

Retraction

Retracted: Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

Copyright © 2023 Security and Communication Networks. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Long, "Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5418332, 9 pages, 2022.

Research Article

Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm

Juan Long 

Chongqing City Vocational College, Yongchuan, Chongqing 402160, China

Correspondence should be addressed to Juan Long; 023382@cqvc.edu.cn

Received 21 April 2022; Accepted 24 May 2022; Published 15 June 2022

Academic Editor: Chin-Ling Chen

Copyright © 2022 Juan Long. 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.

The emergence and development of Chinese insurance companies are affected by their own unique national conditions. The modern marketing concept lags behind, lacks the practical experience of scientifically formulating marketing strategies, and insurance practitioners lack marketing knowledge and the ability to absorb modern marketing achievements to guide practice. Therefore, China's insurance industry inevitably has many problems in insurance marketing. In recent years, with the rapid development of big data (BD) technology, artificial intelligence, and machine learning in engineering and academia, relevant data models have been well developed. The advantages of the decision tree are its good robustness, full sample mining, high precision, fast implementation, fast running speed, and low implementation cost. This paper studies the application of the decision tree classification algorithm under the guidance of BD in insurance marketing planning. The running results of the decision tree classification algorithm model show what factors will affect the accuracy and recall rate of customer churn decision-making. The predicted value and scoring value of users are extracted to test the model, and the results are within a reasonable range. The running time of this model is 2,320.36 s, which is more efficient than the 34 min 25 s of traditional SAS. Therefore, the model can be put into use, and it is necessary to establish a long-term and stable relationship with customers.

1. Introduction

With the development of the insurance industry in China, marketing plays an increasingly important role in the business activities of insurance companies. How to correctly and effectively select and manage marketing channels has become one of the focuses of insurance companies [1]. The potential of the insurance market is huge, but the market competition is also fierce. Therefore, how to narrow the gap between China's insurance industry and the insurance industry of developed countries and enhance the competitive strength of its own insurance market is worthy of our deep thinking. Marketing is an important part of insurance operations. Strengthening the marketing management of the insurance market plays an important role in improving the market competitiveness of China's insurance industry [2, 3]. Because the emergence and development of Chinese insurance companies are affected by their own unique national conditions, the concept of modern marketing lags behind

and lacks the practical experience of scientifically formulating marketing strategies, insurance practitioners lack marketing knowledge and lack the ability to absorb modern marketing achievements and apply them to guide practice, China's insurance industry inevitably has many problems in insurance marketing [4]. It is an urgent need for insurance companies to take integrated marketing as a breakthrough, innovate marketing mechanisms, and get rid of business difficulties. The deepening of financial system reform requires insurance companies to establish a marketing mechanism that truly adapts to the market economic system, which is the internal reason for promoting integrated marketing of insurance companies. As an important part of insurance marketing strategy, the choice of marketing channel is particularly important to China's developing insurance enterprises [5, 6]. Strengthening the marketing and marketing channel selection of Chinese insurance enterprises will not only help the improvement of their own management and technical level but also promote the

formation and enhancement of people's insurance awareness and have a far-reaching impact on the sustainable development of the insurance market [7].

In recent years, with the vigorous development of BD technology, artificial intelligence, and machine learning in engineering and academic circles, the relevant data models have developed very well and fully, and the advantages of the decision tree are its good robustness, full sample mining, high accuracy, fast implementation, fast running speed, and low implementation cost [8, 9]. Therefore, this paper introduces the concept of a BD-guided decision tree classification algorithm to alleviate the problem of heterogeneous data processing in traditional data processing and meet the needs of different data source storage media, introduces the scalable BD analysis model to obtain the characteristics of users' interest migration, applies the algorithm based on the decision tree algorithm model and takes the specific user data of an insurance company as an example to build application scenarios for model training and data prediction, and innovatively introduces the value rate to classify users to solve the problems faced by the company, such as long time, low efficiency, and low accuracy in processing massive user data [10, 11].

Using the BD-guided decision tree classification algorithm to classify insurance marketing planning and find "high-quality and low-quality" customers in a large customer base has important practical significance for life insurance companies to explore the market and avoid business risks [12, 13]. Marketing in the narrow sense is the direct sale of insurance products. In the broad sense, in addition to selling insurance products, marketing also includes a series of business activities such as the development and design of insurance products, the investigation and research of the insurance market, rate formulation, after-sales service, and so on. The business philosophy of insurance companies still stays in the concept of products and promotion but does not implement the modern basic marketing concepts such as "starting from the interests of customers" and "making customers satisfied." The competition of enterprises is still dominated by the competition of insurance premium rate, which seriously distorts the principle that insurance marketing is more suitable for nonprice competition, and is also contrary to the development trend of international insurance industry [14]. The BD-guided decision tree classification algorithm is generalized, and a limited number of discrete data is used to replace continuous data. For example, the age range of the insured is from 6 to 60 years old. In order to find the risk probability of customers of different ages, we are only interested in the age stage of customers and do not need to know the specific age of customers. Eight years old represents the insured aged 6–10; 15 years old represents the insured aged 11–20; and so on [15].

