

Research Article

Dynamic Audit of Internet Finance Based on Machine Learning Algorithm

Jianhui Zhang 

School of Economics and Management, Xi'an University, Xi'an 710065, Shaanxi, China

Correspondence should be addressed to Jianhui Zhang; zhangjianhui@xawl.edu.cn

Received 17 May 2022; Revised 21 July 2022; Accepted 2 August 2022; Published 29 August 2022

Academic Editor: Mian Ahmad Jan

Copyright © 2022 Jianhui Zhang. 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.

Internet finance is a direction for the development of the new system. The existing research mostly analyzes how to build a risk early warning system from a qualitative perspective to promote a more healthy development of Internet finance companies. This paper aims to study the role of machine learning algorithm in the dynamic audit of Internet finance. It proposes Apriori algorithm, data mining, Bayesian network, and other methods, and related experiments are also conducted on the dynamic audit of Internet financial risk based on machine learning algorithm. The experimental results show that in Internet finance, private enterprises have the highest profit rate, which can reach 20%. But the higher the profit, the greater the financial risk. The machine learning algorithm can make a good intelligent identification of the capital circulation of Internet financial enterprises. This can provide timely feedback on abnormal capital problems, which can help auditors better manage the company's capital circulation.

1. Introduction

The fast improvement of economy and science and innovation has together advanced the improvement of humankind into the data age. In the data age, people can more efficiently collect, store, dissect, and process information of different designs and types to make important choices. Thusly, machine learning calculations arise. In the context of big data, it has become easier and more convenient for people to obtain data, but they still face a situation where there are too much data and useful information is too confusing. For this, people are looking for a tool that can quickly process a variety of different types of data in a short period of time and can be filtered and analyzed according to the requirements of the user. Therefore, the number machine learning algorithms related to this problem arise.

The previous financial risk early warning system has the disadvantages of unity and delay. Through the development of big data and the application of machine learning algorithms, people can realize multi-dimensional financial risk analysis in real time and quickly. Simultaneously, in the period of huge information, organizations are confronting the problem of a lot of information and tough choice

making. Utilizing information mining and monetary gamble early admonition frameworks can actually take care of the troublesome issues of corporate decision production continuously, understand the coordination of monetary administrations, and give groundbreaking plans to corporate directors to simply decide.

With the development of society, machine learning algorithms are also constantly developing. Machine learning algorithms are able to extract value from data with four V's: quantity, variety, speed, and authenticity. The Tawalbeh et al. study discusses the role of network healthcare as well as mobile cloud computing and machine learning algorithms in its implementation, and described a cloudlet-based mobile cloud computing infrastructure for medical big data applications [1]. Also in the application research of cloud computing, Massobrio et al. proposed a big data analysis paradigm related to smart cities using cloud computing infrastructure [2]. Xu et al. look at the privacy issues related to data mining from a broader perspective and study various methods that help protect sensitive information [3]. Zhao mainly focuses on the problem of privacy data leakage in current financial data mining. Data mining and privacy protection are combined with existing machine learning

algorithms for studying [4]. In Internet finance, machine learning algorithms are also very widely used, and a large number of researchers have conducted research on Internet finance itself. In Oruc's research, Oruc and Tatar obtained the structural equation model used by online banking. His model can help banks understand the determinants that affect users, and they can also formulate correct policies to attract customers to use online banking [5]. In the research work of Khattak et al., it is evaluating the security of IBS by developing a framework based on in-depth analysis of big data (provided in various formats) and the country's existing security requirements [6]. The purpose of Namahoot and Laohavichien's research on Internet finance is to examine the relationship between the five dimensions of service quality and Thailand's overall behavioral intention to use online banking and use perceived risk and trust as an intermediary to explain the indirect influence between service quality and the behavior intention of using online banking [7]. In the research of these researchers, most of their research results have not been well used in social practice. The innovation of this article lies in the new perspective of topic selection. Since the rise of Internet finance has not been long, scholars' research on Internet finance is still in the exploratory development stage, and many fields have not yet been involved. In this article, the research focuses on Internet finance and its risks, and a series of related experiments and analyses are also carried out.

2. Machine Learning Algorithm and Internet Finance

2.1. Machine Learning. The main goal of machine learning algorithm applications is to help companies make more informative business decisions about [8]. It may include web server logs, Internet clickstream data, social media content and activity reports, text from customer emails, mobile phone call details, and machine data captured by multiple sensors [9]. At present, machine learning algorithms are widely used in many aspects. As shown in Figure 1, the fields where machine learning computing is widely used are shown.

2.2. Data Mining. Information mining ordinarily alludes to the method involved with looking for buried data in a lot of information through calculations. Information mining is typically connected with software engineering, and the above objectives are accomplished through different techniques like insights, online examination and handling, and data recovery [10]. Figure 2 shows a schematic diagram of the data mining process.

A basic and complete information mining engineering is the most common way of consolidating information stockrooms with other data frameworks [11]. The capacity of the information stockroom is to channel, arrange, clean and basically dissect the information from other data frameworks or information assortment devices, and gather the handled information into the capacity unit [12]. Data warehouse and database are two completely different

concepts [13]. The former is a sub-data set extracted from the latter and collected after preliminary screening and processing. Figure 3 shows the data mining structure diagram.

Data mining technology has been applied to the following areas.

2.2.1. Retail Industry. Department stores and large supermarkets such as Walmart are all conducting data mining on the historical data of daily customer purchases. In order to guess the recent consumption habits of customers, predict the general sales of goods in the future or the correlation between different kinds of goods [14]. Data mining technology does not predict customer behavior through the analysis of historical data, but only analyzes a certain information mode implicit in the historical data [15].

2.2.2. E-Commerce Field. Now Taobao customers only need to log in to their account, and the home page of the page will appear on the home page of the page that they have previously entered in the search bar to search for a certain type of product, or the product that the customer has bought [16]. Alibaba's data mining analysis system will lock customers' preferences through data mining technology in many products, recommend products that customers may prefer, and greatly increase transaction opportunities.

2.2.3. Fraud Detection Area. Telecommunications, insurance, and bank credit card departments often face customer fraud, such as vicious credit card overdrafts and misreporting insurance incidents, which have caused a lot of economic losses to the insurance and banking industries. If you can distinguish between normal customers and abnormal customers, predict potential fraud and its customer groups in advance. Even if few predictions can be obtained, they can reduce losses in the telecommunications industry, insurance industry, and banks [17].

