JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 39, 797-807 (2023)
DOI: 10.6688/JISE.202307_39(4).0006

Based on Decision Tree Model to Analyze the Influencing Factors of Customer's Insurance Transactions

CHE-NAN KUO[1], YU-DA LIN[2], DUC-MAN NGUYEN[3] AND YU-HUEI CHENG[4,+]

[1]*Department of Artificial Intelligence*
CTBC Financial Management College
Tainan, 709 Taiwan
[2]*Department of Computer Science and Information Engineering*
National Penghu University of Science and Technology
Magong, 880011 Taiwan
[3]*International School*
Duy Tan University
Danang, 550000 Vietnam
[4]*Department of Information and Communication Engineering*
Chaoyang University of Technology
Taichung, 413310 Taiwan
E-mail: cn.kuo@ctbc.edu.tw[1]*; yudalinemail@gms.npu.edu.tw*[2]*;*
mannd@duytan.edu.vn[3]*; yuhuei.cheng@gmail.com*[4]

In recent years, global digitalization has developed rapidly. Driven by the gradual maturity of various technologies, the popularization of Internet and mobile devices, the Internet of Things and cloud computing services, the growth of various data has greatly increased and diversified data. The value of these data can be used to predict consumer behavior, differentiate user groups, develop effective marketing strategies, and create differentiated competitiveness. To predict consumer behavior in purchasing insurance products, this study collected 4,474 insurance transactions from a bank in Tainan, Taiwan. After data preprocessing, the number of available transactions is 3,430. In these organized transactions, we use the classification of the insurance product as the dependent variable and the features of the customer as the independent variable. Then, correlation analysis was performed by chi-square test, and uncorrelated factors were analyzed. Analyze influencing factors through a decision tree machine learning model. According to the analysis results of the decision tree model, the accuracy rate is almost 70%, and the most important influencing factors are the actual insurance fee and currency. These two influencing factors can be used as a reference for the precise marketing of Tainan Bank in Taiwan.

Keywords: machine learning, data analysis, decision tree, precision marketing, insurance transactions

1. INTRODUCTION

Precision marketing [1-3] is to use data and analytical tools to analyze the current market situation and the composition of customer groups to help companies narrow the scope of target customer groups. Compared with a large number of advertisements in traditional marketing, precision marketing pays more attention to the locking of target customer groups and achieves the highest benefits with the least budget. In traditional marketing thinking, in order to find the target customer group, it is usually through a large number of advertisements to increase the conversion rate and expand the scope of the

Received May 27, 2022; revised July 8, 2022; accepted July 24, 2022.
Communicated by Rung-Ching Chen.
[+] Corresponding author.

advertisement. However, as the cost of advertising increases, so does the cost of marketing, and the benefits of traditional marketing methods are getting lower and lower. Through precision marketing, when formulating marketing strategies, we can first analyze the target customer group, design products and marketing strategies according to customer preferences and habits, reduce marketing budgets, and improve advertising efficiency. To carry out precision marketing, it is necessary to master the list of target customer groups, target potential customers, and subdivide customer groups into different groups. Classification criteria can be factors such as age, region, occupation, *etc.* Marketing plans and budgets are planned according to different races.

To gain information about the potential value of customers, analysts need data on customers' buying behavior at their own firm as well as other firms in the market. Typically, companies only have data on their customers' buying behavior at their own company in their Customer Information File (CIF) [4]. Therefore, models are needed to predict the potential value of customers based on purchasing behavior in the CIF and any available socio-demographic data. Zeithaml points out that a lot of work needs to be done to identify the potential value of current customers [5]. There are already many models to predict individual transactions, and some work has been done to predict the buying patterns of focal suppliers [6]. Kim and Kim describe a model that can estimate the upsell potential of a single product or service provider [7], but apparently, no model is available that predicts the potential value of a customer in a multi-service environment. Verhoef and Donkers introduce a model for predicting the potential value of current customers. They compared the performance of competing models for predicting the potential value of customers. Second, on the management side, they provide CRM managers in multi-service industries with a framework that can be used to predict customer potential. The framework takes into account the data constraints that companies typically have by using only sociodemographic and transactional information from customer databases. The results can then be used as input for customer segmentation [8].