This paper studies and innovates the above problems from the following aspects:

- (1) An insurance marketing planning model based on a BD-guided decision tree classification algorithm is proposed. The decision tree classification algorithm is used to carry out BD theory and marketing mix

strategy in marketing to achieve the purpose of marketing. There are many new types of insurance, especially life insurance, which can basically meet the needs of the insurance market. However, in terms of market demand, the promotion and innovation ideas of insurance products are narrow and single in form. Insurance companies only invest a lot of manpower, material resources, and financial resources in the design of insurance products but neglect or belittle the promotion of products, resulting in many people not knowing insurance and alienating insurance.

- (2) The insurance marketing planning scheme of the BD-guided decision tree classification algorithm is constructed. The insurance marketing process guided by the BD decision tree classification algorithm is a process of seeking the balance between customers' needs and their own profits, and it is necessary to establish a long-term and stable relationship with customers. In order to achieve this effect, insurance companies need to establish a series of control measures such as analyzing the target market, implementing marketing plans and controlling the marketing process and finally achieve the profit target of insurance companies.

The paper is divided into five parts, and the organizational structure is as follows:

The first chapter introduces the research background and current situation of insurance marketing planning and puts forward and summarizes the main tasks of this paper. The second chapter introduces the related work of insurance marketing planning at home and abroad. The third chapter introduces the principle and model of the decision tree classification algorithm. The fourth chapter introduces the realization of the insurance marketing planning model of the BD-guided decision tree classification algorithm and compares the performance of the model through experiments. The fifth chapter is the full-text summary.

2. Related Work

2.1. Research Status at Home and Abroad. Honka and Chintagunta [16] proposed that with the development of China's socialist market economy and the deepening of the reform of the economic system, social security system, and education system, the insurance demand tends to be diversified, the group business is relatively reduced, and the decentralized business is greatly increased [16]. Putz et al. [17] proposed that insurance companies do not pay attention to analyzing the market environment, alienate policyholders, and only focus on developing new business channels, resulting in a decline in the public's trust in insurance agents [17]. Amoako and Okpattah [18] pointed out that the emergence and development of Chinese insurance companies are affected by their own unique national conditions, the concept of modern marketing is still relatively indifferent and lacks of practical experience in scientifically formulating marketing strategies, insurance practitioners lack marketing knowledge and lack the ability to absorb

modern marketing achievements and apply them to guide practice, and China's insurance marketing is destined to have one or another problem [18]. Aksoy [19] proposed that traditional marketing is to invest capital, acquire raw materials, design and produce products in advance, establish sales channels, sell products, and provide after-sales service. It is a traditional form of the value chain. Insurance marketing is the reverse operation of the traditional value chain. Its core mechanism is to take the needs of customers as the starting point, provide corresponding products and services, and establish marketing channels with their own characteristics, so as to form their own products and advantages [19]. Xia et al. [20] pointed out that the overall quality of insurance marketing personnel is not high and the service level of the industry is low, which has seriously damaged the reputation of the insurance industry. Due to the lack of professional ethics or insurance and related knowledge of some employees, there are often violations such as misleading statements, premium rebates, and malicious solicitation when promoting insurance [20]. Karimi et al. [21] proposed that under the circumstances of changing market demand and increasingly fierce market competition, insurance enterprises not only can meet the production and operation of existing products but must constantly develop new types of insurance to meet the needs of customers and market competition, as well as the need to expand market share [21]. Karimi et al. [21] put forward that with the advancement of economic system reform and social and economic development, insurance demand is diversified. Especially, with the development of the private economy and the increase of individual industrial and commercial households, the social demand for insurance is also evolving towards diversification, and the consumer groups are also evolving towards diversification, and the proportion of group consumption is decreasing year by year. Faced with this development trend, marketing is changing in a diversified direction [22]. Wang et al. [23] put forward that the marketing of insurance is the basic goal of selling its insurance products, and more importantly, it also includes various service means derived from insurance products and ideas between products and services, with the ultimate goal of maximizing the value of customers [23]. Nancy et al. [24] put forward that the concept of modern marketing is relatively weak, lacking the time and experience to scientifically formulate marketing strategies, which leads to the lack of marketing knowledge and the ability to absorb modern marketing achievements and apply them to guide practice [24]. Piao et al. [25] put forward that the purpose of insurance marketing is to maintain long-term business relations with customers and to develop through continuous performance growth. Therefore, insurance marketing should select valuable customers through screening and, on this basis, create a long-term win-win relationship with customers through effective means [25].

2.2. Research Status of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm. This paper studies insurance marketing planning through the

BD-guided decision tree classification algorithm. The realization of the marketing objectives of insurance enterprises needs the support of an efficient marketing organization. High performance is accomplished by high-quality business personnel, which is also the result of efficient marketing organization operation. For the decision tree algorithm, there are a large number of features in large data sets, including some redundant, low-quality, and even irrelevant features. Their processing not only consumes a lot of computing resources, leads to the bloated structure of the tree but also risks the accuracy of prediction. Take the customer database of a life insurance company as the data source. A total of 9,238 records of policy information of all customers who opened accounts since 1999 and 1,484 records of claim information of additional insurance claims of the above customers as of October 2005 were extracted. In order to make insurance enterprises have a qualitative leap and a good marketing result, we must improve the existing insurance marketing organization structure and make a fast and efficient response to the market, so as to better provide customers with satisfactory service quality. According to the needs of consumers, design and develop insurance products, effectively adjust marketing strategies, mobilize the enthusiasm and replication quality of marketing personnel, make consumers satisfied with insurance products and services, and enhance consumers' recognition and trust in insurance companies, so as to occupy a dominant position in the fierce insurance market competition. There are 4,940 records in the processed form. Data reduction: the methods of dimension reduction, discretization, and concept layer are adopted to reduce the data. Delete the attributes in the original attribute set that are not related to the mining task to reduce the dimension.

