

## Research Article

# Innovation and Business Model of Life and Health Sharing Platform Based on Algorithm of Naive Bayesian Model

Lisheng Yu 

*Department of Medical and Health Industry Management, School of Management, Putian University, Putian 351100, Fujian, China*

Correspondence should be addressed to Lisheng Yu; [yulisheng555@ptu.edu.cn](mailto:yulisheng555@ptu.edu.cn)

Received 13 August 2022; Revised 23 September 2022; Accepted 29 September 2022; Published 14 October 2022

Academic Editor: M. A. Rashid Sarkar

Copyright © 2022 Lisheng Yu. 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.

Including shared bicycles, shared chefs, shared ideas, shared printing, and even shared human resources, the vocabulary of sharing has quietly entered people's lives, affecting everyone's necessities. "Everyone can participate" and "everything can be shared" are all familiar words in the sharing economy field. The implementation of the sharing economy is carried out by relying on the sharing platform. Knowledge, as a special resource, can exert its maximum benefit only when it is widely disseminated and shared, and the process of knowledge sharing and transformation itself contains knowledge innovation. By linking the multiparticipants of the platform, sharing, cocreation, and win-win are realized. The research value of this paper lies in the use of microservices to solve the problems of high coupling, poor scalability, and difficulty in rapid iteration in traditional monolithic applications, effectively identifying and blocking spam that may appear on the platform through a relatively simple solution. Meanwhile, it has been expected that under the trend of sharing economy, with the development of blockchain and 5G technology, an Internet life and health platform can be built. The upstream and downstream in the medical and health industry can be linked to share resources between the upstream and downstream in the industry, which can create a common nakedness, thus improving the operational efficiency and profitability of the entire medical and health industry. Finally, a spam identification scheme combining the improved Aho-Corasick algorithm (AC) and the naive Bayesian model (NBM) has been proposed, and a comparative experiment was conducted between this scheme and the scheme directly using NBM. The experimental results showed that the macro  $F1$  value of the improved scheme on the bad evaluation dataset of platform A was 3.6% higher than that of NBM alone, and the macro  $F1$  value of the improved scheme was 1.7% higher than that of NBM alone on the B platform review dataset. The overall performance of the improved NBM algorithm was stable and better than the traditional algorithm, which verified the feasibility of the scheme.

## 1. Introduction

The methods and techniques of data management and information management provide the foundation and tools for knowledge-sharing platforms. In order to develop a successful knowledge-sharing platform, these technologies should be effectively integrated and used. Because the realization of knowledge sharing depends on the management of data and information transmission, basic technologies such as database management system, network technology, and communication system are the necessary infrastructure for a knowledge-sharing platform. Each medical institution only stores medical information in its own server, network,

and information system. The information systems of different medical institutions are usually independent of each other, and the phenomenon of "information silo" is formed as a whole, making it difficult for users or institutions to exchange data, which also has become the biggest obstacle to the realization of medical information sharing among institutions. In order to solve the above problems, increase the compatibility and confidentiality of life and health information sharing, and reduce the potential cost of sharing platforms, this paper proposes a design and implementation scheme of a life and health information data-sharing platform based on NBM algorithm, improving the health data compatibility of the platform. At present, there are few

researches on platform business model innovation under the sharing economy of the medical and health industry. This paper delves into the practical application of NBM for text classification. Based on the improved multipattern matching algorithm and NBM algorithm, this paper proposes a spam identification scheme applied to the user's life and health platform, and conducts experiments to verify that it has a high accuracy rate.

Monolithic architecture is generally used in traditional web development, with the advantages of low learning cost and quick development. A single developer can complete the construction of a website. When the scale of the early project was small and the development team was small, it was a good choice to use the monolithic architecture for development. With the increase of business requirements, the expansion of project scale, and the expansion of development teams, projects developed using a monolithic architecture inevitably have many limitations, such as high coupling, poor scalability, and difficulties to iterate quickly. In addition, after research, it is found that most public platforms with publishing functions inevitably have some spam users, who submit unfriendly content such as abusing others, taunting others with insults and exaggeration, and illegal content related to pornography and gambling to the server. If the server does not identify and block these spam contents, it can have a great impact on the atmosphere of the platform and lead to the loss of users.

Based on the above background, this paper proposes a user complaint platform based on AC algorithm and NBM algorithm. The platform divides all functions into a set of microservices, using Dubbo + Zookeeper for microservice development and various technologies to solve problems encountered in the development process and improve platform performance. At the same time, the solution of spam identification is researched, and a spam identification scheme based on AC algorithm and NBM algorithm is proposed to improve the accuracy of spam identification. Finally, a user complaint platform with high performance, high availability, and good user experience can be realized.

The innovation of this paper lies in the following aspects. (1) This paper explores the methods and paths of platform business model innovation in the era of sharing economy. The business model innovation of life and health-sharing platform based on NBM algorithm is also studied, which provides certain ideas and guidance for the medical and health industry from traditional business model to platform business model innovation and at the same time provides certain ideas and references for enterprises in other industries planning to transform and innovate in the context of exploring the sharing economy. (2) This paper uses the canvas analysis method of the life and health-sharing platform business model based on the NBM algorithm to construct a factor analysis framework. Based on the theory of sharing economy and related theories of multilateral platforms, it goes deep into all aspects of enterprise operation, which can guide practice and achieve the true unity of knowledge and action.

## 2. Related Work

Sharing platforms have exploded in recent years, sparking entirely new ways of economic life that in some places surpass traditional models. For example, Beier M believed that sleep was a fundamental biological process critical to human survival and health, but many conditions, such as physical and mental disorders, environment, and lifestyle, can affect healthy sleep. To this end, he designed a virtual research platform that supports data sharing and processing among institutions [1]. Fiorella provided the context for plans for a new data-sharing platform, where various stakeholders are working to share data to make individual patient-level clinical trial data available to the scientific research community [2]. Zhang et al. described the core database resources in the communication of the BIG design studio, which provided a series of services for data precipitation, integration and sharing, laying a solid foundation [3]. Cui et al. discussed the secure management of the E-Clinic database system based on the user roles assigned by users when they registered in the E-Clinic portal. By being used in the E-Clinic portal, it can make the clinic's day-to-day operations more efficient and assist itself in data sharing with authorized personnel. His findings suggested that the E-Clinic portal can overcome future patient growth through an electronic database [4]. To sum up, the academic research on sharing platforms has not yet reached a consensus, and the research on sharing platforms is still in a period of theoretical discussion.

