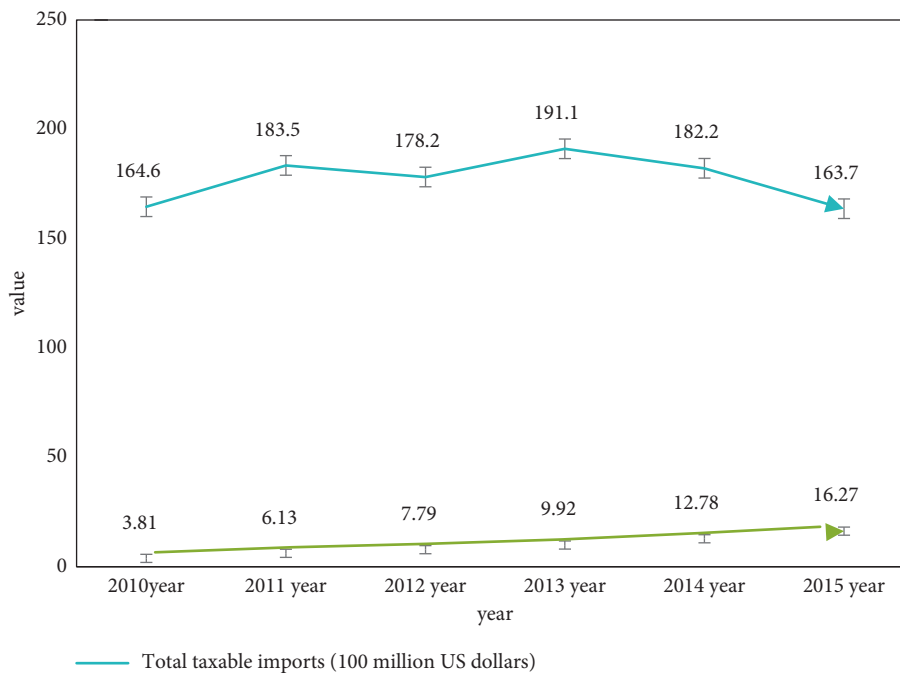


FIGURE 10: The value of China’s imports from country Z during 2007–2015.

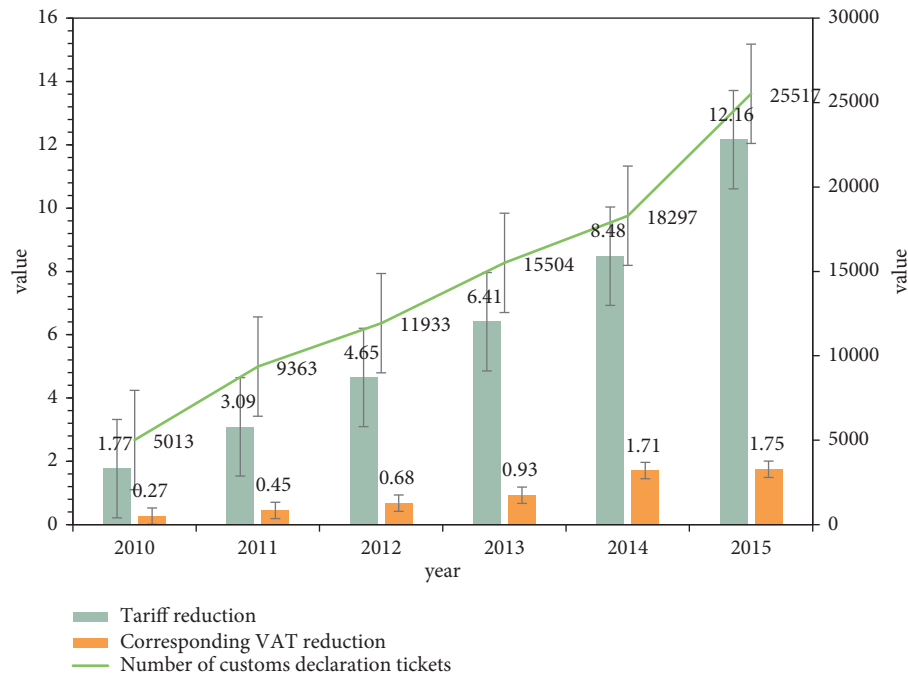


FIGURE 11: Preferential tax amount under China-Z State Free-Trade Agreement 2007–2014.

2010. In 2010, the amount of tariff reduction and exemption was 177 million yuan. In 2015, the amount of tariff reduction and exemption was 1.216 billion yuan. In 2015, the amount of tariff reduction and exemption was 22.94 times that of 2010. From 2010 to 2015, the number of customs declaration votes has been increasing rapidly, from 5031 votes in 2010 to 25,517 votes in 2015. The number of customs declaration votes in 2015 was 21.26 times that of 2009, with an average annual growth rate of 66.45%, as shown in Figure 11.