Due to the rapid development of information technology, it is easy to obtain a large amount of data. For business intelligence, it is very important to analyze data using technological tools to obtain information and further make marketing decisions. Strictly speaking, data is gold, and data likes intelligence. Given the importance of data analysis, this study collected 4,474 insurance transactions from a bank in Tainan, Taiwan. After data preprocessing, the number of available transactions was 3,430. Based on the analysis of the consumer features of insurance transactions, we discuss the important influencing factors of consumer features and insurance product classification. The research results can provide a reference for the precision marketing of Tainan Bank in Taiwan.

2. METHODS

2.1 Decision Tree

Decision trees are often used for data classification or prediction. It is a nonlinear data classification method that can effectively solve the shortcomings of only linear partitioning methods [9-11]. Decision trees can use different features to divide data into multiple single-class subsets and perform multiple single-class hierarchical operations. In the field of machine learning, decision tree is a representative method, which mainly distinguishes data categories according to different tree features, and then builds a predictive model. Therefore, decision trees have the advantages of being easy to explain, easy to use, and computationally efficient.

Decision tree uses tree structure to represent various paths that decision makers can take, uses probability to express possible states of uncertain factors, and obtains the expected value of various action paths through calculation. As data enters the decision tree from the root node, the features in the root node will determine the classification of each record, branch down into two or more classes, and then expand to new nodes. The decision tree model is shown in Fig. 1.

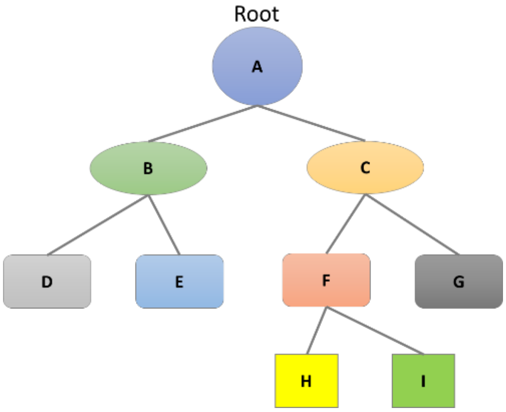


Fig. 1. Decision tree model.

Although a decision tree can clearly show every detail, if there are too many elements in a decision tree, it will become too large and complex. Therefore, it is necessary to prune the decision tree to remove too detailed nodes and branches to improve the presentation and effect of the decision tree. Increasing the size of the decision tree depends on the branching criteria. Common branching criteria include information gain, Gini coefficient, chi-square statistic, and information gain ratio, *etc*.

The first regression tree for decision tree models was developed by Morgan and Sonquist [12]. After that, many algorithms were developed for reference, such as CART [13] using the feature variable with the smallest Gini coefficient as the decision basis. CHAID (Chi-Square Automatic Interaction Detector) [14] uses the chi-square test to select the best features, which must be classified if used for quantitative data. ID3 [15] selects the best feature each time and does not use that feature after selection, which is a simpler and faster method. C4.5 [16] is an algorithm derived from ID3, which improves the problem that ID3 cannot use feature data.

Regarding the advantages of decision trees, Morgan and Sonquist [16] pointed out that decision trees are powerful prediction tools. The attraction of decision trees is that decision trees have rules, which can be expressed in words, making it easy for people to understanding, or translating into a database language such as SQL, makes data records that fall into specific categories easily searchable. Jing-Wei Wu [17] proposed that this method of finding rules based on tree diagrams produces results that are easy for users to understand. Users can analyze the characteristics of consumers through decision trees without any statistical and analytical knowledge.

Decision trees can be applied in many fields for research. The decision tree is used to predict the choice of ward type, and the accuracy of the prediction results is 66.3% [18]; the accuracy of the research results is 90% for predicting employee performance [12]; using fuzzy decision tree to predict Taiwan's weighted stock price trend, the accuracy of the research results is 93.4% after variable adjustment [19]; using three data mining algorithms-neural network, logistic regression and decision tree for predicting the survival of breast cancer patients. The average accuracy of the decision tree model is 87.3%, which is the best among these three models [20].