2.2.4. Risk Control. There are many factors in the financial industry that will have varying degrees of impact on the efficiency of loan repayment and the calculation of customer credit ratings. In order to further prevent the occurrence of financial risk rates, financial institutions often use data mining methods to conduct some surveys on customers. Feature selection and attribute correlation calculation help to identify important and nonrelevant factors.

2.3. Apriori Algorithm. In order to better find the relevance in the transaction data set, the researchers put forward the concept of association rule mining technology. Along with the concept of association rules, it has attracted the attention of many researchers at home and abroad. They have conducted a lot of analysis in this field, many of the data mining algorithms derived from it have not been applied before, and the most eye-catching is the Apriori algorithm [18]. But the Apriori algorithm is not mature enough and has many shortcomings. In order to solve its shortcomings, some good



FIGURE 1: Application of machine learning algorithms.

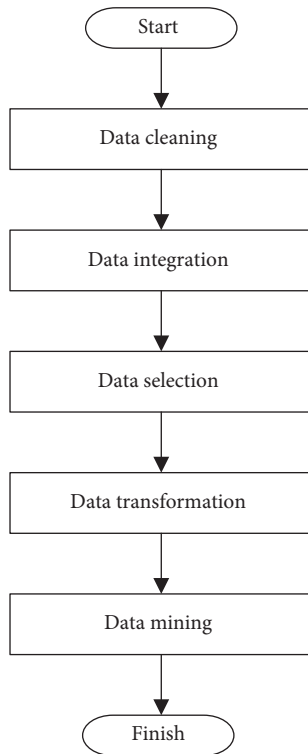


FIGURE 2: Data mining flowchart.

solutions have been proposed. On the basis of previous research algorithms, it provides algorithmic support for association rules, so that it can display its characteristics in practical applications. With the development of the times, association rule analysis technology has also been used in various industries, such as the design and layout of shelves [19]. In the modern electronic information industry, in industries where there are a large number of customers, such as banks and commerce, association rules are the most common application scenarios. It can analyze the data

resources existing in these industries [20]. At the same time, association rules have also made major breakthroughs in economics, medical and health, etc., fully demonstrating its broad prospects and advantages.

The basic idea of the Apriori algorithm: mining transaction data sets, first scan the transaction data set, and count each item set. Then, it is compared with the minimum support count set by the user. If the number of itemsets is larger than the minimum support count, then the itemsets are regarded as a frequent itemset. If it is smaller, delete it; the second is to self-join the frequent itemsets obtained from the first scan to obtain candidate item sets. Then, scan the transaction data set again, and count the candidate item sets obtained by the connection. If it is larger than the minimum support count, then the item set is regarded as a frequent item set; if it is smaller, then it is deleted; repeat this step continuously until there is no $k + 1$ candidate item set, and the final k th item set is the final frequent k item set.

2.4. Bayesian Network. Bayesian network is a graphical model used to express the relationship between variables [21]. Each node in the graphical model of the Bayesian network represents a corresponding variable. If there is a connecting arc between two nodes, it means that the two probability variables are probabilistically dependent; if there is no connecting arc between the two nodes, the two variables depend on each other. Parameter learning is an important part of Bayesian networks. Parameter learning assumes a known network structure and obtains specific probability dependence relationships between variables by analyzing data. The purpose of parameter learning is to obtain the probability density function $G(\alpha|X, \delta)$ of each node. The first is the maximum likelihood estimation method: under a given α , the conditional probability density $G(X|\alpha)$ of the data X is called the likelihood of α . If X is fixed and α changes in its domain, then let

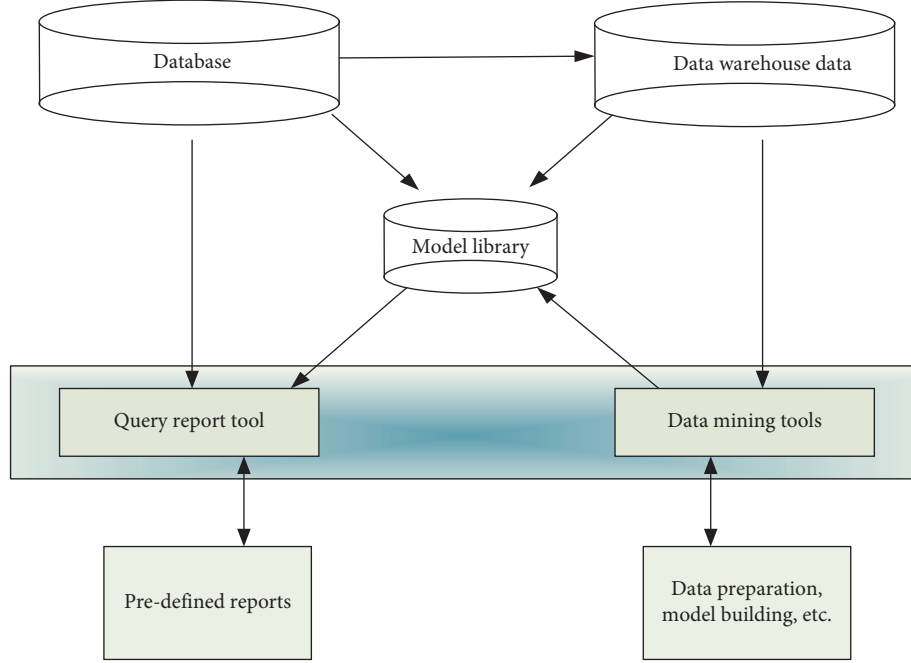


FIGURE 3: Data mining structure diagram.

$$S(\alpha|X) = G(X|\alpha). \quad (1)$$

$S(\alpha|X)$ is called the likelihood function of α . Therefore, the maximum likelihood estimation of the parameter α is the α value corresponding to the maximum value of the function, namely,

$$\alpha^* = \underset{\alpha}{\operatorname{argsup}} S(\alpha|X). \quad (2)$$

Suppose a Bayesian network composed of x variables $N = \{N_1, N_2, \dots, N_n\}$, and node N_i has t_i values. Then, there are z_i value combinations of its parent node $\phi(N_i)$. If node N_i has no parent node, then $z_i = 1$, so the parameters of the Bayesian network are

$$\alpha_{ivc} = G(N_i = L | \phi(N_i) = v). \quad (3)$$

The value of i is $1 \sim n$, and the values of v and c are $1 \sim z_i$ and $1 \sim t_i$ respectively. These parameters are not relatively independent, so for any i and v , the normality of the probability distribution can be obtained:

$$\sum_{L=1}^{t_i} \alpha_{ivc} = \sum_{L=1}^{t_i} G(N_i = L | \phi(N_i) = v) = 1. \quad (4)$$