Among the naive Bayesian classification methods, the naive Bayesian model is also widely used and attracts more and more attention. For example, Chen et al. proposed a local NBM algorithm based on mutual information to better quantify and balance the contribution brought by common neighbors and the interaction between neighbor sets, thereby improving the accuracy of prediction [5]. Indraj and Annapurani can determine the target class of unknown drugs by applying NBM algorithm to the collected data [6]. Zhang et al. adopted NBM algorithm and EM (expectation maximization) algorithm to deal with missing data in software engineering. He also developed two strategies as missing data tolerance and missing data imputation to handle missing data in software datasets [7]. Based on the semisupervised naive Bayes classifier, Xu et al. proposed an improved algorithm based on data classification confidence. The results showed that the improved algorithm can effectively evaluate the air combat situation, which can also improve the classification performance without significantly affecting the efficiency of the classifier [8]. Behrouzian Nejad and Ebrahim Shiri proposed a new image classification technique. The use of heuristics increases the power of the estimator of the learning algorithm and continues to reach the desired state, as well as the classification performed automatically by the model produced by the above steps without user interference [9]. Fan et al. adopted the NBM algorithm of sentiment analysis method to extract sentiment indicators from the content of each online comment and

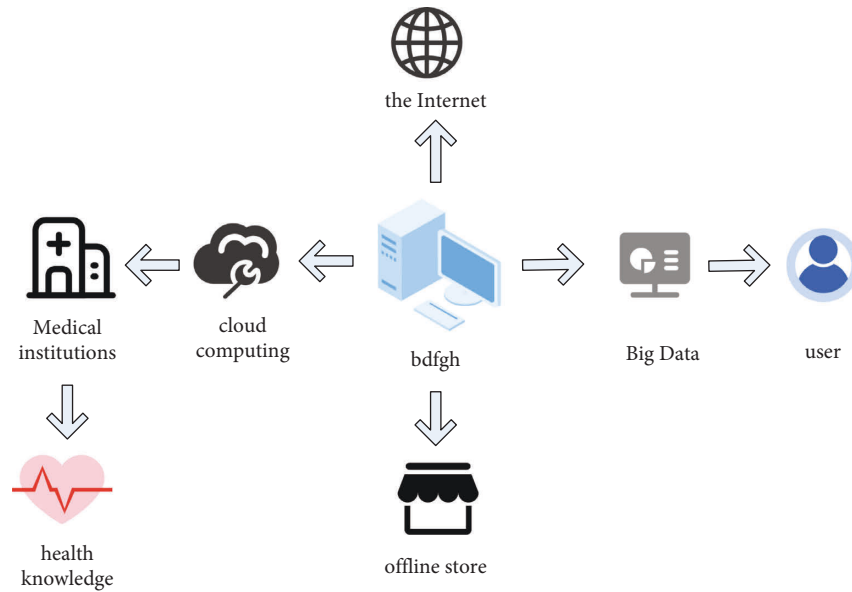


FIGURE 1: Shared platform business model innovation.

incorporate them into the imitation coefficients of the Bass/Norton model to improve the prediction accuracy [10]. To sum up, naive Bayesian method is still under continuous development and improvement in various fields, and there is still a lot of room for improvement and research.

### 3. Integration of Life and Health Data

*3.1. Innovative Ideas of Life and Health Platform Business Model.* Benefiting from the development of information technologies such as the Internet of Things, big data, and 5G in the context of the sharing economy, the medical and health industry has carried out platform business model innovation, linking users, medical institutions, businesses, and other participants, so that the innovation of the value proposition in the value sector and the innovation of important cooperation with doctors, medical institutions, and medical teams have been realized [11]. An innovative platform model with its own channels has been established to achieve channel innovation and obtain transaction commissions, advertising services, and technical service income, as shown in Figure 1.

As can be seen from Figure 1, the platform business model links users, merchants, medical institutions, and other participants, providing them with an interactive mechanism to meet the needs of all users for personalized customization and quality. The life and health platform relies on cloud computing, big data to realize a profitable platform business model by facilitating transactions between multiple parties [12].

*3.2. Web Services and System Architecture Design.* A Web service is essentially an abstract collection of semantically defined computational or physical activities to meet user or business needs. It is actually a software system that can be

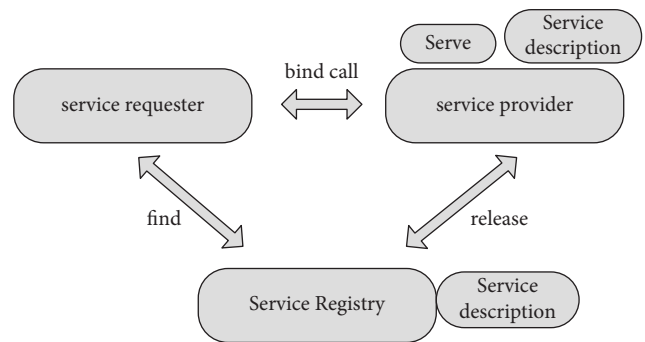


FIGURE 2: Web service structure.

identified by a URI. In the current network environment, there are various systems, and a technology is needed to communicate in these heterogeneous systems. Web service is such a versatile “translator” [13]. The reason why Web services can communicate between heterogeneous systems is that Web services use XML (eXtensible Markup Language) as the carrier of information transmission. XML provides a unified method to define and describe data, which is very suitable for transmission in the Internet [14]. The overall structure of the Web service is shown in Figure 2.

As can be seen from Figure 2, the structure is mainly composed of three important parts: service provider, service requester, and service registration center. Among them, the registration center has a crucial management role in the system. The user describes the required Web service through WSDL (Web Services Description Language). The prepared Web service is published in the registry and UDDI (Universal Description Discovery and Integration). After the release, users can search according to the given address to use the service. During use, the Web service can communicate with the client through SOAP (Simple Object Access Protocol) [15].