2.2 Gini Coefficient

Lorenz proposed the Lorenz curve in 1907 to evaluate the problem of regional income distribution, and proposed the corresponding curve between the cumulative population ratio and the cumulative income ratio. Take a sample of household units from the area to be tested and arrange them from small to large. The horizontal axis is the cumulative percentage of households, the vertical axis is the cumulative percentage, and the sum of households reaches 10% and 20% of the coordinates, respectively. When it reaches 100%, the corresponding cumulative income percentage. In this square chart, the percentage of wealth owned by each hundred-point family is accumulated, and the corresponding points are drawn on the chart to obtain a curve, the Lorentz curve, as shown in Fig. 2.

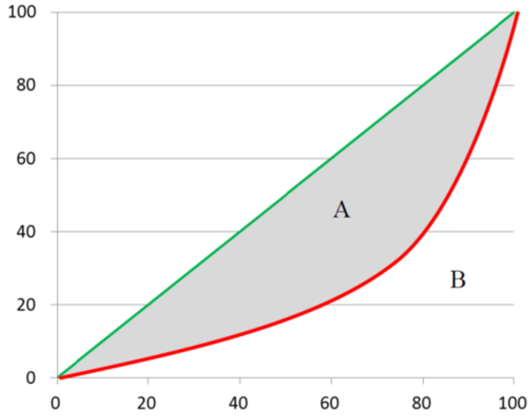


Fig. 2. Lorenz curve sketch map.

At the beginning of the 20th century, the Italian statistician Corrado Gini [11] defined the Gini coefficient based on the Lorenz curve. The Gini coefficient is often used as an indicator to calculate the degree of spread of value, and it is often used to measure the uneven distribution of various statistical values. It has applications in sociology and economics. As shown in Fig. 2, the calculation method of the Gini coefficient is A/(A+B), where the area between the Lorentz curve (red line) and the completely equal distribution (green line) is A, and the area below the green line means B. Therefore, it can be known that when the Gini coefficient is 0, it means that the red line and the green line are completely coincident, that is, the income is completely equally distributed. When the Gini coefficient is 1, it means that the red line completely coincides with the x-axis, that is, the income is completely unequal. Therefore, the Gini coefficient (between 0 and 1) is interpreted as: the closer the value is to 1, the more uneven the distribution of income; the closer the value is to 0, the more evenly the distribution of income. Therefore, when the Gini coefficient is 0, it means complete equality; below 0.2, it means a high degree of equality; 0.2 to 0.3 means it is still equal; 0.3 to 0.4 means tolerable; 0.4 to 0.6 means the gap is too large; above 0.6 means no equality is very high, and a value of 1.0 means no equality at all. When the Gini coefficient is higher than 0.6, society may be turbulent due to competition for power or wealth.

Let Y_i: the income of the ith family, N: the total number of households in the county, G: Gini coefficient, and λ_i: income level. The household income of each county and city needs to be ranked from small to large, according to the level of income, the lowest income is 1, the second is 2, and so on. Then, calculate the Gini coefficient of each county and city according to the following formula.

$$Y = (Y_1, Y_2, \ldots, Y_N), Y_1 \leq Y_2 \leq \ldots \leq Y_N \tag{1}$$

$$Y = \sum_{i=1}^{N} Y_i \tag{2}$$

$$y_i = \frac{Y_i}{Y} \rightarrow \sum_{i=1}^{N} y_i = 1 \tag{3}$$

$$G(Y) = \alpha\mu_y - \beta \tag{4}$$

where $\alpha = \frac{2}{N}, \beta = \frac{N+1}{N}, \mu_y = \sum_{i=1}^{N} \lambda_i y_i = Y; \lambda_1 = 1, \lambda_2 = 2, \ldots, \lambda_N = N.$

2.3 Data Sources

This study collected insurance product transaction data from a bank in Tainan, Taiwan, from 2006 to February 2020, with a total of 4,474 transactions.