For a set of complete data $N = \{N_1, N_2, \dots, N_n\}$ on the assumption of independent and identical distribution of the Bayesian network, the maximum likelihood function corresponding to α is

$$S(\alpha|X) = \log \prod_{l=1}^m G(X_l|\alpha) = \sum_{l=1}^m \log G(X_l|\alpha). \quad (5)$$

In:

$$\begin{aligned} \log G(X_l|\alpha) &= \sum_{i=1}^n \sum_{v=1}^{z_i} \sum_{c=1}^{t_i} \lambda(i, v, c; X_l) \log \alpha_{ivc}, \\ \lambda(i, v, c; X_l) &= \begin{cases} 1, & N_i = L \text{ and } \phi(N_i) = v, \\ 0, & N_i \neq L \text{ and } \phi(N_i) \neq v. \end{cases} \end{aligned} \quad (6)$$

If the sufficient statistic w_{ivc} is the number of samples satisfying $N_i = k$ and $\phi(N_i) = v$ in the data, that is,

$$w_{ivc} = \sum_{l=1}^m \lambda(i, v, c; X_l), \quad (7)$$

$$S(\alpha|X) = \sum_{l=1}^m \log G(X_l|\alpha) = \sum_{l=1}^m \sum_{i=1}^n \sum_{v=1}^{z_i} \sum_{c=1}^{t_i} w_{ivc} \log \alpha_{ivc}. \quad (8)$$

Then, the maximum likelihood estimate corresponding to α is

$$\alpha^* = \begin{cases} \frac{w_{ivc}}{\sum_{L=1}^{z_i} w_{ivc}}, & \text{if } \sum_{L=1}^n w_{ivc} > 0, \\ \frac{1}{z_i}, & \text{if } \sum_{L=1}^n w_{ivc} = 0. \end{cases} \quad (9)$$

The second is the bass estimation method. In the Bayes estimation process, the parameter α is considered as a probability change. First, the probability distribution $G(\alpha)$ represents the predistribution of the previously known information α and then uses the Bayesian expression

likelihood function to combine to synthesize the sample information, namely,

$$G(\alpha|X) \propto G(\alpha)S(\alpha|X). \quad (10)$$

According to the calculation of formula (8), the likelihood function of α , and the relationship between the likelihood function and statistic w_{ivc} can be obtained as shown below.

$$G(\alpha|X) \propto G(\alpha) \prod_{i=1}^n \prod_{v=1}^{z_i} \prod_{c=1}^{t_i} \alpha_{ivc}^{w_{ivc}}. \quad (11)$$

2.5. Least Squares Support Vector Regression Algorithm.

The optimization index of the least squares support vector is the sum of squares, and the loss function adopted is the quadratic loss function $v(m, n, f(m, x)) = |n - f(m, x)|_e^2$. Given a set of data,

$$\% E\{(m_1, n_1), \dots, (m_i, n_i)\} \in (M * N)^i, M = P^t, Y = P. \quad (12)$$

The optimization problem of this set of data is

$$\min \frac{1}{2} \|z\|^2 + \frac{q}{2} \sum_{i=1}^u \ell_i^2. \quad (13)$$

The constraints of this set of data are

$$n_i = z^t m_i + y + \ell_i, \quad i = 1, 2, \dots, u. \quad (14)$$

Construct the Lagrange function according to the given data:

$$L = \frac{1}{2} \|z\|^2 + \frac{q}{2} \sum_{i=1}^u \ell_i^2 - \sum_{i=1}^u \partial_i (z^t m_i + y + \ell_i - n_i). \quad (15)$$

The extreme value of function L satisfies the following conditions:

$$\frac{\partial}{\partial z} L = 0 \quad \frac{\partial}{\partial y} L = 0 \quad \frac{\partial}{\partial \ell_i} L = 0 \quad \frac{\partial}{\partial \partial_i} L = 0. \quad (16)$$

According to the formula, the following equation can be obtained:

$$\begin{aligned} z &= \sum_{i=1}^u \partial_i m_i, \quad i = 1, 2, \dots, u, \\ \sum_{i=1}^u \partial_i &= 0, \quad i = 1, 2, \dots, u, \\ \partial_i &= d\ell_i, \quad i = 1, 2, \dots, u, \\ z^t m_i + y + \ell_i - n_i &= 0, \quad i = 1, 2, \dots, u. \end{aligned} \quad (17)$$

The optimized function can be written as the following:

$$\begin{bmatrix} I & 0 & 0 & -m \\ 0 & 0 & 0 & -\vec{1}^t \\ 0 & 0 & dI & -I \\ m^t & \vec{1} & I & 0 \end{bmatrix} \begin{bmatrix} z \\ y \\ \ell \\ \partial \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ n \end{bmatrix}. \quad (18)$$

Among them, $m = [m_1, \dots, m_u], n = [n_1, \dots, n_u]$; according to these formulas, we can get

$$\begin{bmatrix} 0 & \vec{1}^t \\ \vec{1} & m^t m + d^{-1} I \end{bmatrix} \begin{bmatrix} y \\ \partial \end{bmatrix} = \begin{bmatrix} 0 \\ n \end{bmatrix}. \quad (19)$$

Among them, $z = \sum_i \partial_i m_i$.

Among all the parameters of the least squares support vector machine, only the parameter d is unknown. It is less than the standard support vector regression machine, so the stability of the least squares support vector machine is better, and the solution speed is faster than others [22].

2.6. Internet Finance. Web finance (ITIN) alludes to another monetary plan of action, wherein monetary establishments and Internet organizations use Internet innovation and data and correspondence advancements to acknowledge reserve installment, speculation, and different administrations [23].

2.7. The Development Model of Internet Finance. With the rapid development of Internet finance, Internet finance includes several different development models such as P2P online loan models, third-party settlements, digital currencies, and big data financing.

2.7.1. Crowdfunding. Crowdfunding means to collect funds from the people collectively. Crowdfunding financing generally involves three parties, namely, the initiator, the supporter, and the platform. The working method of crowdfunding supporting is that the initiator conveys an innovative venture that needs assets to the crowdfunding stage, and the crowdfunding stage checks the undertaking. After the task is endorsed, a site page connected with the undertaking can be made on the crowdfunding stage to draw in allies to contribute. The essence of crowdfunding financing is that the initiator uses the project as a unit to raise funds from netizens.

2.7.2. Third-Party Payment. Third-party payment can be divided into a narrow sense and a broad sense. The narrow concept is that nonbanking institutions create an electronic payment model that links users and the bank's payment settlement system by signing an order with the bank. The broad concept is that nonfinancial institutions act as intermediaries of income and expenditure and engage in other aspects of payment services specified by the People's Bank of China.