3. Principle and Model of Decision Tree Classification Algorithm

The training process of the decision tree is iterative splitting and growing from the root node to the bottom, layer by layer. This process seems simple, but if a single layer of iteration is split from the bottom, it will have high space complexity and memory consumption. The flow classification algorithm in the BD environment needs to effectively improve the training speed of the model and reduce resource consumption. The decision tree classification algorithm is a typical classification algorithm, which can analyze and process training data, summarize data rules, generate a decision classification model, and then use this model to analyze and process new data. The high sample size of BD directly leads to the fact that the decision tree cannot be completely built in memory, and it also consumes a lot of operation time. In order to process these data efficiently, a reasonable way is to use the distributed storage and bandwidth resources of the cluster and use tasks parallel methods such as MapReduce or MPI to reduce the computing load. Basic data processing: the original user data is entered to form a basic data lake, and the data is imported into HBase and Oracle databases. The data is preprocessed and selected for table selection, key

selection, and connector layer selection to match the data model. The basic wide table is formed through the preliminary basic processing cleaning and screening of the data table. On the basis of forming the wide table, the data washer is standardized; the sample data is simply described and counted; and the missing values are processed and standardized. The process of constructing a decision tree classification algorithm is similar to the behavior pattern of people making decisions. Overfitting will not only affect the accuracy of prediction but also make the decision rules complicated and difficult to understand. The structure of the decision tree is mainly determined by the measurement index of impurity and the method of postpruning. In the early research, the common methods adopted in these two aspects were introduced and evaluated, and the experimental results on some data sets showed their influence on decision tree size and prediction accuracy. When the initialization of the decision tree classification algorithm model is completed, the window is an empty queue. At this time, the decision tree classification algorithm will start to execute the classification module and the evaluation module. The algorithm reads the input test data stream, predicts every instance in the test data stream in real time through the established decision tree model, matches the predicted result with the premade class label, and then inserts the matched result into the ADDS window to detect the concept drift. The detailed growth process of the decision tree classification algorithm is shown in Figure 1.

The process of constructing a decision tree classification algorithm is similar to the behavior pattern of human decision-making. Given a data set s , which contains the values of multiple attributes and the classification to which it belongs, the first thing to do is to use some statistical methods to select an attribute a as the root node and divide the data set s into multiple subsets according to the value of attribute a . The tree branch and leaf node of the decision tree represent the test output and class label of data attributes, respectively. Therefore, the decision tree can be designed as the structure of the flow chart, and the classification of test data can be obtained by traversing the root and leaf nodes of the tree. The purpose of the decision tree algorithm is to find the classification rules contained in the data. Its core content is to construct a decision tree with high precision and small scale. A single decision tree classifier has limited ability to deal with problems and is prone to overfitting. The basic idea of ensemble learning is to use multiple algorithms or multiple classifiers to produce a better result. Pruning refers to replacing some subtrees with direct leaf nodes, and the class of leaves is represented by the class of most training samples contained in the replaced subtree. It can reduce the complexity of tree structure, improve the over-fitting situation to improve the prediction accuracy, quickly predict the results, and simplify the decision-making process. Back pruning was originally proposed by Breiman. Compared with the front pruning with vision effect problem, it refers to the pruning after the initial construction of a decision tree. In order to solve the problems of limited memory and training

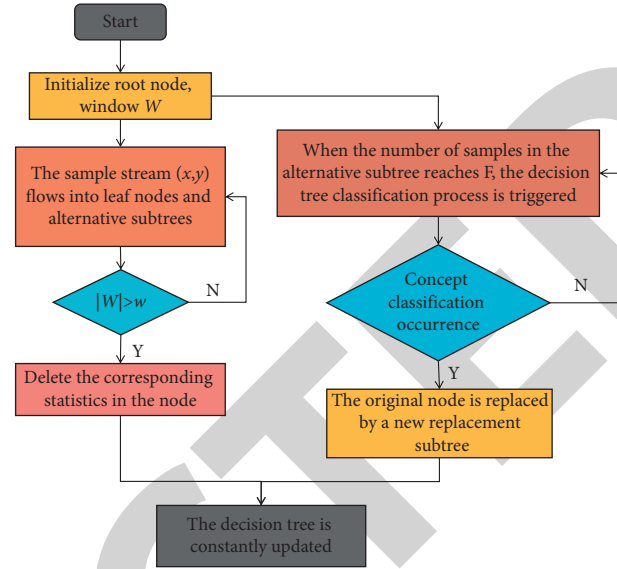


FIGURE 1: Flow chart of decision tree classification algorithm.

efficiency faced by the streaming data classification algorithm, the most effective way is to use the divide and conquer method to decompose the original computing task into several identical subtasks to deal with so that each computer node can balance the load. According to the idea of the divide and conquer method and the growth process of the decision tree, three parallelization strategies can be proposed: task parallelization, horizontal parallelization, and vertical parallelization. We select 1,000 samples in the training data set, and each sample contains the above 4 learning behavior attribute values, which can construct a 4-layer decision tree, and its structure is shown in Figure 2.