2.4 Features in the Data

The study data range is the transaction data of consumers purchasing insurance products in a bank in Tainan, Taiwan. In the transaction data, we use the classification of insurance as the dependent variable, and the independent variables are selected from the features of consumers. The filtered features include: job_id, work_status, insured_income, actual_insurance_fee, age, currency, insured, insure, and insure_and_insured_the_same as shown in Table 1.

2.5 Processing for Missing Value

The total number of records of transaction data is estimated to be 4,474, of which about 193 records are insured and about 851 records have missing values. We cull these useless records from the data to make the data correct. Of the remaining 3,430 records, investment insurance cases accounted for 1,445 and general insurance cases accounted for 1,985.

Table 1. Filtered features.

Feature name	Description
job_id	Occupation type of insure
work_status	Work status of insure
insured_income	Yearly income of insure
actual_insurance_fee	The actual premium paid
age	Age of purchase
currency	Transaction currency
insured	Customer number of insured
insure	Customer number of insure
insure_and_insured_the_same	Whether the insure and the insured are the same person

2.6 Data Classification

Since points are represented in space, all features must be numerical. The features "job_id", "work_status", "insured_income", "actual_insurance_fee", "age", "currency", "insured", "insure" and "insure_and_insured_the_same" need to be converted to numerical values to be expressed in space. Ordinal data are often replaced directly with numerical values, such as medical, military, public education, finance and insurance, manufacturing, services, freelance, home management or retirement or students. Although they are feature classes, we can replace them with 0, 1, 2, 3, 4, 5, 6 because of the order of magnitude. Since the features "job_id", "work_status", "insured_income", "actual_insurance_fee", "age", "currency", "insured", "insure" and "insure_and_insured_the_same" are equivalent, we need to find a way to make them with the origin the distances are equal, and the problem can usually be solved by one-hot encoding method.

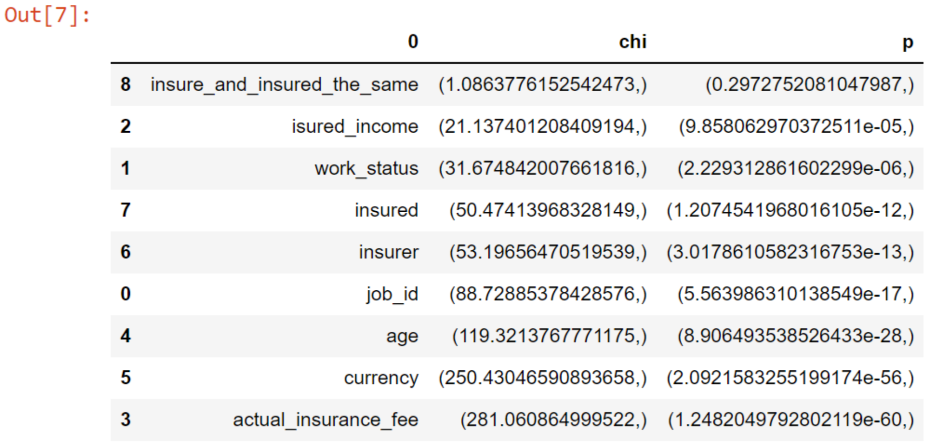
2.7 Data Slicing

After the above data processing, the data can be analyzed through the decision tree model. In terms of decision tree model analysis, data is usually divided into training data and test data when modeling and predicting. The cutting methods used in the data are expected to be 80% training data and 20% test data, respectively. Furthermore, the data between them is replaced multiple times with multiple training sessions via cross-validation to improve the prediction accuracy. However, we still have to be careful to avoid overtraining, which can lead to overfitting problems.

3. RESULTS AND DISCUSSION

3.1 Chi-Square Test to Filter Out Irrelevant Features

According to the chi-square test results in Fig. 4, the P value of the No. 8 feature (insure_ and_insured_the_same) is greater than 0.05, which is not related to the dependent variable (category of insurance), so this feature is filtered out. The p-values for the remaining feature numbers 0 to 7 were less than 0.05, respectively. Therefore, they can be used as independent variables for decision tree model analysis.