2.7.3. Big Data Finance. Huge information finance alludes to the utilization of registering apparatuses, for example, distributed computing by monetary help stages to work out monstrous measures of unstructured information got through network innovation, and to perform comprehensive information analysis on customer data. It is a model to discover customers' transaction consumption habits and accurately predict customer consumption trends to reduce

risks. This model is mainly implemented based on computing tools such as cloud computing and big data, and its requirement is that the amount of unstructured data for computing analysis must be large enough.

2.7.4. Information-Based Financial Institutions. Data-based monetary establishments allude to the utilization of data through specialized means to change and rearrange existing business processes, and to accomplish thorough informatization of the whole monetary foundation start to finish. The comprehensive informatization of financial institutions is an inevitable development trend in the current economic and technological environment, and it is also a necessary development direction to establish its own strategic advantages.

2.8. Basic Theory of Internet Finance. Internet finance has its own basic theories. The basic theories of Internet finance mainly include information asymmetry theory, long tail theory, and transaction cost theory.

2.8.1. Information Asymmetry Theory. The essence of information asymmetry theory is that there is a difference in the information that the two parties who trade in the market have and understand. People who have a large amount of trading market information can better use the information to serve themselves, and the less trading market information they know, the more passive they are in trading. Even traders who know the information can get a certain amount of income by passing information to other traders who do not know the information. In the Internet financial model, the gap in the amount of information held between the two parties in a transaction is getting smaller and smaller, but the party who borrows the funds still has more information than the party who lends the funds.

2.8.2. Long Tail Theory. The long tail theory believes that as long as the storage cost of holding the goods and the cost of obtaining information are quite small, the number of buyers with special needs for the product in the market is very large. Then, products with small sales and products with large sales but limited varieties have the same market share [24]. The long tail theory overturns the traditional rule of focusing only on a small number of buyers who purchase a large number of products. For Internet finance, information asymmetry and transaction costs are significantly reduced. The long tail of Internet finance is more biased toward small and medium-sized enterprises and low-income groups, showing the characteristics of inclusive finance.

2.8.3. Transaction Cost Theory. Transaction cost is the cost of both parties to the transaction in order to achieve a certain purpose. In Internet finance, because Internet finance can use network information technology to replace operating costs and service costs in traditional finance and shorten the enterprise value chain, transaction costs will be greatly

reduced [25]. The decline in transaction costs will promote more and more active financial market activities.

2.9. Dynamic Audit and Early Warning of Internet Financial Risks. Risk early warning refers to the establishment of early warning indicator models and systems by the main body of risk prevention by collecting and analyzing a large amount of data and information and relying on advanced information technology [26]. It monitors the changes in all aspects, hoping that when the indicators change, certain means and methods can be adopted in advance to manage, so that the company can operate as expected. Risk early warning is advanced; that is, when risks have not yet occurred, risks can be analyzed and evaluated through risk early warning models, and some scientific and effective methods can be used to transfer, reduce, or eliminate risks. Table 1 shows the judgment matrix of Internet financial risk evaluation.

In the business activities of enterprises, the method of mixed operation is often used. Table 2 shows the risk judgment matrix of mixed operation.

2.9.1. Internet Financial Risk Dynamic Audit System. The overall structure of the Internet financial risk dynamic audit system is divided into two parts: the network topology of the network audit information system and the functional modules of the network audit system. Figure 4 shows a topology diagram of the network structure of the networked audit information system:

2.9.2. Monitoring and Early Warning. Monitoring and early warning is a continuous real-time dynamic financial risk analysis technology, that is, real-time collection and mining of data, real-time calculation of key indicators, and comparison and analysis with budget values. Once the difference between the two exceeds the set threshold, an early warning will be made in time to remind the manager to find and deal with the problem. Financial crisis early warning is the most important application of data mining financial analysis platform in financial risk forecasting and analysis. Figure 5 shows the structure diagram of a common Internet financial risk early warning analysis system.

Although monitoring and early warning technology can well find some problems that may exist in financial enterprises. However, due to the high production cost of this technology and the high requirements for personnel quality, only a few companies will use this technology.

2.9.3. Audit and Early Warning Client. The audit and early warning client software is an important part of the entire early warning system. It directly faces the audit staff and fully realizes the audit and analysis functions of the fund's operation status. The software can analyze the possible situation of financial funds in the future and discover the key points of the audit according to the preset early warning monitoring indicators. Its specific functions are as follows:

TABLE 1: Internet financial risk judgment matrix.

F	F1	F2	F3	F4
F1	1	1/3	3	6
F2	3	1	3	4
F3	1/3	1/3	1	2
F4	1/6	1/4	1/2	1

TABLE 2: Risk judgment matrix of mixed operation.