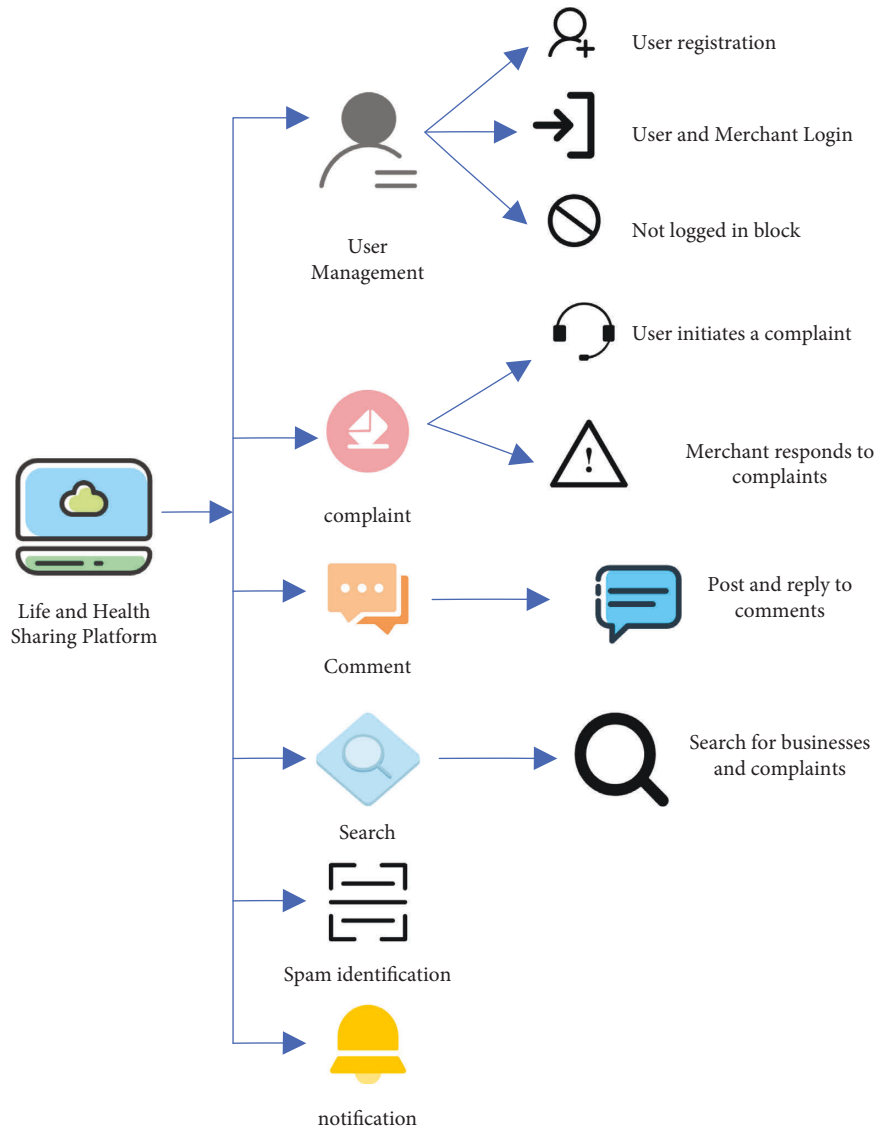


FIGURE 3: Platform function diagram.

The life and health-sharing platform is a Web application, and users can access the platform on the PC (personal computer) side [16]. The sharing platform can establish a more convenient and faster channel for people to protect their own consumer rights, which can help people whose consumer rights have been violated to complain to merchants, exposing the infringing acts of merchants [17]. The functions of this platform are shown in Figure 3.

As shown in Figure 3, the user management function is used to distinguish user roles, and different roles have different functional rights. Each user role exercises different permissions and uses different functions, which makes the platform clear and easy to manage [18]. The complaint function is the core function of this platform. In order to allow users to quickly find the complaints and businesses they are interested in and avoid spending a long time searching by themselves, the platform needs to provide a search function for complaints and businesses, where users can search by entering keywords. The platform is a sharing platform with a publishing function, and inevitably there are

some spam users. If the server does not identify and block these spam contents, it can have a greater impact on the atmosphere of the platform and lead to the loss of users. In response to this problem, this platform needs to identify the spam content published by users and merchants, and intercept the determined spam content.

Traditional development of Web generally uses a monolithic architecture. For example, a single WAR package is developed in Java and deployed to the application server Tomcat. All services such as development and compilation by developers in the same project run in the same process. Later, in order to ensure the high availability of the project, developers deploy the WAR package to multiple Tomcats, and distribute user requests through load balancing servers and related load balancing algorithms [19]. The architecture of this platform is shown in Figure 4.

However, as the overall scale of the project expands, projects implemented using a monolithic architecture inevitably have many limitations. In response to this situation, this article finally uses SpringBoot + Dubbo + Zookeeper to

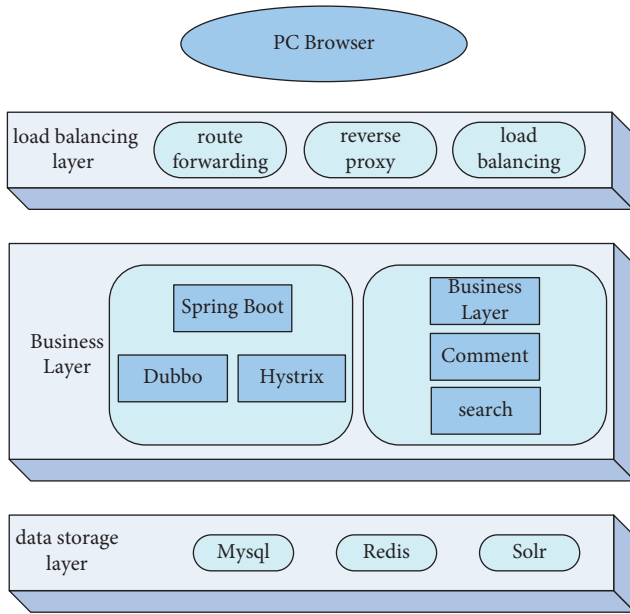


FIGURE 4: Platform architecture diagram.

develop microservices for the user complaint platform so that the coupling between functional modules can be reduced.