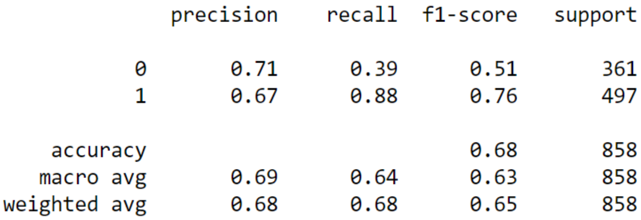
Out[7]:

	0	chi	p
8	insure_and_insured_the_same	(1.0863776152542473,)	(0.2972752081047987,)
2	isured_income	(21.137401208409194,)	(9.858062970372511e-05,)
1	work_status	(31.674842007661816,)	(2.229312861602299e-06,)
7	insured	(50.47413968328149,)	(1.2074541968016105e-12,)
6	insurer	(53.19656470519539,)	(3.0178610582316753e-13,)
0	job_id	(88.72885378428576,)	(5.563986310138549e-17,)
4	age	(119.3213767771175,)	(8.906493538526433e-28,)
5	currency	(250.43046590893658,)	(2.0921583255199174e-56,)
3	actual_insurance_fee	(281.060864999522,)	(1.2482049792802119e-60,)

Fig. 4. Chi-square test analysis.

3.2 Decision Tree Model for Accuracy Analysis

The study adopts the model analysis method as an optimized version of the CART algorithm in the decision tree model. The purity analysis method of each node is expected to use the Gini coefficient (maximum value is 1 and minimum value is 0). When the Gini coefficient is 0, it means that the fulcrum is the end point; if the Gini coefficient is not 0, it means that the fulcrum is not the end point. In addition, the training method of this model uses a 4:1 ratio of training samples to test samples. According to the analysis results in Fig. 5, the accuracy of investment insurance is 0.71, and the accuracy of general insurance is 0.67.

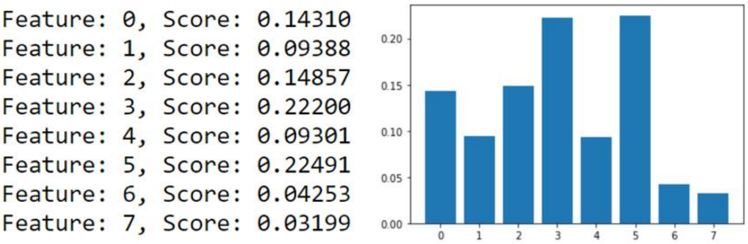


```
              precision    recall  f1-score   support

           0       0.71      0.39      0.51       361
           1       0.67      0.88      0.76       497

    accuracy                           0.68       858
   macro avg       0.69      0.64      0.63       858
weighted avg       0.68      0.68      0.65       858
```

Fig. 5. Chi-square test analysis.



```
Feature: 0, Score: 0.14310
Feature: 1, Score: 0.09388
Feature: 2, Score: 0.14857
Feature: 3, Score: 0.22200
Feature: 4, Score: 0.09301
Feature: 5, Score: 0.22491
Feature: 6, Score: 0.04253
Feature: 7, Score: 0.03199
```

Fig. 6. Feature analysis for features.

3.3 Feature Analysis

According to the analysis results in Fig. 6, the characteristics of each feature from high to low are feature 5: currency, feature 3: actual_insurance_fee, feature 2: insurance_income, feature 0: job_id, feature 1: work_status, feature 4: age, feature 6: insure, and feature 7: Insured. Therefore, feature 5 and 3 are most relevant to the classification of insurance.

3.4 Decision Tree Model Output

According to the analysis result of the decision tree model in Fig. 7, the customer feature can be judged by the root node feature condition, and the left or right branch can be forwarded. When reaching the leaf node, it is the insurance product type recommended for customer's feature. According to the decision tree output, when holding foreign exchange and the actual insurance amount is large, investment insurance is preferred; if the currency is Taiwan dollar, and the actual insurance cost and insurance income are small, general insurance is preferred; holding Taiwan dollar, pay more the actual insurance premium, when the age of the insured is younger or the gender of the insured is female, investment insurance is preferred. When holding Taiwan dollars, the actual insurance cost is low and the insurance income is high, and the insurance is given priority.