F2	F21	F22	F23
F21	1	1/5	1/2
F22	4	1	2
F23	2	1/2	1

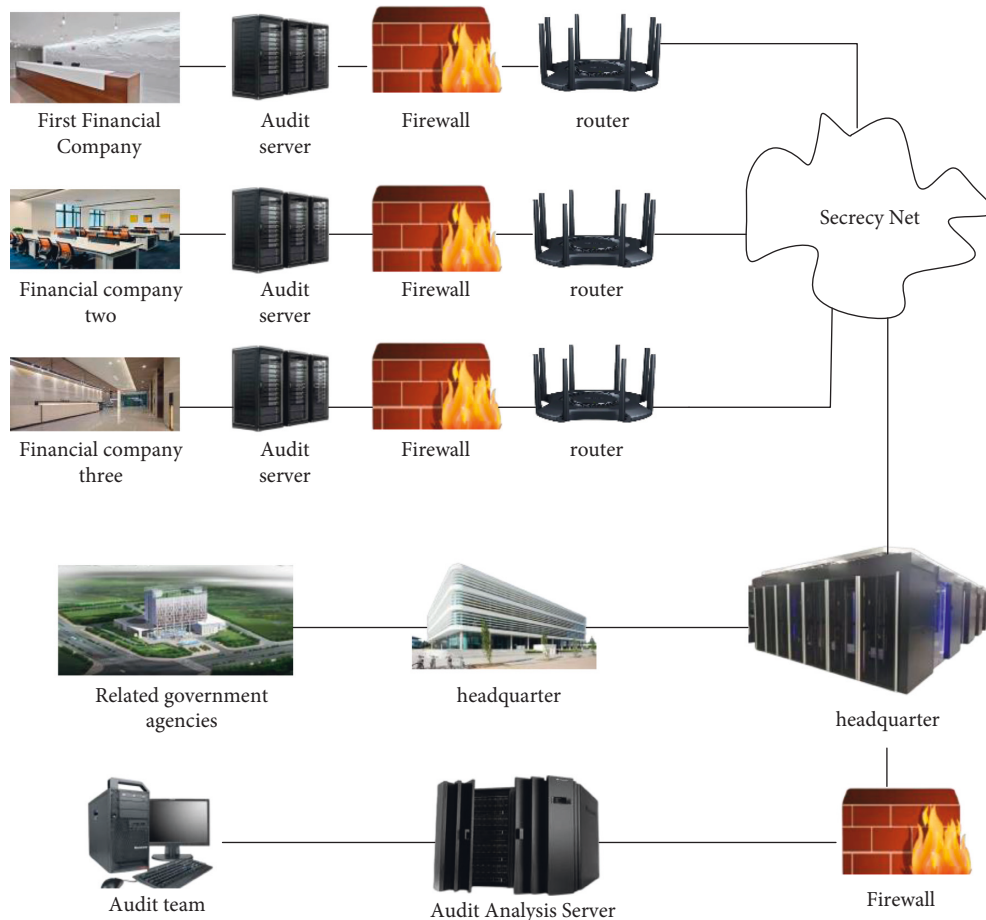


FIGURE 4: Network topology structure diagram of the networked audit information system.

- (1) Query, filter, and graphical analysis of incremental data. Through the analysis of incremental data, auditors can discover the status of data changes.
- (2) Perform audit early warning function on incremental data, and be able to browse and process audit early warning results.
- (3) Multi-angle and multi-level analysis of audit early warning results. Through further analysis, we can discover the focus of the next audit.
- (4) Provide visual analysis function of data and some commonly used auxiliary functions, thus facilitating the analysis of data by auditors.

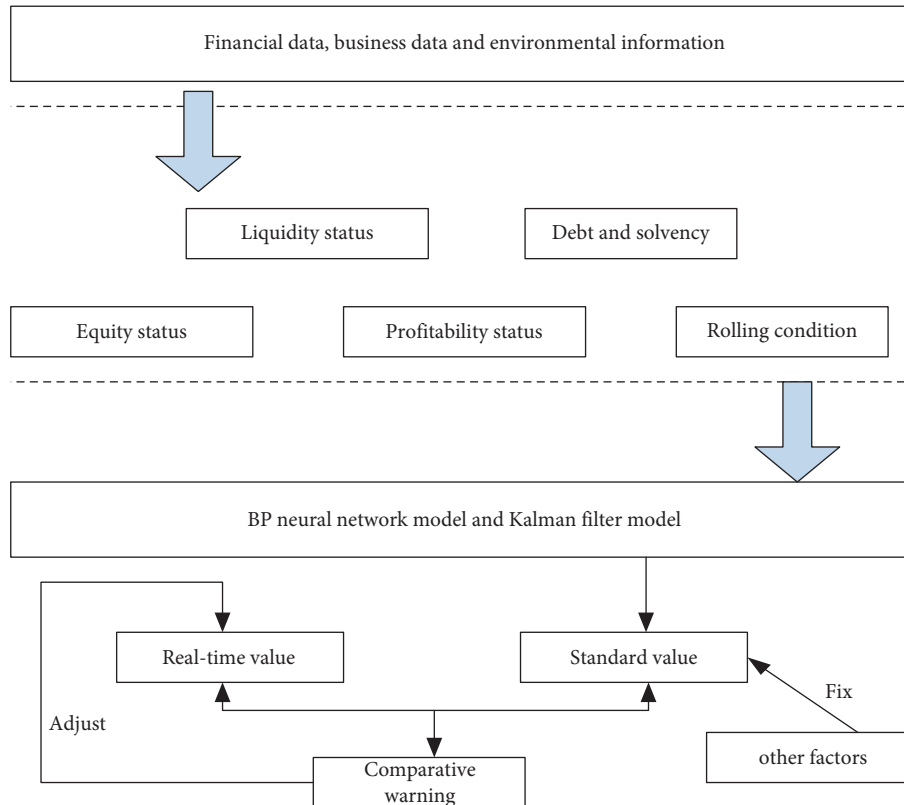


FIGURE 5: Internet financial risk early warning analysis system.

Figure 6 shows the operating process of the audit early warning system.

3. Dynamic Audit Experiment of Internet Finance Risk Based on Machine Learning Algorithm

3.1. The Transaction Scale of the Internet Third-Party Payment Market. The main body of third-party payment is usually a third-party independent institution with strong strength and guaranteed reputation. The business development method is generally through the cooperation with the bank and facilitates the transaction between the two parties to the transaction to realize the network payment mode of the transfer of funds. It mainly completes the settlement of funds through the interface docking with the bank's payment and settlement system. Table 3 shows the changes in the transaction scale of the domestic third-party payment market in recent years.

3.2. Survey on the Status Quo of Internet Financial Transactions. In the course of this investigation and experiment, we will give full play to the role of big data mining and big data analysis. It has conducted surveys and data collation on the order volume data of some financial companies in recent years. Table 4 shows the order volume of Internet finance companies obtained during this survey.

3.3. Internet Financial Market Risk Survey. Due to the large number of Internet financial service platforms, there is greater competition among various platforms, and some platforms attract customers through higher annualized yields, resulting in greater market risks for the platforms; interest rates change with changes in the market. The comprehensive interest rate level in the Internet financial service platform refers to the total level of interest rates that belong to the platform in the current month, which can reflect market risks. The higher the interest rate level, the more interest that needs to be repaid, and the greater the risk. The current status of the comprehensive interest rate level of the Internet financial service platform is shown in Table 5.

4. The Results of the Dynamic Audit Experiment on Internet Financial Risks

4.1. The Survey Results of the Transaction Scale of the Internet Third-Party Payment Market. According to the survey of the transaction scale of the domestic third-party payment market, a graph of the changes in the domestic third-party payment market transaction scale in the past few years can be drawn, as shown in Figure 7.