**3.3. Feasibility of Business Model.** The life and health information data-sharing platform based on naive Bayesian method is an important embodiment of medical and health informatization, and an important part of the development of health informatization. Many developed countries have paid more and more attention to medical and health informatization, and have also invested a lot of financial resources in it [20]. Compared with other countries, China started late, but with the continuous improvement of China’s medical and health level, it has gradually begun to realize the importance of information sharing of medical and health data. Therefore, the financial investment of China in this area has gradually increased, as shown in Figure 5.

As can be seen in Figure 5, China’s expenditure on health care has increased year by year. Total spending in 2021 was about 21% higher than in 2017, while spending on personal hygiene was up about 30.5% and community health was up about 26.3%. It can be seen that China’s emphasis on medical health is becoming more and more obvious. From the analysis, it can be seen that the business model research of the life and health-sharing platform in this paper is in line with the market demand of medical informatization, and the market potential is huge and has a clear market feasibility [21].

Taking the access data of platform A in the past 4 years as an example, through data analysis, the demand for sharing platforms in the emerging stage can be found as shown in Figure 6.

As can be seen from the data in Figure 6, the total number of visits to platform A and the proportion of total number of visitors and user access time fluctuated greatly,

and both the total number of visitors and the total number of visitors from 2019 to 2020 decreased. In other years, it was relatively an upward trend. However, the user access time is increasing year by year. In summary, the business model of the sharing platform has great potential user needs, with a certain market necessity.

**3.4. AC Algorithm.** The related technologies that need to be used in the development of the life and health-sharing platform in this paper need to be borrowed from the classic AC algorithm, and the implementation process of the algorithm provides the theoretical basis and technical support for subsequent research [22]. AC algorithm is a classic multipattern matching algorithm. Implementing the AC algorithm requires two steps: constructing a Trie tree according to the pattern string set, and adding a failure pointer to the constructed Trie tree. Trie tree is often used to deal with string matching problems. Building a Trie from a collection of pattern strings is straightforward. For example, if a pattern string set is {else, hello, her, see, she, so}, the constructed Trie tree is shown in Figure 7.

As shown in Figure 7, in the Trie tree, except for the root node of the Trie tree, every other node represents a character, and the path from the root node to each blue node represents a pattern string [23]. In the AC algorithm, each node of the Trie tree has a failure pointer, as shown in Figure 8.

It can be seen from the Trie tree in Figure 8 that the failure pointer of the root node is equal to null and the failure pointer of the first-level node points to the root node. According to these two initial conditions and the above method, failure pointers can be added to all nodes in the Trie tree.

The AC algorithm introduced here has high efficiency when applied to an English environment. However, when it is directly applied to the Chinese environment or the Chinese-English mixed environment, the space occupied by a node is at least 256 times that of the English environment, and there is a problem of space expansion. Therefore, the AC algorithm cannot be directly applied to a Chinese environment or a Chinese-English mixed environment. Some scholars have done some research on string matching in Chinese environment and Chinese-English mixed environment. The literature [24] proposed the THT (Top-Hat Transform) algorithm. It sets all nodes in the Trie tree as a complete hash table of size 256. English characters are represented by one node, and Chinese characters are represented by two adjacent nodes after encoding. In the matching process, there is no need to perform additional operations on the high and low bytes of Chinese characters, so it has high search efficiency. Because the high and low bytes of Chinese characters are constructed separately, there is no problem of space expansion.

**3.5. Algorithm of Naive Bayesian Model**

**3.5.1. Naive Bayesian Model.** With the theoretical support of Bayes-related mathematical knowledge, the NBM algorithm has extremely high classification efficiency and accuracy

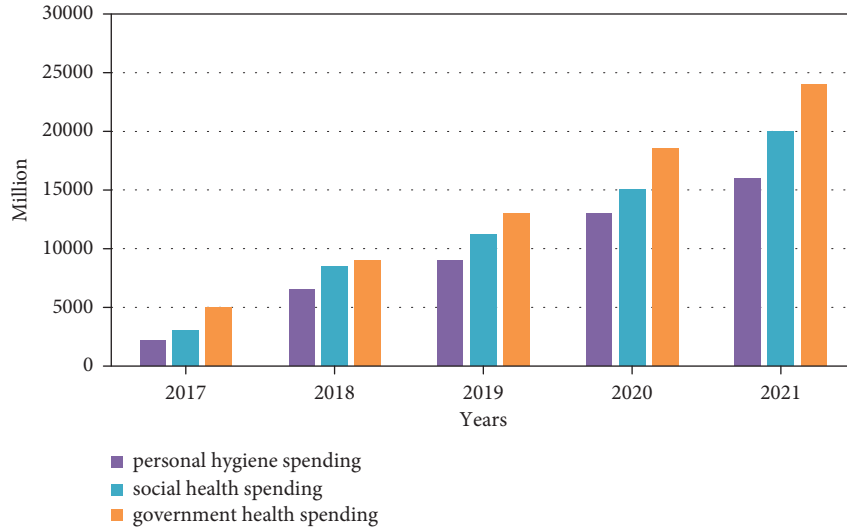


FIGURE 5: China's health care spending in the past five years.