The classification model of the decision tree shows a series of IF-THEN rules. When a new test data enters, the data starts from the root node, goes along the branch to the leaf node, and finally gets the classification result. The data of learners' learning behavior is a series of vectors, each of which has different attribute units. IF-THEN rules can be constructed according to the model to get the classification of learning evaluation.

The memory occupation of each leaf node in the decision tree is $o(dvc)$, where c is the number of classified decisions, d represents the total number of attributes, and v is the maximum number of attribute values for each attribute. It is used to measure the ability of a given attribute to distinguish training samples. The calculation formula of information gain is shown in the following formula:

$$\text{Entropy}(S) = \sum_{i=1}^c -p_i \log(p_i), \quad (1)$$

$$\text{Gain}(S, A) = \text{Entropy}(s) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{S} \text{Entropy}(S_v). \quad (2)$$

Equation (1) is the calculation formula of information entropy, where p_i is the proportion of samples of different

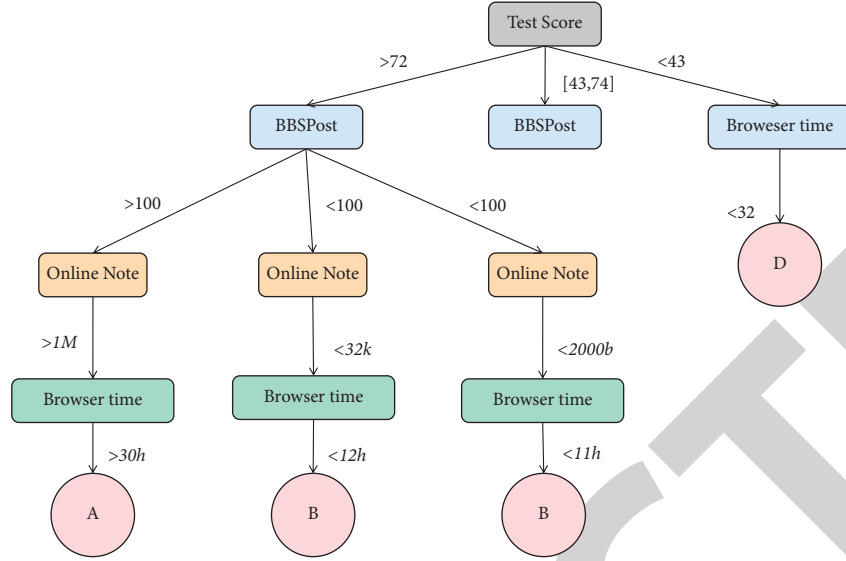


FIGURE 2: Classification model structure of decision tree.

categories C_i in sample set S . Equation (2) is the calculation formula of information gain, where $\text{Values}(A)$ is the set of all possible values of attribute A and S_v is the subset of attribute A in S (i.e., $S_v = \{s \in S | A(s) = v\}$).

With the continuous entry of data flow, the fitting degree of the decision tree model is getting higher and higher; at this time, the classification accuracy of the decision tree should be more and more stable, and the difference between the classification accuracy P_{real} in the sliding window and the optimal classification accuracy P_{optimal} is also getting smaller and smaller. We define $\Delta P = P_{\text{optimal}} - P_{\text{real}}$ to detect the error between the optimal classification accuracy and the real-time classification accuracy in the sliding window, assuming that we give the confidence δ and the confidence interval h for the accuracy error, which can be obtained by the inequality in Hoeffding.

$$P(\bar{X} - E[\bar{X}] \geq h) \leq e^{-2nh^2}. \quad (3)$$

Let \bar{X} be the estimated accuracy and $E[\bar{X}]$ be the true accuracy, then the meaning of the above inequality is that the probability that the deviation between the estimated accuracy and the true accuracy exceeds h is not greater than e^{-2nh^2} .

To find the most accurate Hoeffding upper limit, that is, find the minimum value on the right side of the equation have

$$s = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}. \quad (4)$$

Bring s into the equation and get

$$P(S_n - E[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \quad (5)$$

In the decision tree classification algorithm, Hoeffding inequality can be used to determine the minimum number

of samples needed for the node splitting in the decision tree in a given confidence interval δ .

