The Construction of Internal Tax Audit System of Private Enterprises Based on Python

He Li

lihe6616@163.com

School of Public Finance and Administration, Harbin University of Commerce, Harbin City, Heilongjiang Province, China

Abstract: With the background of digital economy era, China's private enterprises have entered a critical period of transformation and upgrading. It is necessary to enhance the internal management and development of enterprises with the help of digitalization and informatization, avoid and deal with many operational risks, and promote the high-quality development of private economy. Therefore, this paper takes the internal tax audit management of private enterprises as the research object, makes full use of the functions and characteristics of network information technology, data mining technology and computer application technology, and relies on data processing class libraries such as Numpy and Pandas in Python environment to build a Web-based internal tax audit system of private enterprises. The whole system adopts B/S architecture design, and the Web Server is built by Flask framework, and the construction of various business logics and the deployment of API data interfaces are completed according to MVC pattern. The function setting of the system will comprehensively cover the whole process of internal tax audit work of private enterprises, which not only meets the needs of current enterprises and auditors, but also can give full play to the working efficiency of data algorithm models such as K-means and Apriori under data mining technology. The construction of the system not only greatly improves the efficiency of internal tax audit of private enterprises, but also promotes the informatization and intelligent development of internal management of enterprises.

Keywords: Private Enterprises, Internal Tax Audit, Python, Data Mining, Computer Mining.

1 INTRODUCTION

As an important part of China's national economy, private enterprises have played an irreplaceable role in economic development, improving people's livelihood, promoting innovation, deepening reform and opening up, and they are the backbone of China's socialist modernization. Since the 19th National Congress of the Communist Party of China, China's private economy has stood at a new historical starting point. It should actively grasp the theme of the times of high-quality development, adhere to the digital economy as the key driving force for transformation and upgrading, and empower the internal operation and management of enterprises through digitalization and informatization upgrading, which can effectively improve the backward problem of enterprise operation risk management mode, improve management efficiency, and ensure the smooth realization of organizational goals. [5]

BDEDM 2023, January 06-08, Changsha, People's Republic of China

DOI 10.4108/eai.6-1-2023.2330318

The internal tax audit is an important part of many audit activities of enterprises, and it is also a necessary means for enterprises to take the initiative to prevent corporate tax risks. The internal tax audit inherits the basic characteristics of the internal audit work of enterprises, and is completely independent of other routine economic activities of enterprises such as accounting and finance. It aims at accounting and auditing the amount of tax paid by enterprises and the amount of tax payable on various incomes through standard audit methods, so as to ensure that enterprises pay taxes in accordance with the tax law in the process of tax payment, and make reasonable tax reduction or exemption. [2] As the front node of the external tax audit, the internal tax audit of an enterprise's level of prevention and control is not only related to the immediate economic benefits of the enterprise, but also can provide the necessary data support for the future development of the enterprise, thus helping the enterprise to make timely decisions and adjustments within the scope of tax policies, and promote the rapid improvement of its own market competitiveness. However, the internal tax audit of private enterprises still faces many difficulties. Firstly, due to the particularity of private enterprises and their own limitations, there are obvious deficiencies in the establishment of management system and the handling of personnel hierarchy; Secondly, the lack of corresponding policies and regulations to support and guide, so that the internal tax audit work is ignored; Thirdly, the professional quality of internal tax auditors in private enterprises is not high, the audit mode is backward, and the routine on-site audit efficiency is low, which makes it more difficult to cope with the current complex electronic data and information, resulting in high labor cost. In view of this, this paper holds that private enterprises should adhere to innovation drive, actively implement the best operation mode of internal tax audit, and make full use of the application advantages of network information technology, data mining technology and computer application technology to build a Web-based internal tax audit system for private enterprises, and put forward a set of feasible solutions for internal tax audit of private enterprises. This system will cover all aspects of audit demand analysis, audit plan formulation, audit plan execution, audit report processing, etc., and apply K-means, Apriori and other data analysis and processing algorithm models under data mining to it, which will greatly improve the working efficiency of internal tax audit, speed up the information construction process and promote the scientific development of internal tax audit in private enterprises.