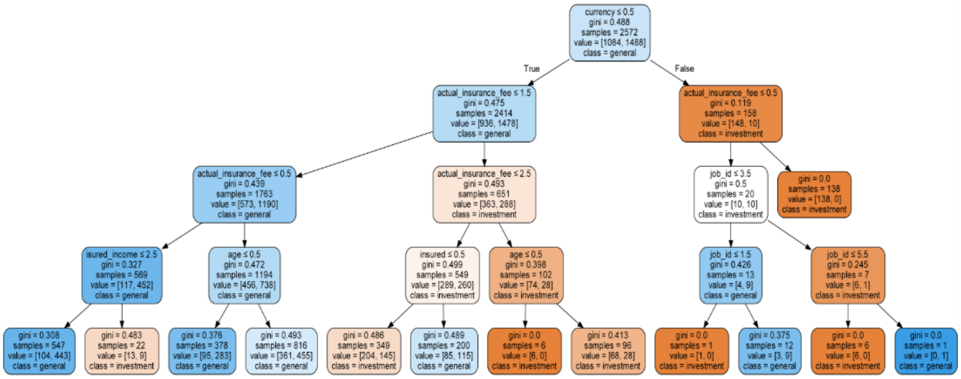

currency ≤ 0.5 gini = 0.488 samples = 2572 value = [1084, 1488] class = general
True
False
actual_insurance_fee ≤ 1.5 gini = 0.475 samples = 2414 value = [936, 1478] class = general
actual_insurance_fee ≤ 0.5 gini = 0.119 samples = 158 value = [148, 10] class = investment
actual_insurance_fee ≤ 0.5 gini = 0.439 samples = 1763 value = [573, 1190] class = general
actual_insurance_fee ≤ 2.5 gini = 0.493 samples = 651 value = [363, 288] class = investment
job_id ≤ 3.5 gini = 0.5 samples = 20 value = [10, 10] class = investment
gini = 0.0 samples = 138 value = [138, 0] class = investment
isured_income ≤ 2.5 gini = 0.327 samples = 569 value = [117, 452] class = general
age ≤ 0.5 gini = 0.472 samples = 1194 value = [456, 738] class = general
insured ≤ 0.5 gini = 0.499 samples = 549 value = [289, 260] class = investment
age ≤ 0.5 gini = 0.398 samples = 102 value = [74, 28] class = investment
job_id ≤ 1.5 gini = 0.426 samples = 13 value = [4, 9] class = general
job_id ≤ 5.5 gini = 0.245 samples = 7 value = [6, 1] class = investment
gini = 0.308 samples = 547 value = [104, 443] class = general
gini = 0.483 samples = 22 value = [13, 9] class = investment
gini = 0.376 samples = 378 value = [95, 283] class = general
gini = 0.493 samples = 816 value = [361, 455] class = general
gini = 0.486 samples = 349 value = [204, 145] class = investment
gini = 0.489 samples = 200 value = [85, 115] class = general
gini = 0.0 samples = 6 value = [6, 0] class = investment
gini = 0.413 samples = 96 value = [68, 28] class = investment
gini = 0.0 samples = 1 value = [1, 0] class = investment
gini = 0.375 samples = 12 value = [3, 9] class = general
gini = 0.0 samples = 6 value = [6, 0] class = investment
gini = 0.0 samples = 1 value = [0, 1] class = general

Fig. 7. The analysis result of decision tree model.

4. CONCLUSIONS

The study collected 4,474 insurance transactions from a bank in Taiwan Tainan. After the data pre-processing, the available number of transactions is 3,430. In these organized transactions, we explored the important influence factors between consumer features and the classification of insurance products. According to the analysis result of the decision tree model, the accuracy rate almost close to 70%, and the most important influence factors are the actual insurance fee and currency. These two influence factors can be used as a reference for the bank in Taiwan Tainan to precise the marketing. In addition, we can identify consumer groups by the decision tree model. For example, those who hold foreign currency and have more assets are more inclined to investment insurance. Therefore, investment insurance can be marketed to high-asset foreign currency customers, and general insurance to customers with less assets in the foreign exchange department of the bank; investment insurance is marketed to young people with assets, but those with low assets and high annual income can also suitable for investment insurance; general insurance can be marketed for low assets and for young people of holding Taiwan currency; high assets customers have the ability to undertake higher-risk insurance products than low assets customers. Customers with foreign currency holdings, occupations in medical care, military education and high incomes also like high-risk insurance products.