With the continuous development of Internet technology and continuous improvement of network infrastructure, domestic third-party payment services have developed rapidly in recent years. It can be seen from Figure 7 that a

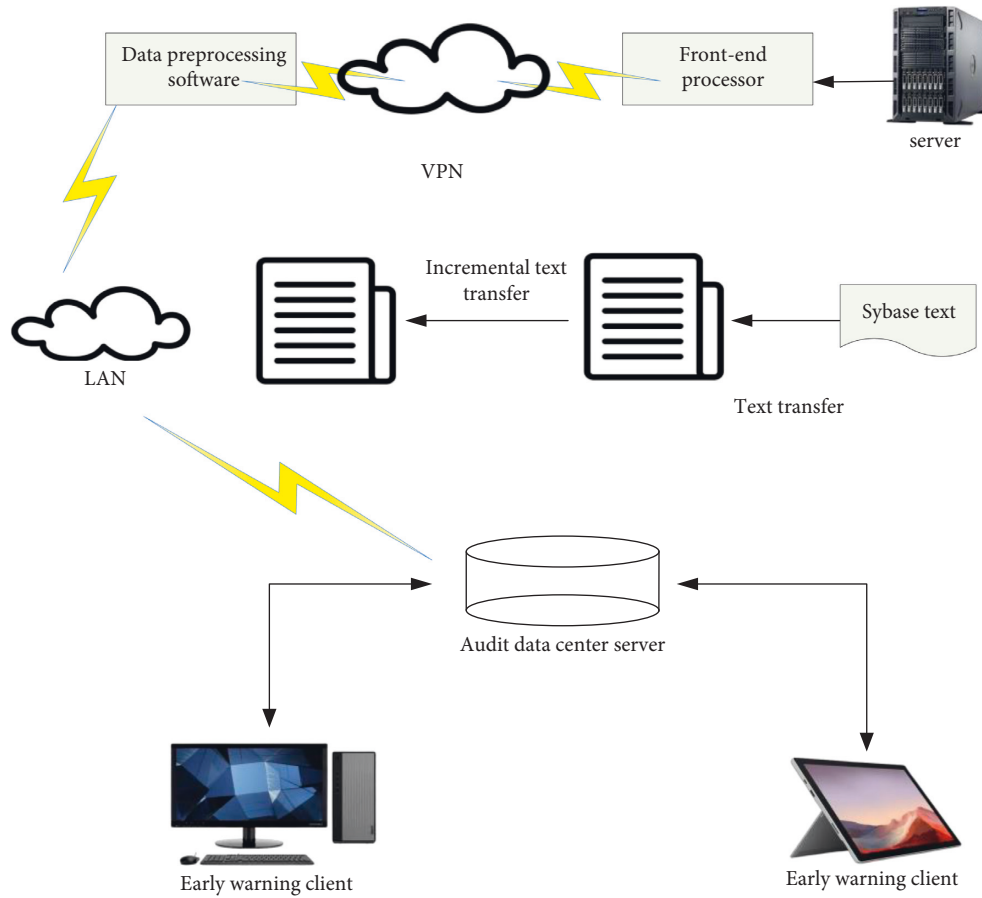


FIGURE 6: Operation process of early warning system.

TABLE 3: Transaction scale of the third-party payment market.

Years	Transaction size (trillion yuan)	Growth rate (%)	Market share (%)
2012	12.4	47.6	12.3
2013	17.3	38.7	16.9
2014	23.6	35.9	21.8
2015	31.0	33.4	26.9
2016	57.1	87.5	38.5

TABLE 4: Order volume situation.

	Private	Listed company	State-owned	VCs	Bank
2012	1256	12.3	3.8	65.9	32.1
2013	1354	48.6	13.6	198	38.6
2014	1652	98.1	21.5	256	98.5
2015	3647	221	29.8	468	162.8
2016	3897	362	37.6	654	201.6
2017	4561	554	50.1	960	268.9

TABLE 5: Current situation of comprehensive rate of return.

	Private	Listed company	State-owned	VCs	Bank
2012	11.6	12.6	8.6	11.1	6.4
2013	14.8	13.1	9.1	11.6	6.9
2014	19.6	14.2	9.6	12.0	7.5
2015	15.2	13.3	8.5	11.3	8.2
2016	13.6	11.2	9.9	12.4	7.8
2017	12.1	10.9	10.2	10.6	6.9

total of 12.4 trillion yuan of transactions were completed in 2012 and 57.1 trillion yuan in 2016. Compared with 2012, it has increased nearly four times.

4.2. Analysis of the Current Survey Results of Internet Financial Transactions Based on Machine Learning Algorithm. In this research, a statistical analysis of data is carried out on the

current transaction status of Internet finance. According to the data in Table 4, a statistical graph of the transaction volume of the domestic Internet financial market in recent years can be obtained. The details are shown in Figure 8.

It can be seen from Figure 8 that the transaction volume of private financial service platforms has grown from 125.6 billion yuan in 2012 to 456.1 billion yuan in 2017. The huge transaction volume truly reflects the characteristics of

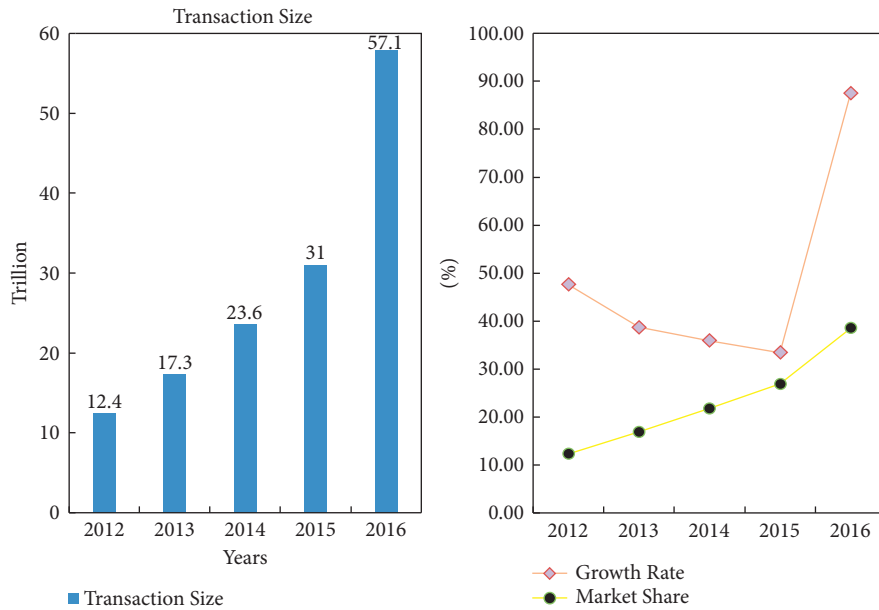


FIGURE 7: Changes in the transaction scale of the third-party payment market.