According to symmetry, there is

$$P(-\bar{X} + E[\bar{X}] \geq h) \leq e^{-2nh^2}. \quad (6)$$

It can be obtained by formulas (5) and (6):

$$P(|\bar{X} - E[\bar{X}]| \geq h) \geq 2e^{-2nh^2}. \quad (7)$$

That is, for a given confidence δ , within the confidence interval where the expected accuracy $E[\bar{X}]$ width is $2h$, there is

$$\delta \leq 2e^{-2nh^2}. \quad (8)$$

A parallel window scheme based on a decision tree classification algorithm initializes multiple windows, divides real-time sample streams into S_1 and S_2 , monitors the hidden information distribution of the two streams, and obtains the formula according to the Poisson process and Hoeffding inequality.

$$P_r\{X \geq (1 + \varepsilon)E[X]\} \leq \exp\{-((1 + \varepsilon)1n(1 + \varepsilon) - \varepsilon)E[X]\}, \quad (9)$$

where $E(X)$ is the period value of the sample flow and ε is the limit of concept drift that needs to be detected. According to the Taylor series formula and Poisson process, the following formulas can be expected:

$$1n(1 + x) = \sum (-1)^{n+1} \cdot \frac{X^n}{n} = X - \frac{X^2}{2} + \frac{X^3}{3} \dots, \quad (10)$$

$$E[Y_n] = nE[X_n] = n\lambda_x. \quad (11)$$

The threshold for judging the conceptual drift of the sample flow is $\varepsilon = \sqrt{2\lambda 1n(2/\delta)(1/n_1 + 1/n_2)}$. In this paper, the sample flow is divided into two streams S_1 and S_2 ; then n_1 is the number of samples in the sample flow S_1 ; and n_2 is

the number of samples in the sample flow S_2 . δ is the confidence of Hoeffding Bounds, and λ is the mean value of samples in the sample flow.

4. Realization of Insurance Marketing Planning

4.1. Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm. From the perspective of the development history of the world insurance industry, China's insurance industry has been developing for a short time. However, with the deepening of the reform and opening up and the continuous improvement and development of the market economy system, China's insurance industry has been developing in the form of a blowout, especially since the reform and opening up, and its development speed and scale have far exceeded people's imagination, and the insurance market has made remarkable achievements. The insurance marketing of the decision tree classification algorithm obtains its own BD benefits on the premise of meeting customers' needs. It requires every member of the organization to think of customers and do their best to create more value for customers. The characteristics of insurance marketing and the particularity of insurance products determine the characteristics of insurance marketing. We can summarize the marketing characteristics of insurance products as follows:

① Change potential demand into actual demand

Most people's demand for insurance is potential, and insurance products are invisible and intangible abstract goods. Most people seem to have no urgency for it, especially for life insurance products. Therefore, the insurance marketer must change the potential demand of the insured into the actual demand through active marketing.

② Turn negative demand into a positive demand

Because most insurance products are related to people's life and death, for many people, their demand for insurance products is a negative demand. That is to say, people take negative evasive attitudes and behaviors towards insurance products because they do not like or understand them.

③ Change one-way communication into two-way communication

As a marketer of insurance products, one-way communication must be changed into two-way communication. That is to say, through active marketing, the information to be conveyed by enterprises will be transmitted to consumers through information media in a way that consumers can understand and accept, and consumers' feedback on information will be tracked and paid attention to, so as to collect consumers' opinions and reactions on the provided insurance products and timely adjust and improve service strategies to achieve customer satisfaction.

Now that China has joined the WTO, in order to gain a firm foothold in the further open Chinese insurance market

and remain invincible in the competition, many insurance companies use BD to meet the needs of economic development and enhance their ability to connect with the international insurance market. Modern marketing theory believes that using the decision tree classification algorithm to classify products is divided into three forms: tangible products, intangible labor services, and social behavior. In order to achieve this goal, fundamentally speaking, we should use a decision tree classification algorithm to carry out BD theory and marketing combination strategy in marketing in order to improve consumers' awareness. There are many new types of insurance, especially life insurance, which can basically meet the needs of the insurance market. However, from the perspective of market demand, the promotion and innovation ideas of insurance products are narrow, and the form is single. Insurance companies only invest a lot of human, material, and financial resources in the design of insurance products but ignore or despise the promotion of products, resulting in many people not knowing and alienating insurance. Insurance is an economic behavior in which the decision tree classification algorithm guided by BD predicts the possible uncertain events and collects the insurance premium, establishes the insurance fund, and transfers the risk from the insured to the insurer in the form of contract, and the majority participating in the insurance jointly share the loss of a few. Therefore, insurance marketing has different characteristics from other commodities. The process of insurance marketing guided by the BD decision tree classification algorithm is a process of seeking the balance between customer needs and their own profits. It is necessary to establish a long-term and stable relationship with customers. In order to achieve this effect, insurance companies need to establish a series of control measures such as analyzing the target market, implementing marketing plans, and controlling the marketing process and finally achieve the profit goal of the insurance company.

4.2. Experimental Results and Analysis. In this experiment, the data selection basically processes the row and column dimension data of the wide table. Because a wide table with about 25 attribute columns is generated from the data in the actual process, the data selection can avoid the disaster of high-dimensional data in data processing, and some data are normalized in the data processing process to adapt to the matching degree of the model as shown in Table 1.

It can be concluded from Table 1 that the running platform of this model is based on the Hadoop distributed file system, and its good high fault tolerance and high throughput data access are more suitable for the application of large-scale data sets. The application environment of this model is based on the basic running environment of HDFS, using Python data processing language, and the operating system version is centosrelease6 5 (final), set up 6 clusters. Relevant information of each device are as follows: Intel (R), e5606, @ 2.13 GHz, 2128.000 mHz, and cache size: 8,192 kb.

