

Research Article

Application of C4.5 Algorithm in Insurance and Financial Services Using Data Mining Methods

Shihao Xuanyuan,¹ Shiang Xuanyuan ,² and Ye Yue³

¹School of Economics, Beijing Technology and Business University, Beijing 100048, China

²Graduate School of Arts and Sciences, Columbia University in the City of New York, New York 10027, USA

³School of Finance and Trade, Liaoning University, Shenyang 110036, China

Correspondence should be addressed to Shiang Xuanyuan; sx2292@columbia.edu

Received 1 June 2022; Revised 1 July 2022; Accepted 25 July 2022; Published 23 August 2022

Academic Editor: Mian Ahmad Jan

Copyright © 2022 Shihao Xuanyuan et al. 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 insurance financial management information system has accumulated a large amount of data as the insurance financial system has improved and the number of people investing in insurance has increased rapidly. The performance of the insurance agency significantly contributes to the industry's growth, which leads to economic prosperity. Different financial ratios were developed to investigate it, taking into consideration the insurance provider's stability, insolvency, profitability, and leverage. The profitability of organizations and insurers is used to evaluate the general effectiveness. In order to achieve this goal, this study examines the impact of insolvency, leverage, stability, scope, and impartiality of capital on the efficiency of Chinese life insurers. The study of financial statements examines a company's overall financial health throughout time. It is a method of identifying a company's financial assets and liabilities by integrating a statement of financial position and balance sheet features. It provides a systematic approach to assessing and evaluating the company's predicament. Using the experimental results, the scores of several insurance firms are compared, and their performance is described based on these results. The effective use of these data to assist decision-makers in developing more reasonable financial insurance investment policies have emerged as a significant challenge that must be addressed. This study utilized the decision tree C 4.5 mining algorithm to analyze insurance financial system data, identify key factors influencing insurance finance, and assist decision-makers in optimizing policy parameters. Finally, the consequence of an increase is analyzed using a previously unseen method to assess the precision of the prediction result.

1. Introduction

Insurance is a form of financial protection against a range of financial hardships. A contract among the involved individuals: the insurer and the protected recipient, defines this security. The insurance company is the organization that sells the policy, and the insured is the individual or organization that purchases the policy for the advantages it provides. In exchange for financial compensation, referred to as a high value, the insurance promises to absorb the responsibility of a protected entity against future eventualities [1]. In the event of an unexpected incident, the insurance company is required to pay the demand to the policyholder, i.e., the benefits are paid in full to the

beneficiaries as specified in the company's policy. Insurance policies vary depending on the type of event covered. Auto insurance, health insurance, travel insurance, property insurance, and life insurance are just a few of the many lines of business within the insurance organization. Therefore, the researcher is investigating various insurance and financial techniques in conjunction with various types of life security using artificial intelligence, big data mining, etc. [2]. Data analysis is the process of extracting potentially useful knowledge models or rules from large amounts of data. Data mining analysis employs association rules, classification, clustering, and other techniques. Data classification has a significant goal and task in data mining. As an important method of data mining classification, the decision tree data

mining algorithm makes efforts on deducing the decision tree representation's evaluation rules from a set of unordered and regular cases. It has features like high data analysis efficiency, intuitiveness, and simplicity of use [3].

The application of data mining in biomedicine mainly focuses on the research of molecular biology, especially genetic engineering. Its work in molecular biology can be divided into two types: one is to locate gene strings with certain functions from the DNA sequences of various organisms; functional proteins are similar to higher-order structural sequences. Database marketing and shopping basket analysis are two types of data mining applications used in marketing. The former's goal is to find new clients and offer them items using techniques including interactive searches, data segmentation, and model prediction [4]. The latter's role is to study market transaction data in order to perform an analysis of client purchasing behavior, which aids in determining shop shelf layout and encouraging sales. Data mining is mostly utilized in the banking industry for credit fraud modeling and trend analysis, as well as prediction, revenue analysis, risk assessment, and helping direct marketing efforts.