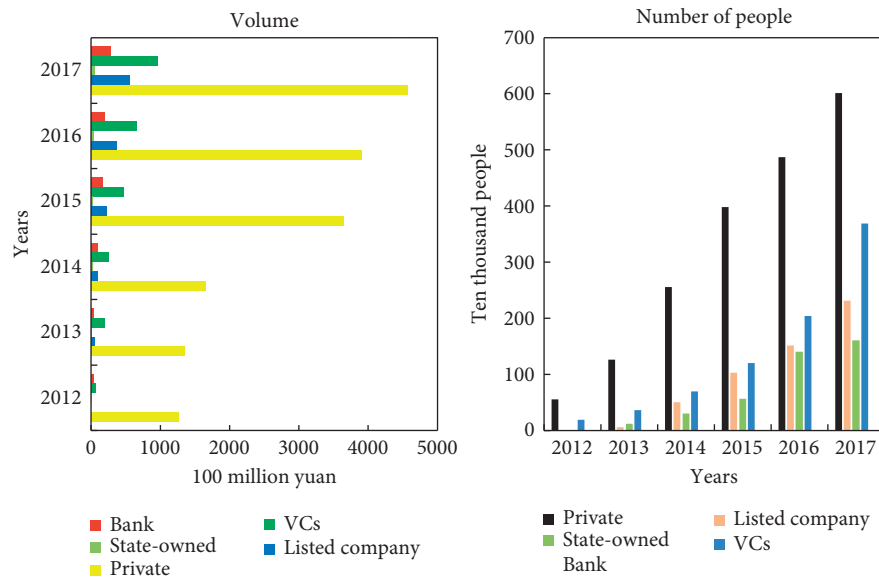


FIGURE 8: Internet financial market transaction volume.

inclusive finance in Internet finance. However, combined with the two comprehensive data of the number of Internet financial platforms and transaction volume, it can be judged that the transaction volume of a single private financial service platform is lower than that of other departments.

4.3. *The Results of the Internet Financial Market Risk Survey.* Market risk refers to the risk of a sharp decline in the share of the Internet financial service platform due to market changes during the production and operation of the Internet financial service platform. In the investigation of the risks of

the Internet financial market, this experiment sorted out the survey data. According to the data in Table 5 and some other experimental data, a graph of the changes in the market rate of return of Internet finance in recent years can be drawn. The details are shown in Figure 9.

It can be seen from Figure 9 that from 2012 to 2017, the comprehensive interest rate level of various platforms has shown a downward trend. However, the comprehensive interest rate level of private Internet finance companies is still significantly higher than that of other departments, reaching a profit rate level of more than 20% in 2014. However, the greater the value of the comprehensive interest rate level, the greater the market risk.

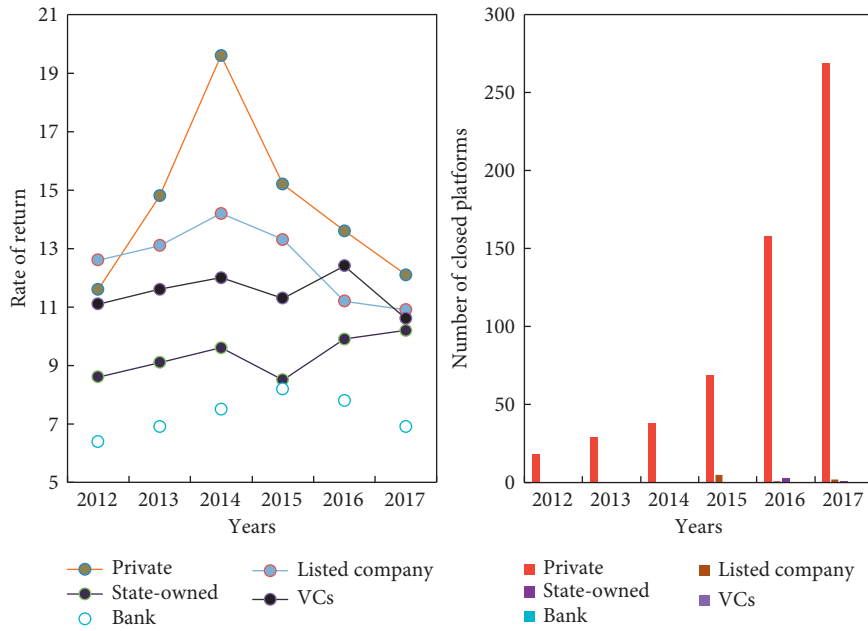


FIGURE 9: Changes in Internet financial yield.

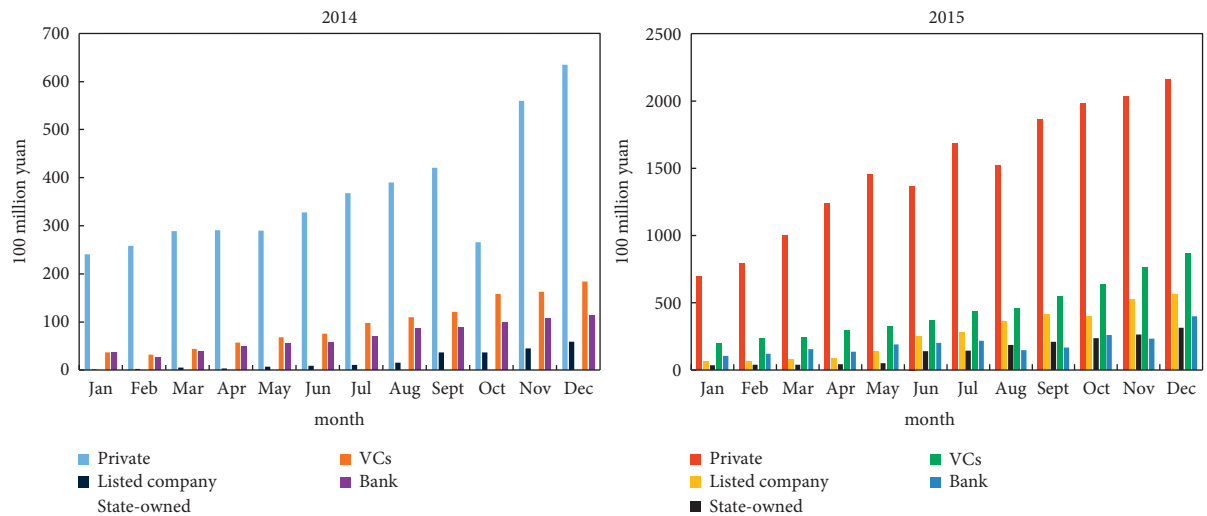


FIGURE 10: Historical outstanding balances of various platforms.

Figure 10 shows the current status of historical outstanding balances of various platforms in 2014 and 2015.

It can be seen from Figure 10 that the pending balances of various platforms are increasing, especially private enterprises have the largest pending balances. In 2015, the average monthly outstanding balance exceeded 100 billion yuan. The higher the balance, the greater the repayment pressure of the enterprise and the greater the credit risk. If the loan cannot be recovered on time and the platform’s capital chain is broken, the company may go bankrupt.