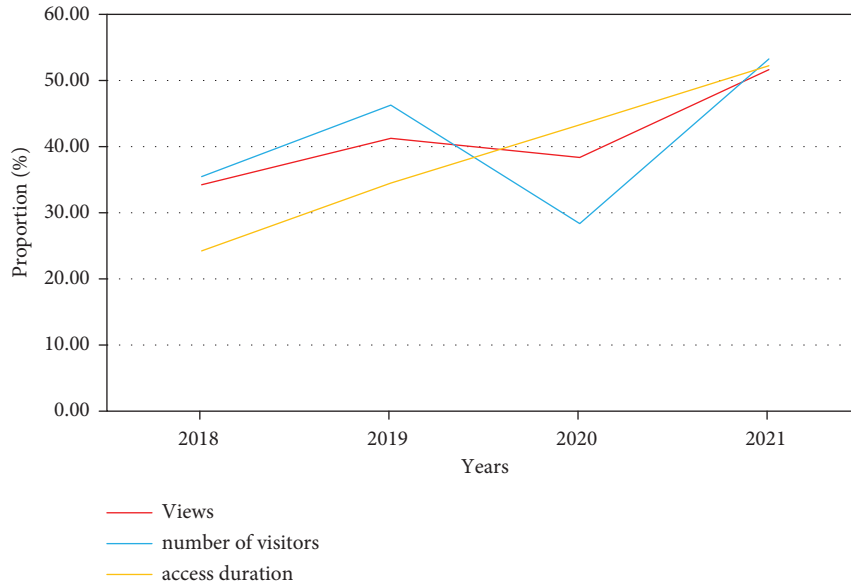


FIGURE 6: Proportion of visits to platform A and the number of visitors and time.

compared with other algorithms on the premise that the feature items are not correlated with each other. In general, when calculating the probability, each feature item in the category affects each other. The algorithm must satisfy the conditional independence assumption. That is, each feature item of the category is an independent event.

This paper uses the NBM algorithm to calculate the probability of occurrence of the document on the platform for all category options after the preword segmentation of the document to be processed. In theory, the category with the highest probability of occurrence is the classification result of the text document to be processed. According to the NBM classifier, the probability that test document  $X_r$  to be classified is classified into category  $C_i$  is

$$P\left(\frac{C_i}{X_r}\right) = \frac{P(C_i)P(X_i/C_r)}{P(X_i)} = \frac{P(C_i)P(t_1, t_2, \dots, t_m/C_i)}{P(t_1, t_2, \dots, t_m)}. \quad (1)$$

In formula (1),  $P(t_1, t_2, \dots, t_m)$  is constant for all categories, so it can be ignored in the calculation process. Formula (1) is simplified as shown in the following formula :

$$C(X_r) = \operatorname{argmax}_{C_i \in C} P(C_i)P\left(\frac{t_1, t_2, t_m}{C_i}\right). \quad (2)$$

The NBM classification algorithm must satisfy the premise that each condition is independent of each other. That is, in the specified category, the appearance of each feature item has no associated influence, which is independent of each other. Then there is

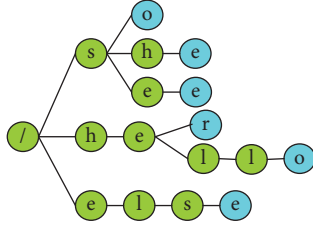


FIGURE 7: Building a Trie tree from a collection of pattern strings.

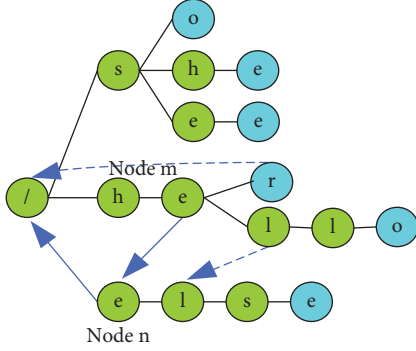


FIGURE 8: Hierarchical traversal adding failure pointer.

$$P\left(\frac{t_1, t_2, t_m}{C_i}\right) = \prod_{j=1}^m P\left(\frac{t_j}{C_i}\right). \quad (3)$$

Formula (2) can be rewritten from formula (3), as shown in the following formula:

$$C(X_r) = \operatorname{argmax}_{C_i \in C} P(C_i) \prod_{j=1}^m P\left(\frac{t_j}{C_i}\right). \quad (4)$$

Then the space vector of the test document to be classified is rewritten as follows:

$$X_r = \{(t_1, w_1), (t_2, w_2), \dots, (t_j, w_j), \dots, (t_m, w_m)\}, w_j \in \{0, 1\}. \quad (5)$$

In formula (5),  $w_j$  indicates the weight of feature item  $t_j$ , that is, whether the feature item  $t_j$  appears in the document  $X_r$  to be tested.

In the Bernoulli model, it is assumed that the event is that whether  $t_j$  appears in the document, and the number of feature items is  $m$ . The results are produced by subjecting this event to  $m$  Bernoulli trials. The probability that the test document  $X_r$  to be classified is classified into category  $C_i$  is

$$P\left(\frac{X_r}{C_i}\right) = \prod_{j=1}^m \left( w_j P\left(\frac{t_j}{C_i}\right) + (1 - w_j) \right). \quad (6)$$

In formula (6),  $P(t_j/C_i)$  is the probability that feature item  $t_j$  appears when document  $X_r$  to be tested belongs to category  $C_i$ . Formula (6) represents the probability that document  $X_r$  to be tested belongs to category  $C_i$ . The calculation method of  $P(t_j/C_i)$  here adopts the most widely used document frequency method:

$$P\left(\frac{t_j}{C_i}\right) = \frac{n_{ji}}{n_i}. \quad (7)$$

In formula (7), the calculation method of  $P(t_j/C_i)$  adopts the most common word frequency method to calculate as

$$P\left(\frac{t_j}{C_i}\right) = \frac{N_{ji}}{\sum_{j=1}^{|V|} N_{ji}}. \quad (8)$$

In formula (8),  $N_{ji}$  is the total number of occurrences of feature item  $t_j$  in all documents of the category  $C_i$ .  $|V|$  is the length of the word list of the document  $X_r$  to be tested (excluding repeated words), that is, the number of non-repetitive feature items contained in the document.