According to the reviews, data mining algorithms are frequently used to detect fraudulent insurance and financial policies by examining connections or linkages between various claim records, and the study developed a strategy for identifying insurance claims [5]. Kareem et al. applied data mining classification rules to evaluate related features and help control information disparities in false claims, thereby reducing health insurance fraud. The study provided an excellent explanation of why identifying health insurance fraud is one of the most challenging problems in the insurance industry. However, it did not give us the whole dataset descriptions utilized in their research [6]. Lin et al. proposed a framework in this work that could determine each characteristic in a given dataset. To analyze potential clients, these authors employed a sampling method in conjunction with a massive insurance firm data-mining training algorithm and proposed a collective random forest (RF) algorithm. They obtained insurance information about the company since China's life insurance establishment to use the specified technique. Furthermore, researchers evaluate the algorithm's performance using measurements and G-mean. The experimental result shows that, after being compared to the standard artificial methodology, the collective RF algorithm outperforms the support vector machine (SVM) and other classification models in efficiency and reliability within the unbalanced dataset [7]. This article presented several performance measures that may help determine the reliability of the techniques used in this study. Using instructional methods, the authors developed an early warning system for at-risk students. They used the orange data mining tool to conduct their research study. The findings of this study are critical in establishing elements for early warning systems and student achievement assessments that may be built for e-learning. Simultaneously, it will assist academics in selecting algorithms and preprocessing strategies for instructive data analysis. This study will assist us to identify numerous approaches in the orange miner tool that

will let us do our experiments on the national insurance dataset [8].

The process of mining theoretically beneficial material and information from large amounts of data is based on data mining. The decision tree classification method in data mining technology is used to discover hidden relationships and rules in data. It provides a theoretical foundation for policymakers to set and adjust parameters and also analyze and research several factors that limit fairness issues. To help decision makers make decisions, we should set the best parameters [9]. The research aims to see how machine learning and data mining algorithms might help insurance companies identify trends in different types of insurance claim evaluation categories. That is precisely what the whole research paper provides. In this study, the insurance data are used to perform claim analysis using a variety of categorization approaches. The feature selection techniques were utilized to decrease the complexity of the data and improve the results of the study.

The main contribution of the article is as follows:

- (i) Firstly, we present a conceptual framework for insurance and financial data mining methods that includes a comparison of performance measures between those used in insurance and those used in financial data.
- (ii) Secondly, data mining is the process of extracting potentially useful material and information from huge amount of information. In data mining technology, the decision tree classification approach is used to uncover hidden relationships and rules in data.
- (iii) Thirdly, after data preparation, parameter and class selection, decision tree building and pruning, analysis and assessment, and rule creation, the classification data mining process is complete.
- (iv) Finally, the decision tree C 4.5 mining technique was used to evaluate insurance financial system data, identify significant elements impacting insurance finance, and aid decision-makers in improving policy parameters in this paper.

The rest of this article is organized as follows: Section 2 shows related work, Section 3 shows the insurance and financial scheme detection and data mining, Section 4 shows the principle of the 2C4.5 algorithm, and Section 5 shows the algorithm application and experimental analysis of the decision. Finally, in Section 6, the research work is concluded.

2. Related Work

According to the literature review, insurance systems have undergone numerous significant changes in society, even during the global period, as proposed in this study. Rising stress in everyday life increases insurance demand. The authors of this study aim to determine how data mining benefits insurance companies, how its approaches improve insurance results, and how data mining aids decision-

making utilizing insurance data. Secondary research, observations from many periodicals, and studies, among other sources were used in the theoretical study [10]. According to Devale et al. information discovery has been constructed in financial firms to improve decision-making by employing knowledge as a strategic factor. The goal of this study is to investigate the application of various data mining approaches for data discovery inside the insurance industry. Current software is ineffective when it comes to showing data with these characteristics. The decision maker can outline the insurance activities' development in proposed data mining strategies to enable the existing life protection division's particular capabilities [11]. Yeo et al. analyzed insurance prices using mathematical optimization tools and data mining methods. In a competitive insurance market, one of the most essential factors in attracting clients is pricing. They employed K-means clustering algorithms to categorize customers, as well as a neural network to assess each classification's value perceptions [12]. Bian et al. evaluate the driver's risk level using driver behavior relevant information and a bagging-based classification model to assist the insurance firm in identifying the most acceptable business payment mechanism for various insurance policies. Customers' needs can be identified by collecting information and data from product customers and analyzing it using data mining techniques. The information gathered could be put to good use to help the organization progress [13]. Kumar et al. investigate using data mining and the analytical hierarchical process (AHP) to provide product recommendations. To begin with, clients of the insurance business were separated into groups built on their age and revenue. The AHP was then used to control the comparative weights of a set of factors in order to choose the best products for every cluster [14].