5. Conclusions

In recent years, under the rapid development of mobile communication technology and machine learning

algorithms, Internet finance has achieved breakthrough development. Because Internet finance has the dual attributes of Internet and finance, Internet finance not only faces all the risks of traditional finance, but also faces some major risks. This article analyzes the current situation and existing problems of Internet finance, investigates some market transactions and profit margins of Internet finance, and fully carried out research on the dynamic audit and early warning of Internet financial risks. The research results of this paper show that it is necessary to strengthen the analysis of corporate financial information data and construct a dynamic audit and early warning system for Internet financial risks as soon as possible. This can be the first to find abnormalities in corporate funds, prevent the abnormal operation of the corporate capital chain, and further prevent the occurrence of financial crises.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by Scientific Research Program Funded by Shaanxi Provincial Education Department (Program no. 13JK0160).

References

- [1] L. A. Tawalbeh, R. Mehmood, E. Benkhelifa, and H. Song, "Mobile cloud computing model and big data analysis for healthcare applications," *IEEE Access*, vol. 4, no. 99, pp. 6171–6180, 2016.
- [2] R. Massobrio, S. Nesmachnow, A. Tchernykh, A. Avetisyan, and G. Radchenko, "Towards a cloud computing paradigm for big data analysis in smart cities," *Programming and Computer Software*, vol. 44, no. 3, pp. 181–189, 2018.
- [3] L. Xu, C. Jiang, and J. Wang, "Information security in big data: privacy and data mining," *IEEE Access*, vol. 2, no. 2, pp. 1149–1176, 2014.
- [4] Y. Zhao, "Research on wireless distributed financial risk data stream mining based on dual privacy protection," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, pp. 241–253, 2020.
- [5] Ö Ege Oruç and C. Tatar, "An investigation of factors that affect internet banking usage based on structural equation modeling," *Computers in Human Behavior*, vol. 66, pp. 232–235, 2017.
- [6] S. Khatkhat, S. Jan, I. Ahmad, Z. Wadud, and F. Q. Khan, "An effective security assessment approach for Internet banking services via deep analysis of multimedia data," *Multimedia Systems*, vol. 27, no. 4, pp. 733–751, 2021.
- [7] K. S. Namahoot and T. Laohavichien, "Assessing the intentions to use internet banking: the role of perceived risk and trust as mediating factors," *International Journal of Bank Marketing*, vol. 36, no. 2, pp. 256–276, 2018.
- [8] Y. Huang, T. Li, C. Luo, H. Fujita, and S. Hornig, "Matrix-based dynamic updating rough fuzzy approximations for data mining," *Knowledge-Based Systems*, vol. 119, pp. 273–283, 2017.
- [9] S. Sengan, P. V. Sagar, O. I. Khalaf, and R. Dhanapal, "The optimization of reconfigured real-time datasets for improving classification performance of machine learning algorithms," *Mathematics in Engineering, Science and Aerospace (MESA)*, vol. 12, 1 page, 2021.
- [10] B. Tang, Z. Chen, G. Hefferman et al., "Incorporating intelligence in fog computing for big data analysis in smart cities," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2140–2150, 2017.
- [11] R. Ranjan, S. Garg, A. R. Khoskbar, E. Solaiman, P. James, and D. Georgakopoulos, "Orchestrating BigData analysis workflows," *IEEE Cloud Computing*, vol. 4, no. 3, pp. 20–28, 2017.
- [12] D. Santi, E. Magnani, M. Michelangeli et al., "Seasonal variation of semen parameters correlates with environmental temperature and air pollution: a big data analysis over 6 years," *Environmental Pollution*, vol. 235, pp. 806–813, 2018.
- [13] M. Ismail, N. E. El-Rashidy, and N. Moustafa, "Mobile cloud database security: problems and solutions," *Fusion: Practice and Applications*, vol. 7, no. No. 1, pp. 15–29, 2021.
- [14] H. Yan, "Cocustering of multidimensional big data: a useful tool for genomic, financial, and other data analysis," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 3, no. 2, pp. 23–30, 2017.
- [15] M. K. Hassan, A. K. Hassan, and A. I. Eldesouky, "MSJEP classifier: "modified strong jumping emerging patterns" for fast efficient mining and for handling attributes whose values are associated with taxonomies," *Journal of Intelligent Systems and Internet of Things*, no. 2, pp. 37–53, 2019.
- [16] P. P. Sfikakis, V. K. Bournia, P. Sidiropoulos et al., "Biologic treatment for rheumatic disease: real-world big data analysis from the Greek country-wide prescription database," *Clinical & Experimental Rheumatology*, vol. 35, no. 4, pp. 579–585, 2017.
- [17] S. Roh, "Big data analysis of public acceptance of nuclear power in Korea," *Nuclear Engineering and Technology*, vol. 49, no. 4, pp. 850–854, 2017.
- [18] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.
- [19] X. S. Yan and L. Zheng, "Fundamental analysis and the cross-section of stock returns: a data-mining approach," *Review of Financial Studies*, vol. 30, no. 4, pp. 1382–1423, 2017.
- [20] J. B. Varley, A. Miglio, V. A. Ha, M. J.G. M. Van Setten, and G. Hautier, "High-throughput design of non-oxide p-type transparent conducting materials: data mining, search strategy, and identification of boron phosphide," *Chemistry of Materials*, vol. 29, no. 6, pp. 2568–2573, 2017.
- [21] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116, 2017.
- [22] H. Hong, P. Tsangaratos, I. Ilia, J. Liu, A. X. Zhu, and W. Chen, "Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China," *Science of the Total Environment*, vol. 625, no. 1, pp. 575–588, 2018.
- [23] M. Rocchetti, G. Delnevo, L. Casini, and P. Salomoni, "A cautionary tale for machine learning design: why we still need human-assisted big data analysis," *Mobile Networks and Applications*, vol. 25, no. 3, pp. 1075–1083, 2020.
- [24] S. Xiao, S. H. Ma, G. Li, and S. K. Mukhopadhyay, "European option pricing with a fast fourier transform algorithm for big data analysis," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1219–1231, 2016.
- [25] B. Li and A. Arreola-Risa, "Financial risk, inventory decision and process improvement for a firm with random capacity," *European Journal of Operational Research*, vol. 260, no. 1, pp. 183–194, 2017.
- [26] X. Meng, "Analysis of marketing forms in the era of network economy," *International Journal of Educational Technology*, vol. 2, no. 1, pp. 43–45, 2021.