2 INTRODUCTION OF KEY TECHNOLOGIES

2.1 Web Technology

The essence of Web is a service program running on the Internet, and it is a distributed system. The overall operation of the Web architecture depends on the client and the server, that is, the user makes a "request" to the server through the client and parses the received content; The server is used to save content and "respond" to requests from browsers. [3] According to the Web architecture, all technologies applied in the construction process are collectively referred to as Web technologies.

2.2 Python

The Python is a high-level scripting language that combines interpretive, compiler, interactive and object-oriented. As for Web application development, Python has many mature Web

development template technologies. As a lightweight Web framework, Flask framework is more flexible and extensible than other frameworks. The operation process of Flask framework is shown in Figure 1. When a user sends an HTTP request to the Server through the client browser, WSGI Server forwards the request to WSGI App, which completes logical processing of the request and returns the processing result to the server. The WSGI App can include multiple stacked middleware, which need to implement both the Server and the App at the same time, so it can play a regulatory role between the WSGI server and the WSGI application: for the server, the middleware acts as the application, and for the application, the middleware acts as the server. [1]

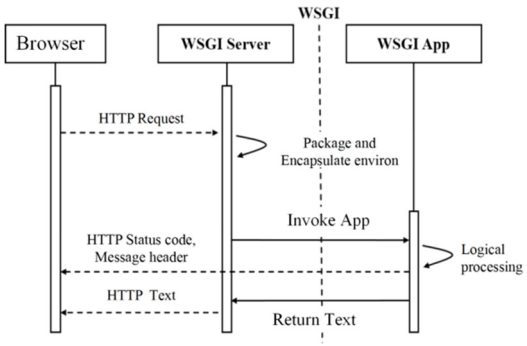

Figure 1: Flask operation flow chart (original)

2.3 Data Mining Technology

As a kind of computer science and technology, data mining is a processing method for big data, which aims to extract information and knowledge that people don't know in advance but have potential usefulness from a large number of, incomplete, noisy, fuzzy and random actual data. [4] The whole process involves three important links: data preparation, data mining implementation, result expression and interpretation, as shown in Figure 2. Among them, the object of data mining, that is, the data source, needs to take into account the characteristics of large quantity, multi-type, authenticity and applicability, and its quality directly affects the accuracy of the final result. The main steps of data mining include: data acquisition, data preprocessing, model establishment and operation, and final result expression.

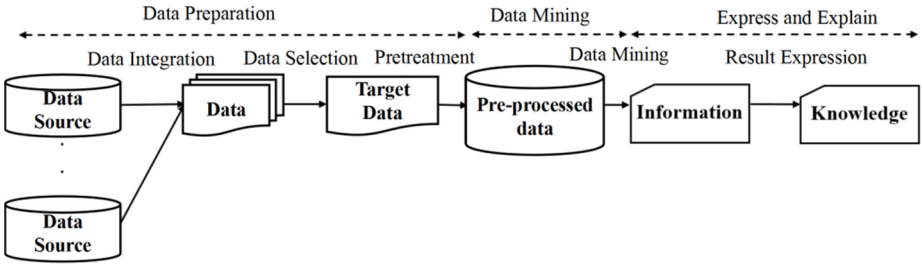

Figure 2: Basic flow chart of data mining (original)

The construction of data mining model is the core of the whole data mining work. The construction of data mining model corresponds to data analysis method, which not only determines the application direction of data mining results, but also determines the construction of data mining model. The common data analysis methods include classification analysis, prediction analysis, cluster analysis, valuation analysis and correlation analysis, as shown in Table 1.