3. Insurance and Financial Scheme Detection and Data Mining

3.1. Bank Scheme. A bank scheme can be defined as anyone deliberately implementing, or seeking to implement, a strategy or contrivance to deceive a financial company: to collect money, credits, funds, stocks, investments, or additional assets possessed by and further down the protection or control of a financial company by using false, fraudulent pretenses, promises, or representations [15]. A bank scheme can be defined as unauthorized card use, uncommon operation activity, or communications made with an inactive card. A significant misstatement, financial deception, or deception concerning the potential mortgage or property on which a benefactor or investor relies to account or acquire is defined as a mortgage scheme. If you want to engage in a card and credit card scheme, you should get a loan. Unauthorized card use, unusual transaction activity, and transactions on an inactive card are all case studies of it now [16]. According to the FBI, money laundering is the act of criminals concealing or concealing the benefits of their crimes, or converting those monies into services and products. It gives criminals undue economic power by allowing them to infuse their illegal cash into the system,

corroding financial organizations and the cash source. According to Gao and Ye, money laundering is the process by which criminals launder dirty disgusting money to conceal its illegal origins and make it appear legitimate and clean [17].

3.2. Insurance Scheme. Customers, brokerages, insurance company workers, healthcare specialists, and others can perpetrate insurance products at many phases of the insurance process, including eligibility, claims, billing, rating, and application. Crops, automotive insurance schemes, and healthcare are the subjects of this study. According to the FBI, charging for medically unnecessary services, services not given, upcoding of services, upcoding of products, kickbacks, unbundling, excessive services, and duplicate claims are among the most prevalent kinds of the scheme. Crop insurance fraud occurs when policyholders fake or exaggerate crop losses due to natural disasters or income losses due to agricultural commodity price declines [18, 19]. Automobile insurance fraud encompasses staged accidents, needless repairs, and manufactured personal injuries. Figure 1 shows a framework of data mining applications.

4. The Principle of the 2C4.5 Algorithm

To create decision tree classifiers, the C4.5 algorithm is employed. The data gain values of each descriptive feature are compared in this procedure, and the attribute with the greatest value is chosen for categorization. The C4.5 algorithm creates decision trees based on the concept of information gain, with each classification's decision being linked to the target classification. The best way to assess uncertainty is to use entropy [20, 21].

$$s = - \sum_1 (P_i * \log(P_i)). \quad (1)$$

In this article, the effective reduction of the information descendant is referred to as information gain. Using this method, it is possible to determine which types of variables are chosen for classification at what level. Assume that there have been two classes, P and N, and that record set S comprises x P information and y N records. The following is the amount of data required to decide which category a record in record set S belongs to

$$\text{Info}(s) = \text{Info}(s_p, s_n) = -\frac{x}{x+y} \log \frac{x}{x+y} - \frac{y}{x+y} \log \frac{y}{x+y}. \quad (2)$$

Considering that variables D are used as the decision tree's root, the record set S is separated into subclasses $\{s_1, s_2, \dots, s_k\}$ with each s_i ($i = 1, 2, \dots, k$) containing x_i data from class P and y_i records from class N. Then, there is the quantity of data required to categorize all of the subclasses:

$$\text{Info}(D, S) = \sum_{i=1}^k \frac{x_i + y_i}{x + y} \text{Info}(s_{ip}, s_{in}). \quad (3)$$

If variable D is chosen as the classification node, the value of its data increments must be greater than the values of the other variables. The variable D data increment is