**3.5.2. Improvement of the Model Calibration Method of the Naive Bayesian Method.** In the traditional  $m$  estimation formula,  $m$  is the number  $|V|$  of document feature items contained in class  $C_i$ . The coefficient  $a$  (that is, the probability of feature item  $t_j$  appearing in all categories) is introduced at the  $m$  value to reduce the  $m$  value. Then, the following formula (9) can be obtained:

$$ml = a \times m = P\left(\frac{t_j}{C}\right) \times m = P\left(\frac{t_j}{C}\right) \times |V|. \quad (9)$$

From this, the  $p$ -value can be calculated as follows:

$$P = \frac{1}{ml} = \frac{1}{P(t_j/C) \times |V|}. \quad (10)$$

Using the  $p$  value and the  $ml$  value used to replace the  $m$  value, the posterior probability calculation formula of the improved  $m$  estimate is obtained as

$$P\left(\frac{t_j}{C_i}\right) = \frac{N_{ji} + 1}{\sum_{j=1}^{|V|} N_{ji} + P(t_j/C_i) \times |V|}. \quad (11)$$

**3.5.3. Polynomial Model.** NBM is a simplification based on the Bayesian method, which assumes that the attributes of a given target value are independent of each other. Although this simplification sacrifices the classification effect to a certain extent, it greatly simplifies the complexity of the problem in practical applications. Assuming that a piece of text in the platform data contains  $m$  feature words, there are a total of  $n$  text categories, and it is necessary to determine which category the text belongs to. Using the conditional probability formulas as formulas (12) and (13):

$$P\left(\frac{Ak}{B}\right) = \frac{P(A, Bk)}{P(B)} \quad (1 \leq k \leq n), \quad (12)$$

$$P\left(\frac{Bk}{A}\right) = \frac{P(A, Bk)}{P(Bk)} \quad (1 \leq k \leq n). \quad (13)$$

Formula (14) can be obtained then:

$$P\left(\frac{Ak}{B}\right) = \frac{P(B, Ak) * P(Ak)}{P(B)} \quad (1 \leq k \leq n). \quad (14)$$

Among them, A represents the category and B represents the text. When using this formula to calculate the probability of each category, since the denominator is the same, the problem can be transformed into calculating the numerator corresponding to each category, and the category with the largest numerator is the judgment result.

For the first half of the numerator, since the training set cannot contain all the text, it is very difficult to find the probability of this text appearing in a class. It cannot be directly set to 0, which is obviously not the correct value. At this point, formula (15) needs to be used:

$$P\left(\frac{B}{Ak}\right) = \prod_{x=1}^m P\left(\frac{Cx}{Ak}\right) \quad (1 \leq k \leq n). \quad (15)$$

Thus the probability of this text appearing in each category can be obtained.

**3.5.4. Evaluation of Classification Effect.** Since the feature conditional independence assumption of naive Bayesian has a certain influence on the classification results, feature weighting is used in many cases to reduce this influence. After this paper uses the classifier to classify the platform data, the final classification results can be analyzed and evaluated. Commonly used evaluation parameters include recall, precision, F1 value, and macro average.

The formulas for precision, recall, and F1 value are as follows:

$$\begin{aligned} P &= \frac{MP}{MP + FP}, \\ R &= \frac{MP}{MP + FN}, \\ F1 &= \frac{2 * P * R}{P + R}. \end{aligned} \quad (16)$$

Among them, MP represents the number of positive examples judged as positive examples. FP is the number of negative examples judged as positive examples. FN is the number of positive examples judged as negative examples, and TN is the number of negative examples judged as negative examples.

For multiclassification problems, macro averaging can evaluate the overall effect of the classification. The formulas for the macro precision, macro recall, and macro F1 value are as follows:

$$\begin{aligned} P_{\text{macro}} &= \frac{1}{n} \sum_{i=1}^n P_i, \\ R_{\text{macro}} &= \frac{1}{n} \sum_{i=1}^n R_i, \\ F1_{\text{macro}} &= \frac{2 * P_{\text{macro}} * R_{\text{macro}}}{P_{\text{macro}} + R_{\text{macro}}}. \end{aligned} \quad (17)$$

TABLE 1: Recognition results of platform A's negative review dataset.

Category	Improved			Naive bayes		
	algorithm + naive Bayes			P	R	F1
	P	R	F1			
Normal	0.8586	0.9616	0.9072	0.8852	0.9163	0.9005
Advertise	0.9475	0.9089	0.9278	0.9166	0.8838	0.8999
Ridicule	0.9460	0.8782	0.9109	0.8694	0.8656	0.8665
Abuse	0.9302	0.9233	0.9267	0.8822	0.8866	0.8844

TABLE 2: Recognition results of platform B's negative review dataset.

Category	Improved			Naive Bayes		
	algorithm + naive Bayes			P	R	F1
	P	R	F1			
Normal	0.8757	0.9436	0.9084	0.8656	0.9239	0.8938
Advertise	0.9042	0.8909	0.9058	0.8796	0.8757	0.8776
Ridicule	0.9222	0.8731	0.8968	0.9019	0.8718	0.8867
Abuse	0.9204	0.9135	0.9169	0.9198	0.8916	0.9055

## 4. Experimental Design and Results of Life-Sharing Platform

**4.1. Sharing Platform Experimental Data Results.** In this paper, two sets of experiments were designed to compare the naive Bayesian recognition scheme and the improved THT algorithm combined with the NBM algorithm using the negative review data-set of platform A and the review dataset of platform B, respectively. Naive Bayes adopted polynomial model, and weight calculation method adopted TFIDF weight method. The feature dimension value was 1000. The datasets of the two platforms had 4500 paragraphs of text, respectively, of which the normal category contained 4000 paragraphs, and the mocking, abusive, and advertising categories each contained 500 paragraphs. The 4100 paragraphs of text were used as the training set, and the remaining 400 paragraphs of text were used as the test set. Each group conducted 10 experiments, and the experimental results were averaged. The experimental language was Python3.6. The processor was i5-7200U. The frequency was 2.50 GHz. The memory was 8G, and the operating system was Windows10. The precision, recall, and F1 values on the two datasets are shown in Tables 1 and 2.

As can be seen from Tables 1 and 2, for the F1 value on the four categories, the new scheme proposed in this paper has a certain improvement over using naive Bayes alone. The macro precision, macro recall, and macro F1 values on the above two platform datasets are shown in Table 3.