Table 1: Types of Common Data Mining Algorithms

Classify	Data analysis method	Data mining model
Guided data mining	Classification analysis	Decision Tree, Random Forest, Neural Network
	predictive parsing	Regression tree and rough set method
	Valuation analysis	SVM、Bayesian method
Unsupervised data mining	Correlation analysis	Pearson correlation coefficient, Aprior algorithm
	Cluster analysis	K-means clustering algorithm, pedigree clustering

2.4 Development Process

According to the system development requirements and the use requirements of the above key technologies, complete the configuration and deployment of the development environment. The overall design and development environment of private enterprise internal tax audit system is divided into two parts. The first is to use data processing class libraries such as Numpy and Pandas to complete the construction and training of algorithm models related to data mining in Python language environment; The second is to use Flask framework to build a Web Server, and integrate and package each functional module to form a Web application.

Firstly, according to the business process and organizational structure of private enterprises, the tax-related departments and business links are determined, and the selection of tax internal audit data sources and the collection of basic data are completed. The data source is mainly composed of financial information (vouchers, accounts, tables), corporate tax information, some corporate basic information and internal management information. All kinds of target data are summarized through data copying and transmission, and stored in a unified Mysql database system. The data in the whole is in a state of disorder and disorder, which needs to be preprocessed to improve the data quality and provide convenience for subsequent data mining. The data preprocessing mainly includes data cleaning, denoising and basic operation, and finally the basic financial and tax details with relatively uniform format and type, as well as key indicators such as tax rate, quick ratio, asset-liability ratio and expense ratio are obtained.

Secondly, according to the data information, the development of tax risk data mining function module is completed. As the core function module of the system, on the one hand, it can analyze and mine massive data information, and quickly obtain the identification and investigation of tax risks; On the other hand, it can integrate matplotlib data function class library to complete the visual expression of the final result. During the actual development process, the underlying

operating system is Linux CentOS 4.7, and the compilation environment is Visual Studio 2019. The development environment is Pyton 4.2.0, and the Anaconda integrated distribution can be selected for installation and deployment to shorten the time. In addition, the deployment and import of Numpy and Pandas need the help of PyCharm tool. The construction of all kinds of algorithm models can be completely realized by Python code, as shown in Figure 3, which is the key code of K-means clustering algorithm.

```
import random
import numpy as np
import matplotlib.pyplot as plt
k = 3,md = 0,ROUND_LIMIT = 10,THRESHOLD = 1e-10,melons = [],clusters = []
f = open('melons.txt', 'r')
while True:
    for melon in melons:
        c = np.argmin(
            list(map( lambda vec: np.linalg.norm(melon - vec, ord = 2), mean_vectors))    )
        clusters[c].append(melon)
    for i in range(k):
        new_vector = np.zeros((1,2))
        for melon in clusters[i]:
            new_vector += melon
        new_vector /= len(clusters[i])
        change += np.linalg.norm(mean_vectors[i] - new_vector, ord = 2)
        mean_vectors[i] = new_vector
    if md > ROUND_LIMIT or change < THRESHOLD:
```

Figure 3: Implementation code of K-means clustering algorithm(original)

Finally, for the development of the whole system on the Web Server side, the web server chooses Nginx server, version Nginx/Windows -1.12.2, project development language Python4.2.0, development tool PyCharm 2018.3.1 x64 and database MySQL5.7 to complete the construction and support of the system database system. The whole server is implemented by Flask2.0.1 framework. Through the introduction of the above key technical theories, the overall environment of the system development, the configuration of related software and tools are determined, and the technical feasibility of the overall project of the internal tax audit system of private enterprises is also clarified.

3 FUNCTION REALIZATION

3.1 Initiation of Tax Audit

In this function module, the internal audit users of private enterprises can independently complete the comprehensive query of enterprise tax data. Compared with the decentralization of all kinds of financial statements, personnel information, production information and management information under the traditional audit procedures, the system can realize the centralized presentation of all kinds of data information. When the user initiates the tax audit, the system will automatically generate the work project process guide, which will be completed

online from project determination, planning, personnel determination and final audit notice issuance, thus reducing the complicated document exchange of the traditional audit mode and greatly improving the working efficiency of the internal tax audit.