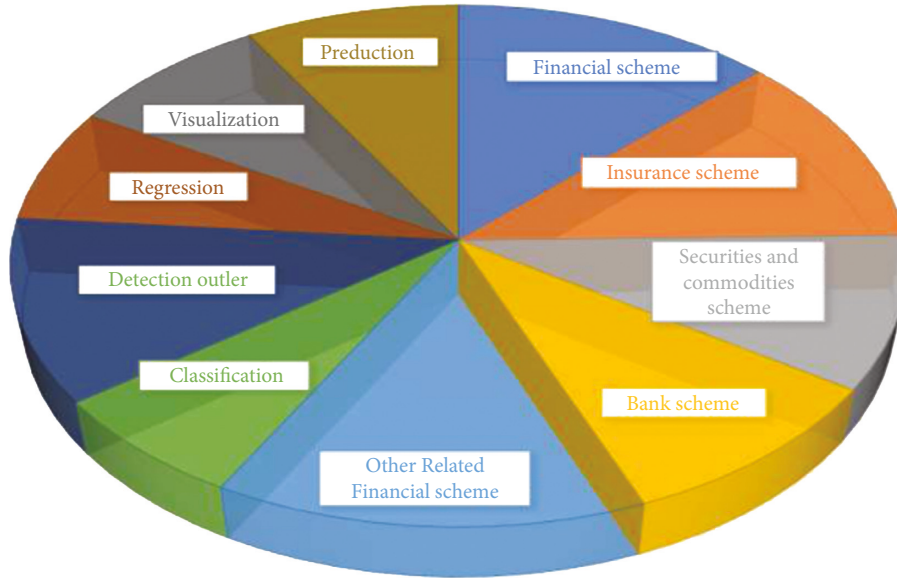


FIGURE 1: The framework for categorizing data mining applications.

$$\text{Gain}(D)_+ = \text{Info}(s) - \text{Info}(A, S). \quad (4)$$

A generic definition of the information gain function may be derived from this

$$\text{Gain}(D, S) = \text{Info}(s) - \text{Info}(D, S). \quad (5)$$

4.1. Pruning of Decision Trees. A pruning strategy for the decision tree is adopted when the fully grown decision tree is obtained. In this way, branch anomalies caused by noisy data and isolated nodes are eliminated. To address the problem of overadaptation of training data, decision trees are reduced. Statistically, approaches are typically used in the pruning process to remove the most unreliable branches and enhance the speed of identification and characterization, or the capacity to accurately classify data. The goal is to eliminate outliers and noise from the training set. The prepruning method and the postpruning approach are the two most common methods for pruning branches [22, 23].

4.1.1. Front Trimming Method. This approach is accomplished by halting the tree's formation in advance, i.e., deciding. Once a branch is terminated and the current node becomes a leaf node, it is important to continue dividing or splitting the subset of training samples at that node. Statistical significance detection or information gain can be used to assess branch development while building a decision tree. If the samples on a node are divided, the samples in the lower node will fall below a certain threshold. Then, we continued dissecting the sample set; selecting a suitable threshold is frequently challenging. A threshold that is too high will result in oversimplification of the decision tree, while a threshold that is too low will result in the failure to prune redundant branches.

4.1.2. Postpruning Method. This method is a popular decision tree-pruning strategy. The input of the postpruning algorithm is an unpruned tree T , and the output is a pruned

decision tree T_1 , which is the tree obtained after pruning one or more subtrees in T . The cost-complexity-based pruning algorithm is an exposed pruning method in which the bottom unpruned node becomes a leaf. It is designated as the category with the most categories in the samples it contains. The anticipated error rate is calculated after each nonleaf node in the tree is pruned, as well as the predicted error rate after a node is not pruned depending on the weight of a separate branch and the error rate of the separate branch. If pruning increases the projected classification error rate, the trimming will be unrestricted, and the branches of the consistent node will be maintained; then, the consistent node branch will be pruned and removed [24].

An independent test data set is used to evaluate the results after generating a sequence of trimmed decision tree applicants. The classification precision of the clipped decision trees is evaluated, and the tree with the lowest expected classification error rate and decision tree is maintained. In addition to the increased classifying error rate, the decision tree's embedding length can be used for decision tree pruning.