The running results of the decision tree classification algorithm model show which factors affect the decision of customer churn, and more valuable customer information

TABLE 1: Description and statistics of sample data.

Customer value rate category	Number of samples (PCs.)	Average age (years)	First year annualized premium	Income
4	1,251	43.0045	7,335.0658	104,969.65
3	5,381	38.1036	8,051.2526	170,750.08
2	924	41.4388	3,437.5842	81,722.66
1	277	45.9191	3,310.5149	99,677.58
0	10,456	49.1984	2,846.6048	46,312.29

TABLE 2: Full sample data operation model.

TN	FN	FP	TP	TestErr	Recall	Precision
363,235	27,784	11,216	57,168	0.0849024	0.67291584	0.67291569

can be obtained through evaluation. The evaluation methods include accuracy rate, recall rate, PR, ROC, and so on. Among them, the real TP: the sample type is correctly classified by the data model to predict the number of hits in the right class; FN: the sample type is misjudged by the data model as the quantity of other types; FP: the number of samples that do not belong to the correct category and are misjudged as the correct category by the data model; and true negative (TN): the sample type belongs to the correct category and is misjudged by the data model as the quantity of other types. The results of the whole sample data running model are shown in Table 2.

According to Table 2, the accuracy and recall rate are used here to extract the predicted value and scoring value of users for the model test. The obtained values are within the reasonable value range. The running time of the model is 2,320.36 s, which is more efficient than the 34 min 25 s run out of the traditional SAS. Therefore, this model can be put into use. In this experiment, in the field of agricultural economics, many foreign scholars have done a lot of empirical research on farmers' attitudes towards risk, especially in developing countries, and generally put forward the conclusion that farmers are risk-averse. In order to make this empirical conclusion in line with the reality of contemporary Chinese farmers, we visited the farmers in Fengshu village, Low Ping village, Liangshan village, and Baoyuan village in the Hongtang area of Xiangxiang City. Three surveys were conducted and compared in this experiment. The experimental results are shown in Figures 3–5.

As can be seen from Figures 3–5, the average importance of risk preference is about 8.5%; the average importance of risk aversion is about 6.3%; the average importance of risk aversion is about 4.6%; the average importance of risk aversion is about 10.2%; the average importance of risk aversion is about 7.1%; and the average importance of risk aversion is about 4.6%. Generally speaking, older people tend to avoid risks, and the proportion of young farmers who tend to take risks is much higher than that of older people.

In this experiment, the matching between the premium income of China's agricultural insurance and that of property insurance from 2015 to 2021 was analyzed, and two experimental investigations were conducted to compare them. The experimental results are shown in Figures 6 and 7.

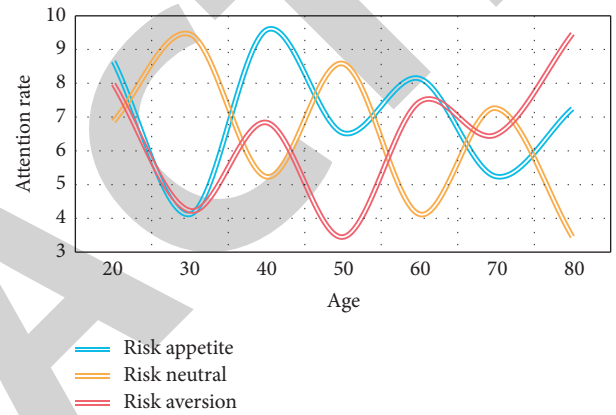


FIGURE 3: The farmers' attention to risks from the age structure.

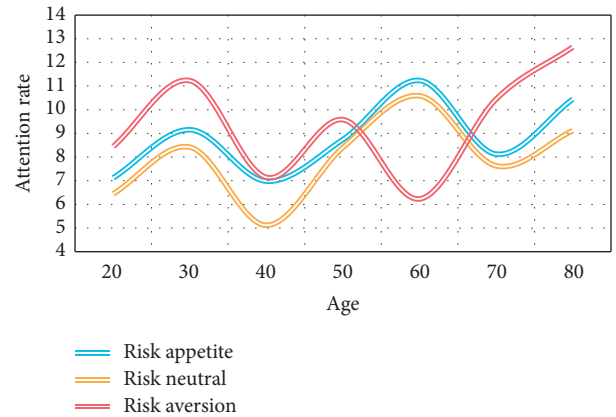


FIGURE 4: The farmers' attention to risks from the age structure.

As can be seen from Figures 6 and 7, the premium of China's agricultural insurance premium income in property insurance premium income from 2015 to 2017 is generally too low. Before the implementation of the policy pilot, the premium did not exceed 8,500. After the implementation of the policy pilot, although there was an obvious increase, it was less than 9,500. China's agricultural insurance premium income from 2018 to 2021 does not match the development of property insurance premium income. Before 2019, the property insurance premium increased by a large margin year by year, while the agricultural insurance premium has

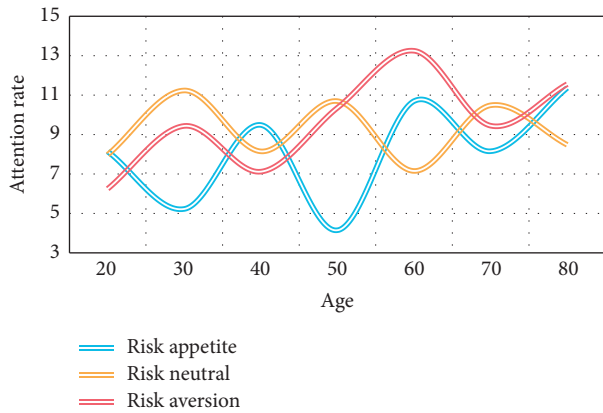


FIGURE 5: The farmers' attention to risks from the age structure.