From Table 3, it can be seen that on the platform A's bad evaluation dataset, the macro F1 value of the new scheme was 3.6% higher than that of using naive Bayes alone. On the B platform review dataset, the macro F1 score of the new scheme was 1.7% higher than that of naive Bayes alone. To sum up, compared with the traditional naive Bayesian scheme, the identification scheme combining the improved AC algorithm and NBM algorithm proposed in this paper



TABLE 3: Macro averaging.

Dataset	Improved algorithm + naive Bayes			Naive Bayes		
	$P_{macro}$	$R_{macro}$	$F1_{macro}$	$P_{macro}$	$R_{macro}$	$F1_{macro}$
A platform	0.9206	0.9178	0.9193	0.8884	0.8881	0.8882
B Platform	0.9056	0.9053	0.9055	0.8917	0.8908	0.8912

TABLE 4: Experimental result data.

Frequency	Number of records	Traditional algorithm running time (s)	Improve algorithm running time (s)
1	314086	14	35
2	801681	39	41
3	1302894	62	44
4	1910308	99	48
5	2503509	151	55
6	3000000	Not enough storage	57

TABLE 5: Classification results of users.

Frequency	Number of records	A	B	C	D
1	314086	587	31591	112534	15595
2	801681	1022	53642	303897	65805
3	1302894	1864	75293	447681	100219
4	1910308	2326	126431	769802	187084
5	2503509	2792	167201	989804	240135
6	3000000	3012	185866	1164961	298365

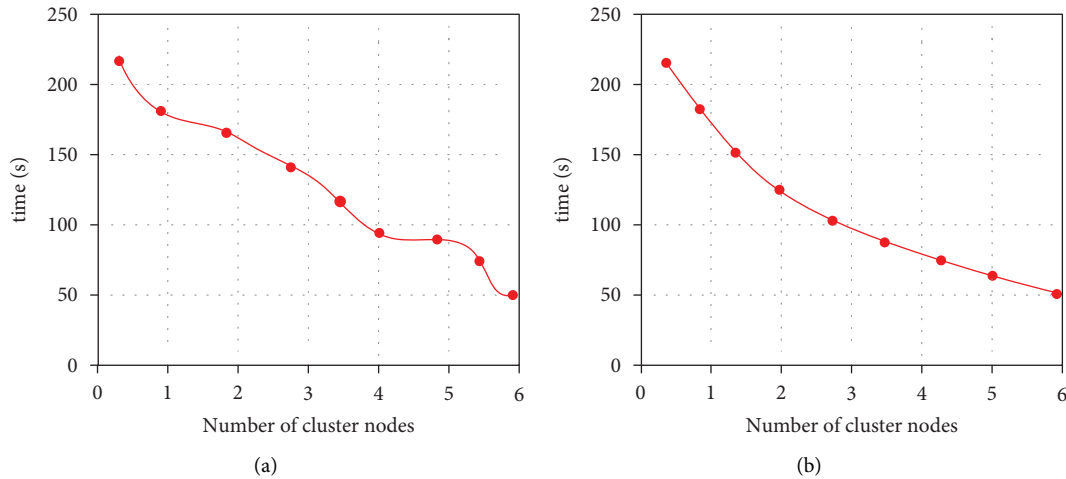


FIGURE 9: Comparison of node count results. (a) Results of the number of nodes of the traditional algorithm. (b) Results of the number of nodes of the improved NBM algorithm.

had a certain improvement in the spam content identification effect, which showed the effectiveness and feasibility of the new scheme.

#### 4.2. Single Machine and Improved NBM Algorithm Test.

In order to compare the running efficiency of the traditional serial naive Bayesian method and the improved NBM algorithm, the time required to process the same amount of data in a data computing node environment in the same configuration environment was analyzed. A total of 6 experiments were carried out, and the datasets used ranged

from 300,000 to 3 million. The experimental results are shown in Table 4.

At the same time, the corresponding user classification results were also obtained. According to the previous classification standards, they could be divided into four categories: A, B, C, and D. The classification results shown in Table 5 were obtained.

From the data in Tables 4 and 5, it can be seen that when the amount of data was relatively small, the processing time of the NBM algorithm based on it was similar to that of the traditional NBM algorithm, and even the time was slightly higher than that of the traditional algorithm. However, with

the increase of the training dataset, the advantages of the improved NBM algorithm became more and more obvious, and the time used was significantly less than that based on the ordinary NBM algorithm. When the number of data reached more than 10,000, the traditional serial NBM algorithm was abnormal due to insufficient system memory, but the improved NBM algorithm worked normally at this time.

*4.3. Improved NBM Algorithm Performance Test in Different Node Cluster Environment.* The content of the experiment was to compare the execution time of the traditional algorithm and the improved NBM algorithm under the condition of different numbers of data nodes in the cluster environment. This time included text data preprocessing time, data training time, and testing time. The data used was the platform data collected by the data acquisition system, and different data nodes were used to participate in the cluster cross-classification test experiment. The experimental results are shown in Figure 9.

From the analysis of the experimental results in Figure 9(a), it can be seen that when the traditional algorithm processed the same amount of data and when the number of data nodes in the cluster increased, the time for the system to process tasks fluctuated and decreased, and the overall performance of the cluster was not stable. As can be seen from Figure 9(b), using the improved NBM algorithm, the system processing time was continuously decreasing, and the overall performance of the cluster was stable, showing good scalability. It can be seen that the parallel naive Bayesian method based on the method had better efficiency and accuracy than traditional algorithms when dealing with massive data.

## 5. Conclusions

In recent years, the Internet of Things technology has developed rapidly. As the core technology in the Internet of Things architecture, the NBM algorithm has been applied in many fields. Aiming at the phenomenon that many consumers may encounter many difficulties when searching for information, this paper has developed a life and health-sharing platform based on NBM algorithm and AC algorithm according to actual needs through extensive research. This paper proposed the principle, physical structure, and logical structure of building a life and health medical data-sharing platform, and how to transform the existing information system to reduce costs to the greatest extent to achieve medical and health data sharing. In the study of the practical application of naive Bayes in text classification, an improved THT algorithm and NBM combined spam identification scheme was proposed, and the relevant experiments were carried out to verify its feasibility. However, the life and health-sharing platform based on the NBM algorithm proposed in this paper still has shortcomings in some places. It is necessary to continue to optimize the scheme combining the improved AC algorithm and the NBM algorithm, and find a more suitable weight calculation

method to further improve the recognition accuracy. On the other hand, the possibility of combining the improved AC algorithm with other classifiers needs to be studied.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

## Acknowledgments

The study was funded by Putian Life and Health Industry Development Plan (2022–2035), Project No.: 2022AHX162.