4.2. Decision Tree Rule Extraction. The decision tree can then be used to directly extract the corresponding decision rules after pruning. Decision trees are intuitive and simple to understand because the classification rules are expressed in the form and each rule is a path from the root. The leaf node then represents the specific conclusion. The nodes and edges above the leaf node represent the condition value of the corresponding condition [25]. Figure 2 depicts the direction from the decision tree to the decision rule.

5. Algorithm Application and Experimental Analysis of the Decision Trees (C 4.5)

5.1. Processing of Data. The operator information of the data table item is required by the C4.5 algorithm. It uses a type definition file, which is an ASCII file with a suffix of names,

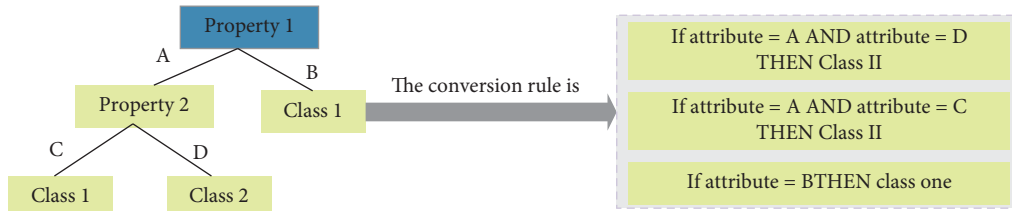


FIGURE 2: Transformation from decision tree to decision rule.

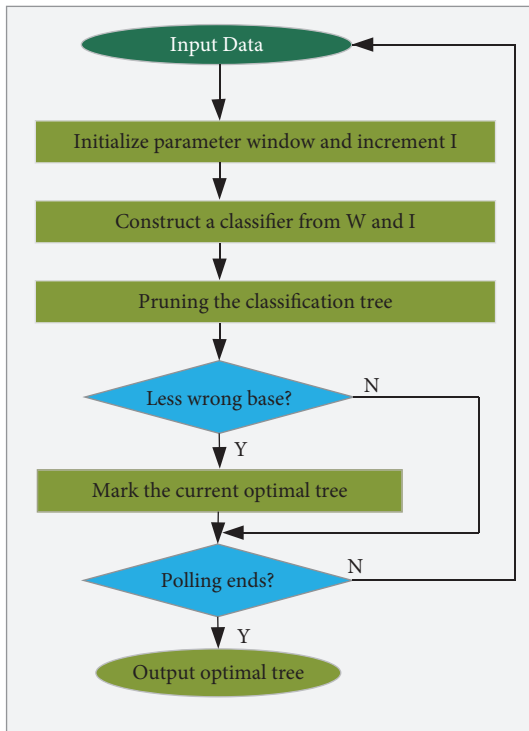


FIGURE 3: Mining algorithm processes.

to record the type of each attribute item or the range of potential values. According to the type description, the C4.5 algorithm will compute the gain of each feature item.

The computer calculates the gain value of each descriptive feature in a round-robin manner, compares the gain values of each attribute, chooses the attribute with the highest gain value for classifications, and ultimately builds an ideal decision tree. The program flow of the mining algorithm is shown in Figure 3. First, the initial variables of the program are set according to the initial input data: the window size and the value of the increment, and then different classification trees are generated in a continuous cycle. The pruned error rates are compared to find the best classifier.

5.2. Constructing a Decision Tree. According to the method of maintaining the correct rate of the judgment classification method, this article randomly selects two-thirds of the data from the preprocessing data as the training data of the C4.5 algorithm, obtains a decision tree from the training data, and outputs easy-to-understand rules.