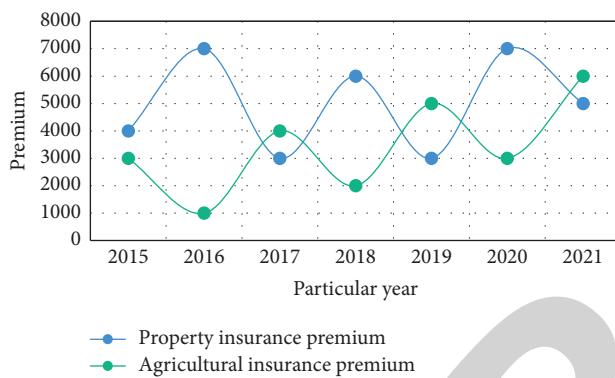


FIGURE 6: Trend correlation chart of agricultural insurance premium income and property insurance premium income in China.

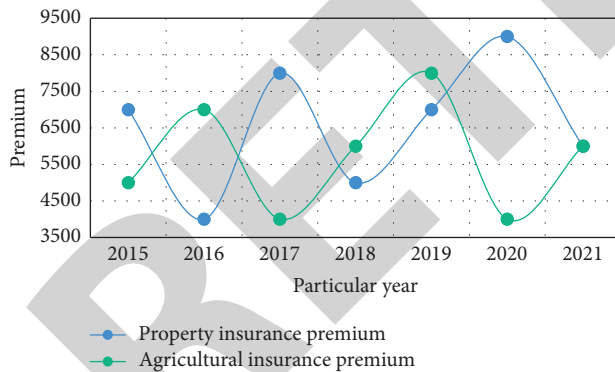


FIGURE 7: Trend correlation chart of agricultural insurance premium income and property insurance premium income in China.

been at a very low level year later, the development trends of the two are relatively close, but in terms of quantity, the premium income of agricultural insurance is still too low, and the premium income of property insurance is too low.

5. Conclusions

Through the research on the basic characteristics of decision tree classification algorithm, the combination point of decision tree classification algorithm and insurance user churn

rate is found, and the big data model of insurance marketing planning is established; The model is based on open source HDFS environment and has good scalability. Insurance agents play an important role in obtaining decentralized insurance consumption business, supplementing the sales capacity of insurance companies, and promoting the construction of a marketing network of insurance companies and have become an important channel for the business source of China's insurance market. According to the nondirectional classification of the nature of employment, insurance agents can be divided into personal agents, part-time agents, and professional agents. After selecting the target big data market, insurance companies under the decision tree classification algorithm should design different insurance types and marketing schemes for each target market to meet the insurance needs of different consumers. In this paper, for the big-data-guided decision tree classification algorithm, the use of conventional processing methods in data processing will inevitably lead to the loss of some data and the decline of prediction accuracy. The insurance marketing planning of the decision tree classification algorithm guided by big data is studied. The operation results of the decision tree classification algorithm model show which factors affect the decision-making of customer churn. The accuracy and recall rate are used to extract the predicted value and scoring value of users for the model test. The values are within the range of reasonable values, and the operation time of the model is 2,320.36 s, Compared with the 34 min 25 s of traditional SAS, it is more efficient. Therefore, this model can be put into use. Insurance is an economic behavior that predicts possible uncertain events and collects insurance premiums through big-data-guided decision tree classification algorithm, establishes an insurance fund, and transfers risks from the insured to the insurer in the form of contract, and the majority participating in insurance jointly share the losses of a few. Therefore, insurance marketing has different characteristics from other commodities.

Data Availability

The labeled datasets used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the 2020 Chongqing Municipal Education Commission Humanities and Social Sciences (No. 20SKGH337), 2020 Chongqing City Vocational College (No. XJSK202001009), and Chongqing City Vocational College Smart Retail Collaborative Innovation Centre.

References

- [1] J. Hong and S. H. Seog, "Life insurance settlement and the monopolistic insurance market," *Insurance: Mathematics and Economics*, vol. 81, pp. 36–50, 2018.