## References

- [1] M. Beier, C. Jansen, G. Mayer et al., "Multicenter data sharing for collaboration in sleep medicine," *Future Generation Computer Systems*, vol. 67, no. feb, pp. 466–480, 2017.
- [2] D. Fiorella, "Is data sharing the future of clinical trials?" *Journal of Neurointerventional Surgery*, vol. 9, no. 7, pp. 624–625, 2017.
- [3] Y. S. Zhang, L. Xia, J. Sang et al., "The BIG Data Center's database resources," *Hereditas*, vol. 40, no. 11, pp. 1039–1043, 2018.
- [4] L. Cui, X. Wang, and S. C. Su, "Techniques used on E-clinic portal access control," *International Journal of Computational Intelligence Research*, vol. 14, no. 1, pp. 79–88, 2018.
- [5] Z. Chen, R. Wang, and Z. Liu, "A novel complex network link prediction framework via combining mutual information with local naive Bayes," *Chaos*, vol. 29, no. 11, 113110 pages, 2019.
- [6] B. Indraja and K. Annapurani, "Classification of medicines using naive bayes classifier," *Research Journal of Pharmacy and Technology*, vol. 11, no. 5, pp. 1940–1944, 2018.
- [7] W. Zhang, Y. Jiang, G. Yin, and L. Yu, "Handling missing values in software effort data based on naive Bayes and EM algorithm," *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice*, vol. 37, no. 11, pp. 2965–2974, 2017.
- [8] X. Xu, R. Yang, and Y. Fu, "Situation assessment for air combat based on novel semi-supervised naive Bayes," *Journal of Systems Engineering and Electronics*, vol. 29, no. 4, pp. 768–779, 2018.
- [9] M. Behrouzian Nejad and M. Ebrahim Shiri, "A new enhanced learning approach to automatic image classification based on Salp Swarm Algorithm," *Computer Systems Science and Engineering*, vol. 34, no. 2, pp. 91–100, 2019.
- [10] Z. P. Fan, Y. J. Che, and Z. Y. Chen, "Product sales forecasting using online reviews and historical sales data: a method combining the Bass model and sentiment analysis," *Journal of Business Research*, vol. 74, no. May, pp. 90–100, 2017.
- [11] C. Mele, T. Russo-Spena, M. L. Marzullo, and A. Ruggiero, "Boundary work in value co-creation practices: the mediating role of cognitive assistants," *Journal of Service Management*, vol. 33, no. 2, pp. 342–362, 2022.
- [12] S. Sengan, K. R. Ganga Rama, and M. Osamah Ibrahim Khalaf, "Rajesh Babu" Markov mathematical analysis for comprehensive real-time data-driven in healthcare,"

- Mathematics in Engineering, Science and Aerospace (MESA)*, vol. 12, 1 page, 2021.
- [13] Y. Huang, N. D. Chatooh, T. Qi et al., "Health-related quality of life and its associated factors in Chinese middle-aged women," *Climacteric*, vol. 21, no. 5, pp. 483–490, 2018.
  - [14] N. Bidi and Z. Elberrichi, "Best features selection for biomedical data classification using seven spot ladybird optimization algorithm," *International Journal of Applied Metaheuristic Computing*, vol. 9, no. 3, pp. 75–87, 2018.
  - [15] S. Jeong, M. M. Tentzeris, and S. Kim, "Machine learning approach for wirelessly powered RFID-based backscattering sensor system," *IEEE Journal of Radio Frequency Identification*, vol. 4, no. 3, pp. 186–194, 2020.
  - [16] Z. Lv and H. Ko, "Introduction to the special issue on recent trends in medical data security for e-health applications," in *Proceedings of the ACM Transactions on Multimedia Computing Communications and Applications*, New York, NY, United States, May 2021.
  - [17] T. Liu and L. Ning, "Deep convolutional neural network and weighted bayesian model for evaluation of college foreign language multimedia teaching," *Wireless Communications and Mobile Computing*, vol. 2021, no. 3, Article ID 1859065, 7 pages, 2021.
  - [18] K. L. Dias, M. A. Pongelupe, W. M. Caminhas, and L. de Errico, "An innovative approach for real-time network traffic classification," *Computer Networks*, vol. 158, no. 20, pp. 143–157, 2019.
  - [19] A. Gadgil, V. Jagtap, and P. Kulkarni, "Smart refrigerator with recipe assistance," *International Journal of Synthetic Emotions*, vol. 8, no. 2, pp. 51–61, 2017.
  - [20] F. Alam and S. Pachauri, "Comparative study of J48, naive bayes and one-R classification technique for credit card fraud detection using WEKA," *Advances in Computational Sciences and Technology*, vol. 10, no. 6, 1743 pages, 2017.
  - [21] A. A. Elngar, K. M. Sagayam, A. A. Elngar, and A. A. Elngar, "Augmenting security for electronic patient health record (ePHR) monitoring system using cryptographic key management schemes," *Fusion: Practice and Applications*, vol. 5, no. 2, pp. 42–52, 2021.
  - [22] E. Yigit, K. Sabanci, A. Toktas, and A. Kayabasi, "A study on visual features of leaves in plant identification using artificial intelligence techniques," *Computers and Electronics in Agriculture*, vol. 156, no. 2, pp. 369–377, 2019.
  - [23] L. Zhang, "Design of a sports culture data fusion system based on a data mining algorithm," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 75–86, 2020.
  - [24] B. Ning, J. Wu, and H. Feng, "Spam message classification based on the naive bayes classification algorithm," *IAENG International Journal of Computer Science*, vol. 46, no. 1, pp. 46–53, 2019.