- (1) Install and import the packages.

```
import pandas as pd.
import NumPy as np.
import matplotlib.pyplot as plt
import seaborn as sns
```

- (2) Import the data

```
df=pd. reads ("Insurance industry decision tree
case \data\ma_resp_data_temp.csv")
df. head ()
df. shape.
df. info ()
```

- (3) Split the data into train and test

```
clf= DecisionTreeClassifier ()
clf. fit (X_train, y_train).
clf. score (X_train, y_train).
0.5297687199450424.
clf.get_depth ()
42.
```

- (4) This is the classification accuracy of the decision tree model obtained without any parameter adjustment. Next, we will adjust the parameters of the model.

```
test_score = []
train_score = []
CV_score = []
for i in range (2,42): dtc= DecisionTreeClassifier
(max depth = i, random state = 100)
dtc.fit (X_train, y_train).
CV_score. append (cross_val_score (dtc, X_train,
y_train, cv = 5, n_jobs = -1). mean ())
train_score. append (dtc. score (X_train, y_train))
test_score. append (dtc. score (X_test, y_test))
```

- (5) To this end, the learning curve adjusts a single parameter, and selects the maximum depth max_depth for parameter adjustment, as shown in Figure 3.

```
plt. figure (dpi = 150). plt. plot (range (2, 42), test_score, label = "test_score"). plt. plot (range (2, 42),
train_score, label = "train_score"). plt. plot (range (2,42), cv_score, label = "CV_score ")
plt. legend ()
plt. show ()
print ("The best score is {}". format (np. max (cv_score)), "optimal depth: {}". format (np. argmax (cv_score) +2))
```

TABLE 1: The marketing-based insurance and financing of the investments.

Transaction	Items
T1	Market-based life insurance
T2	Based on the market
T3	Clint investment
T4	Market-based, tax-benefitting, and investment
T5	Market benefits based on taxation
T6	Market-based, tax advantage
T7	Life insurance, market-based insurance, tax benefits, and investment
T8	Tax benefits and life insurance
T9	Life insurance, market-based, and tax benefits
T10	Market-based life insurance and tax advantages

TABLE 2: Data of the insurance and financial industry.

KBM INDV ID	Resp flag	CA00	CA3	CA06	CA11	CA16	C210poo	C210psu
281478	0	4	0	5	1	1	52	71
290485	0	0	0	0	0	0	44	84
299949	0	0	0	0	0	0	38	48
314635	0	0	4	0	0	0	45	73
363702	0	0	0	0	0	0	32	14

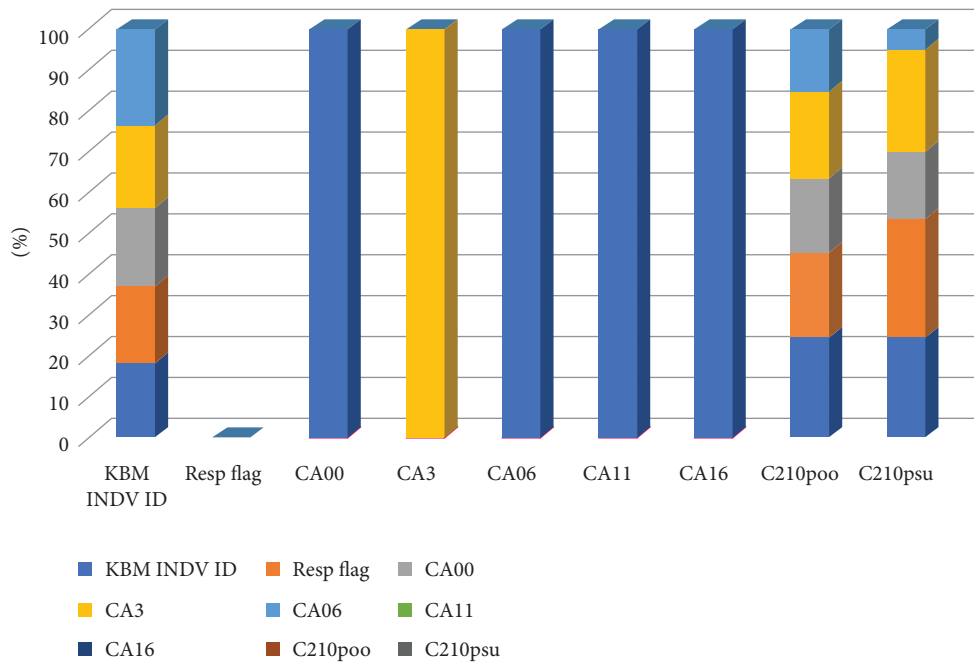


FIGURE 4: Comparison of all data of the insurance and financial industry.