5. Conclusions

It can be seen from this case that cluster analysis plays an important role in scientific clustering and rapid discovery of abnormal enterprises in tax source monitoring, and it can provide effective guidance for tax officials to manage different types of enterprises in a targeted manner. Data used in this section come from third-party data. Due to the lack of complete informatization of the current tax system, it can only artificially and objectively obtain the required information from the relevant platform. Obviously, when the third phase of the Golden Tax is completed and after being put into use, the use of big data technology proposed in the article can easily solve the current difficult problem of data acquisition. Therefore, it can be said that cluster analysis will help achieve more scientific and professional tax source classification management in the current environment. Then, in the context of big data, based on the completion of the big data platform and on the premise of mastering big data technology, the algorithm of cluster analysis will have more room to play. It can even be said that the era of big data is exactly it. The best time to unleash your great potential.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by University Students Innovation and Entrepreneurship Program of Zhanjiang University of Science and Technology in 2020: A Study on University Students Innovation and Entrepreneurship Tax Policy in the Context of New Crown Epidemic—A Case Study of Foshan, Guangdong Province (Project no. 2021ZKYDCA16).

References

- [1] G. Khatwani and P. R. Srivastava, "Impact of information technology on information search channel selection for consumers," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 63–80, 2018.
- [2] I. Kitouni, D. Benmerzoug, and F. Lezzar, "Smart agricultural enterprise system based on integration of Internet of things and agent technology," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 64–82, 2018.
- [3] S. K. Biswas, D. Devi, and M. Chakraborty, "A hybrid case based reasoning model for classification in Internet of things (iot) environment," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 104–122, 2018.
- [4] H. Hamidi and M. Jahanshahifard, "The role of the internet of things in the improvement and expansion of business," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 24–44, 2018.

- [5] C. Zhang, "Review of managing the risks of offshore banking and its implications for the China (Shanghai) Pilot free trade zone," *Journal of Shanghai Jiaotong University*, vol. 21, no. 01, pp. 33–43, 2016.
- [6] E. Politou, E. Alepis, and C. Patsakis, "Profiling tax and financial behaviour with big data under the GDPR," *Computer Law & Security Report*, vol. 35, no. 3, pp. 306–329, 2019.
- [7] A. Pjw and B. Kcl, "Unstructured big data analytics for retrieving e-commerce logistics knowledge," *Telematics and Informatics*, vol. 35, no. 1, pp. 237–244, 2018.
- [8] X. Lyu, W. Ni, H. Tian et al., "Optimal schedule of mobile edge computing for Internet of things using partial information," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2606–2615, 2017.
- [9] Z. Obermeyer and E. J. Emanuel, "Predicting the future-big data, machine learning, and clinical medicine," *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.
- [10] W. Gang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: certain Investigations for research and applications," *International Journal of Production Economics*, vol. 176, pp. 98–110, 2016.
- [11] E. D. Siew, R. K. Basu, H. Wunsch et al., "Optimizing administrative datasets to examine acute kidney injury in the era of big data: workgroup statement from the 15(th) ADQI consensus conference," *Canadian Journal of Kidney Health and Disease*, vol. 3, no. 1, pp. 1–12, 2016.
- [12] S. Athey, "Beyond prediction: using big data for policy problems," *Science*, vol. 355, no. 6324, pp. 483–485, 2017.
- [13] M. Chen, Y. Ma, J. Song, C. F. Lai, and B. Hu, "Smart clothing: connecting human with clouds and big data for sustainable health monitoring," *Mobile Networks and Applications*, vol. 21, no. 5, pp. 1–21, 2016.
- [14] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the internet of things using big data analytics," *Computer Networks*, vol. 101, no. C, pp. 63–80, 2016.
- [15] I. A. T. Hashem, V. Chang, N. B. Anuar et al., "The role of big data in smart city," *International Journal of Information Management*, vol. 36, no. 5, pp. 748–758, 2016.
- [16] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: from big data to big insights," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 215–225, 2016.
- [17] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: privacy and data mining," *IEEE Access*, vol. 2, no. 2, pp. 1149–1176, 2017.
- [18] S. Akter and S. F. Wamba, "Big data analytics in E-commerce: a systematic review and agenda for future research," *Electronic Markets*, vol. 26, no. 2, pp. 173–194, 2016.
- [19] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: healthcare cyber-physical system Assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2017.
- [20] E. Baccarelli, N. Cordeschi, A. Mei, M. Panella, and J. Stefa, "Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: review, challenges, and a case study," *Computers & Chemical Engineering*, vol. 91, no. 2, pp. 182–194, 2016.
- [21] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of big data based on its essential features," *Library Review*, vol. 65, no. 3, pp. 122–135, 2016.
- [22] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
- [23] E. Bastug, M. Bennis, E. Zeydan et al., "Big data meets telcos: a proactive caching perspective," *Journal of Communications and Networks*, vol. 17, no. 6, pp. 549–557, 2016.
- [24] E. Zeydan, E. Bastug, M. Bennis et al., "Big data caching for networking: moving from cloud to edge," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 36–42, 2016.
- [25] H. Xing, A. Qian, R. C. Qiu, W. Huang, L. Piao, and H. Liu, "A big data architecture design for smart grids based on random matrix theory," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 674–686, 2017.