3.2 Tax Risk Identification

In this function module, internal audit users can identify and check the tax risks of related projects, data information and departments according to the tax audit plan. For example, after the user selects the basic information of tax audit department, project, time, etc. in the system in turn, according to the data rules, the tax rate of the project, quick ratio, net interest rate of assets, expense ratio, gross profit margin and other indicators are taken as the key features, and the K-means clustering algorithm model is selected to complete the data classification, thus generating the classification standard, which can be compared according to the standard. If the two groups of data are similar, it can be judged that the tax risk is low, whereas the tax risk is high. Table 2 shows the algorithm formulas of some indexes, and Table 3 shows the variance analysis diagram of some K-means clustering algorithms, in which the significant difference value (Sig) is the criterion for judging whether the enterprise has tax risks, which can reduce the scope of the follow-up on-site audit and improve the efficiency.

Table 2: Calculation formulas of some indexes

Number	Index name	Algorithm formula
1	Tax rate	VAT tax rate: $Q_1 = \frac{M_1(\text{Tax increment})}{\text{K}(\text{Actual sales revenue})}$
		Income tax rate: $Q_2 = \frac{M_2(\text{Income tax amount})}{\text{K}(\text{Actual sales revenue})}$
2	Quick ratio	$W = \frac{SS(\text{quickassets})}{\text{SF}(\text{current liabilities})} \times 100\%$

Table 3: Variance analysis results of K-means clustering algorithm

Index	Cluster		Error		Test value (F)	Significant difference (Sig)
	MSE	Degree of freedom (df)	MSE	Degree of freedom (df)		
Tax burden rate	0.000	1.0	0.000	41.0	0.091	0.761
Quick ratio	0.000	1.0	0.000	41.0	1.667	0.201
Asset load rate	19.414	1.0	0.166	41.0	120.611	0.000
Net interest rate of assets	18.477	1.0	0.315	41.0	58.564	0.000

3.3 Tax Audit Report

According to the tax risk identification and investigation results, auditors can conduct targeted on-site tax audits and make detailed records of tax problems found during the period. After the on-site audit is completed, the user will summarize the data according to the tax audit results and tax categories, and complete the compilation of the Tax Audit Statement in the system to form a preliminary audit opinion. The final tax audit report will be released in the system after the audit is correct.

4 CONCLUSIONS

This paper aims at promoting the internal tax audit mode of private enterprises, and makes full use of the functional characteristics of network information technology, data mining technology and computer application technology to build a Web-based internal tax audit system for private enterprises. The functions of the system will cover all aspects of audit demand analysis, audit plan formulation, audit plan execution, audit report processing, etc., with concise, convenient and efficient operation, the efficiency of internal tax audit will be greatly improved, the accuracy of tax risk investigation and judgment will be increased, the informatization construction of tax audit will be accelerated, and the scientific development of internal management of private enterprises will be promoted.

REFERENCES

[1] Cai Min. (2019) The Web development and application analysis of Python language. Wireless Internet Technology.02.

[2] Fang Lei. (2022) The Strategy of Reducing Enterprise Tax Risk by Internal Tax Audit. Accounting Learning.06.

[3] Li Guanghong. (2020) The Web front-end development and design. Computer & Network.08.

[4]Wang Lili. (2021) The application of data mining technology under the background of big data. Computer & Network.10.

[5] Zhou Lihua. (2021) The bottleneck and realistic path of high-quality development of private economy. Reformation & Strategy.07.

[6] Liu Wuping. (2022) The Application Research of Data Mining Technology. Wireless Internet Technology.01.

[7] Gu Ronglong,Zhao Wenjie etc. (2022) The application of data mining technology in the era of big data. Technology innovation and application.02.

[8] Chen Tianyou. (2019) The Application Research of Data Mining Technology in Tax Audit. Finance and Economics University Of Jiangxi.01.

[9] Cai Guanming. (2022) Suggestions on strengthening internal tax audit in enterprise tax risk prevention and control. Modern Business.02.

[10] Yu Zhong,Chen Jie etc. (2016) A summary of research on the application of data mining methods in tax risk management. Tianjin Economy.06.