The optimal score is 0.651036326061516 and the optimal depth is 0.8.

5.3. *Analysis of the Results.* In marketing research, financial institutions may employ association rules. The data examined in this case are material on the protection that customers obtain. The insurance provider can create a classification model that specifies which insurance is acquired when a policy is purchased. The company goals to benefit from the association amid several policies sold for varied goals based on all these realities. The same company

with customers having two insurance plans is far more likely to renew than those with only insurance. A customer with multiple policies is less likely to transfer than a customer with fewer policies. A company adds value and improves customer satisfaction by offering significant discounts and packaged products to customers, such as life insurance and investment plans, reducing the likelihood of the customer switching to a competitor. Table 1 shows the marketing-based insurance and financing of the investments.

The insurance company may design a sector-specific taxable profit, payment method, and insurance amount. These kinds of patterns can be recorded in a database. When

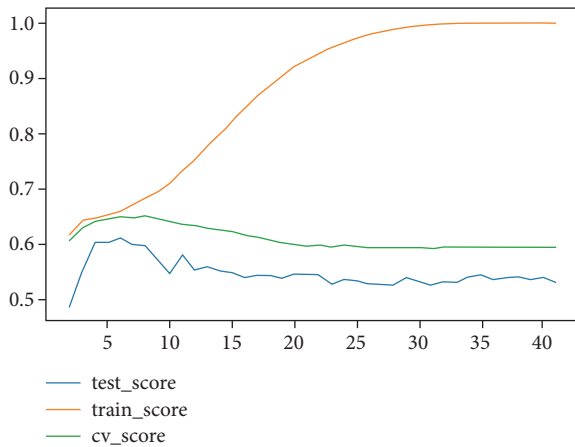


FIGURE 5: Adjusted parameters.

a consumer calls to purchase insurance, the agent can get information such as the client's age and income. This design may be associated with database records, and the agent can offer payment modes, payment amounts, and policy durations to customers based on the matching patterns. Table 2 and Figure 4 show the data for the insurance and financial industries.

Through the training of the C4.5 algorithm, the following decision classification tree is obtained, as shown in Figure 5. Y represents the insurance financial data security category; N represents the insurance financial data damage category. Using the training set and test set method to classify the accuracy test, the test result is that the correct identification rate for insurance finance is 96.25%.

6. Conclusion

Data analytics is being used in a variety of businesses throughout the world. Data mining and machine learning have a lot of potential for giving firms a competitive advantage over their competitors. This research is available in a variety of disciplines and uses a variety of analytical methodologies. In data mining, a decision tree is a common algorithm tool. The C4.5 algorithm is a decision tree algorithm with numerous applications and a high frequency. The classification data mining process is completed after data preprocessing, parameter and class selection, decision tree construction and pruning, analysis and evaluation, and rule generation. This article investigates the application of data mining techniques to insurance finance data statistics. Some factors affecting the insurance industry are initially obtained, and after experiments, the effect is relatively good, but there is no in-depth research on other influences in this experiment. Therefore, reinforce learning while gradually fixing its flaws in the future learning process. Possibly, comparing results with various categorization methods might be part of future research. Customer segmentation learning efficiency and performance might also be evaluated using computational complexity analysis. Other industries that might benefit from the strategy recommended in this study include retail, healthcare, food, and bookshops.

Data Availability

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

Conflicts of Interest

The author declares that he has no conflicts of interest.