- [2] M. J. del Rio Olivares, K. Wittkowski, J. Aspara, T. Falk, and P. Mattila, "Relational price discounts: consumers' meta-cognitions and nonlinear effects of initial discounts on customer retention," *Journal of Marketing*, vol. 82, no. 1, pp. 115–131, 2018.
- [3] J. Dexe, U. Franke, and A. Rad, "Transparency and insurance professionals: a study of Swedish insurance practice attitudes and future development," *Geneva Papers on Risk and Insurance-Issues and Practice*, vol. 1, pp. 1–26, 2021.
- [4] A. Ra and Z. B. Wei, "Developing benchmarks to capture relative performance for sales force incentives decisions: lessons from B2B insurance industry - ScienceDirect," *Industrial Marketing Management*, vol. 92, pp. 55–71, 2021.
- [5] A. M. Roncancio, K. K. Ward, C. C. Carmack, B. T. Muñoz, M. A. Cano, and F. Cribbs, "Using social marketing theory as a framework for understanding and increasing HPV vaccine series completion among hispanic adolescents: a qualitative study," *Journal of Community Health*, vol. 42, no. 1, pp. 169–178, 2017.
- [6] M. H. J. Lin, S. N. Cross, W. J. Jones, and T. L. Childers, "Applying EEG in consumer neuroscience," *European Journal of Marketing*, vol. 52, no. 1/2, pp. 66–91, 2018.
- [7] S. H. Seog and J. Hong, "The efficiency effects of life settlement on the life insurance market," *Pacific-Basin Finance Journal*, vol. 56, pp. 395–412, 2019.
- [8] J. Abraham, C. Drake, D. W. Sacks, and K. Simon, "Demand for health insurance marketplace plans was highly elastic in 2014-2015," *Economics Letters*, vol. 159, pp. 69–73, 2017.
- [9] A. Nikiforov and E. Pilotte, "Security design, market risk and round quotes in the treasury bond market," *Applied Economics Letters*, vol. 26, no. 12, pp. 971–977, 2019.
- [10] J. Romaniuk, J. Dawes, and M. Nenycz-Thiel, "Modeling brand market share change in emerging markets," *International Marketing Review*, vol. 35, no. 5, pp. 785–805, 2018.
- [11] C. Demangeot, E. Kipnis, C. Pullig et al., "Constructing a bridge to multicultural marketplace well-being: a consumer-centered framework for marketer action," *Journal of Business Research*, vol. 100, pp. 339–353, 2019.
- [12] T. Stimpfling, N. Bélanger, O. Cherkaoui, A. Béliveau, L. Béliveau, and Y. Savaria, "Extensions to decision-tree based packet classification algorithms to address new classification paradigms," *Computer Networks*, vol. 122, pp. 83–95, 2017.
- [13] J. Yuan, Z. Zhou, N. Zhou, and G. Zhan, "Product market competition, market munificence and firms' unethical behavior," *Chinese Management Studies*, vol. 13, no. 2, pp. 468–488, 2019.
- [14] M. A. Adibi, "Single and multiple outputs decision tree classification using bi-level discrete-continues genetic algorithm," *Pattern Recognition Letters*, vol. 128, pp. 190–196, 2019.
- [15] E. Ventura-Molina, A. Alarcón-Paredes, M. Aldape-Pérez, C. Yáñez-Márquez, and G. Adolfo Alonso, "Gene selection for enhanced classification on microarray data using a weighted k-NN based algorithm," *Intelligent Data Analysis*, vol. 23, no. 1, pp. 241–253, 2019.
- [16] E. Honka and P. Chintagunta, "Simultaneous or sequential? Search strategies in the U.S. auto insurance industry," *Operations Research*, vol. 58, no. 3, pp. 255–257, 2018.
- [17] F. Putz, F. Murphy, M. Mullins, and L. O. Malley, "Connected automated vehicles and insurance: analysing future market-structure from a business ecosystem perspective," *Technology in Society*, vol. 59, Article ID 101182, 2019.
- [18] G. K. Amoako and B. K. Okpattah, "Unleashing salesforce performance: the impacts of personal branding and technology in an emerging market," *Technology in Society*, vol. 54, pp. 20–26, 2018.
- [19] H. Aksoy, "How do innovation culture, marketing innovation and product innovation affect the market performance of small and medium-sized enterprises (SMEs)?" *Technology in Society*, vol. 51, pp. 133–141, 2017.
- [20] H. Xia, J. Tang, J. Qiao, J. Zhang, and W. Yu, "DF classification algorithm for constructing a small sample size of data-oriented DF regression model," *Neural Computing & Applications*, vol. 34, no. 4, pp. 2785–2810, 2022.
- [21] F. Karimi, S. Sultana, A. S. Babakan, and S. Suthaharan, "Urban expansion modeling using an enhanced decision tree algorithm," *GeoInformatica*, vol. 2019, pp. 1–17, 2019.
- [22] Z. Mirzamomen and M. R. Kangavari, "Evolving fuzzy min-max neural network based decision trees for data stream classification," *Neural Processing Letters*, vol. 45, no. 1, pp. 341–363, 2017.
- [23] Y. Wang, S.-T. Xia, and J. Wu, "A less-greedy two-term Tsallis Entropy Information Metric approach for decision tree classification," *Knowledge-Based Systems*, vol. 120, pp. 34–42, 2017.
- [24] P. Nancy, S. Muthurajkumar, S. Ganapathy, S. V. N. Santhosh Kumar, M. Selvi, and K. Arputharaj, "Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks," *IET Communications*, vol. 14, no. 5, pp. 888–895, 2020.
- [25] M. Piao, C. H. Jin, J. Y. Lee, and J.-Y. Byun, "Decision tree ensemble-based wafer Map failure pattern recognition based on radon transform-based features," *IEEE Transactions on Semiconductor Manufacturing*, vol. 31, no. 2, pp. 250–257, 2018.