References

- [1] Y. M. Wang and L. Xue, "Application of decision tree algorithm in early entrepreneurial project screening," *Scientific Programming*, vol. 2022, Article ID 3584196, 9 pages, 2022.
- [2] Y. Si, "Construction and application of enterprise internal audit data analysis model based on decision tree algorithm," *Discrete Dynamics in Nature and Society*, vol. 2022, Article ID 4892046, 11 pages, 2022.
- [3] B. Çığışar and D. Ünal, "Comparison of data mining classification algorithms determining the default risk," *Scientific Programming*, vol. 20198 pages, Article ID 8706505, 2019.
- [4] C. C. Lien, C. C. Ho, and Y. M. Tsai, "Applying fuzzy decision tree to infer abnormal accessing of insurance customer data," vol. 2, pp. 801–805, in *Proceedings of the 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, vol. 2, IEEE, Shanghai, China, July, 2011.
- [5] N. V. Chawla, "C4.5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure," in *Proceedings of the ICML* vol. 3, p. 66, CIBC, Toronto, ON, Canada, 2003, August.
- [6] R. Bhowmik, "Detecting auto insurance fraud by data mining techniques," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 4, pp. 156–162, 2011.
- [7] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, "An ensemble random forest algorithm for insurance big data analysis," *IEEE Access*, vol. 5, pp. 16568–16575, 2017.
- [8] F. Wu, X. Liu, Y. Wang, X. Li, and M. Zhou, "Research on evaluation model of hospital informatization level based on decision tree algorithm," *Security and Communication Networks*, vol. 2022, Article ID 3777474, 9 pages, 2022.
- [9] Z. Zheng, G. Zhang, and H. Shao, "Data mining and knowledge discovery: retrospect and prospect," *Information and Control*, vol. 23, no. 5, pp. 1–35, 1999.
- [10] K. Umamaheswari and Dr. S. Janakiraman, "Role of data mining in insurance industry," *An international journal of advanced computer technology*, vol. 3, no. 6, pp. 961–966, 2014.
- [11] A. B. Devale and Dr. R. V. Kulkarni, "Applications of data mining techniques in life insurance," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 2, no. No.4, pp. 31–40, July 2012.
- [12] A. C. Yeo, K. A. Smith, R. J. Willis, and M. Brooks, "A mathematical programming approach to optimize insurance premium pricing within a data mining," 2002.
- [13] Y. Bian, C. Yang, J. L. Zhao, and L. Liang, "Good drivers pay less: a study of usage-based vehicle insurance models," *Transportation Research Part A: Policy and Practice*, vol. 107, pp. 20–34, 2018.
- [14] P. Kumar and D. Singh, "Integrating data mining and AHP for life insurance product recommendation, Communications in Computer and Information Science," *Integrating Data Mining and AHP for Life Insurance Product Recommendation, Computational Intelligence and Information Technology*, Springer- Verlag Berlin, Heidelberg, pp. 596–602, 2011.

- [15] Culs Cornell University Law School, *White-collar crime: an overview*, Tianjing University, Tianjing, China, 2009.
- [16] Fbi Federal Bureau of Investigation, *Financial Crimes Report to the Public Fiscal Year* Department of Justice, Washington, D C, U S A, 2007.
- [17] Z. Gao and M. Ye, "A framework for data mining-based anti-money laundering research," *Journal of Money Laundering Control*, vol. 10, no. 2, pp. 170–179, 2007.
- [18] J. L. Kaminski, "Insurance Fraud, OLR Research Report," 2004, <http://www.cga.ct.gov/2005/%20rpt/2005-R-0025.html>.
- [19] *Coalition against Insurance Fraud*, "Learn about Fraud, Science Publisher, Hunan, China.
- [20] X. N. Li and S. Olafsson, "Discovering dispatching rules using data mining," *Journal of Scheduling*, vol. 8, no. 12, pp. 221–225, 2005.
- [21] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 106–181, 1996.
- [22] Roco, *Theoretical Method and Application Research of Data Mining in Database*, pp. 1–17, Hunan University, Hunan, China, 2004.
- [23] B. Wang, "Data mining technology and its application status," *Statistics and Decision Making*, vol. 41, no. 5, pp. 16–32, 2006.
- [24] L. Liu, G. Tu, and F. Yang, "Research on the application of data mining method based on decision tree in CRM," *Computing Technology and Automation*, vol. 25, no. 1, pp. 367–369, 2006.
- [25] B. Wang, "Analysis of the main methods and research directions of data mining," *Computer Simulation*, vol. 10, pp. 12–15, 2005.