ACKNOWLEDGEMENTS

This work was supported in part by the Ministry of Science and Technology (MOST) in Taiwan under grant MOST110-2221-E-432-001, MOST110-2218-E-005-016, MOST110-2622-E-324-005, MOST110-2821-C-324-001-ES, MOST111-2218-E-005-009, and MOST 111-2221-E-432-001.

REFERENCES

1. J. Zabin and G. Brebach, *Precision Marketing: The New Rules for Attracting, Retaining, and Leveraging Profitable Customers*, John Wiley & Sons, 2004.
2. X. Yang, H. Li, L. Ni, and T. Li, "Application of artificial intelligence in precision marketing," *Journal of Organizational and End User Computing*, Vol. 33, 2021, pp. 209-219.
3. S. Zhao and J. Ma, "Research on precision marketing data source system based on big data," *International Journal of Advanced Media and Communication*, Vol. 7, 2017, pp. 93-100.
4. P. C. Verhoef, P. N. Spring, J. C. Hoekstra, and P. S. Leeflang, "The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands," *Decision Support Systems*, Vol. 34, 2003, pp. 471-481.
5. V. A. Zeithaml, "Service quality, profitability, and the economic worth of customers: what we know and what we need to learn," *Journal of the Academy of Marketing Science*, Vol. 28, 2000, pp. 67-85.
6. D. C. Schmittlein and R. A. Peterson, "Customer base analysis: An industrial purchase process application," *Marketing Science*, Vol. 13, 1994, pp. 41-67.
7. B. D. Kim and S. O. Kim, "Measuring upselling potential of life insurance customers: Application of a stochastic frontier model," *Journal of Interactive Marketing*, Vol. 13, 1999, pp. 2-9.
8. P. C. Verhoef and B. Donkers, "Predicting customer potential value an application in the insurance industry," *Decision Support Systems*, Vol. 32, 2001, pp. 189-199.
9. A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics*, Vol. 18, 2004, pp. 275-285.
10. J. R. Quinlan, "Learning decision tree classifiers," *ACM Computing Surveys*, Vol. 28, 1996, pp. 71-72.
11. Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, Vol. 27, 2015, p. 130.

12. J. N. Morgan and J. A. Sonquist, “Problems in the analysis of survey data, and a proposal,” *Journal of the American Statistical Association*, Vol. 58, 1963, pp. 415-434.
13. G. V. Kass, “An exploratory technique for investigating large quantities of categorical data,” *Journal of the Royal Statistical Society*, Vol. 29, 1980, pp. 119-127.
14. C. Gini, *Variabilità e Mutabilità: Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche*, Cuppini, 1912.
15. J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, Vol. 1, 1986, pp. 81-106.
16. G. S. Linoff and M. J. Berry, *Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management*, John Wiley & Sons, 2011.
17. C.-W. Wu, “Application of decision tree and association rules for the analysis of student achievement – A case study of a vocational school in Tainan,” Master Thesis, Department of Information Management, Southern Taiwan University of Science and Technology, 2014.
18. W.-L. Chiang, “Investigation on the selection of ward types with decision trees: Case study of a regional hospital,” Master Thesis, Department of Information Management, National Chung Cheng University, 2019.
19. G.-Y. Chen, “Decision tree, logis regression and neural network predicts employee performance comparative study,” Master Thesis, Institute of Human Resource Management, National Central University, 2017.
20. Y. F. Lu, “Predicting breast cancer patients’ survivability: The comparison of using three data mining methods – Artificial neural network, logistic regression and decision tree,” Master Thesis, School of Public Health, National Defense Medical College, 2006.

Che-Nan Kuo received the BS degree in the Department of Computer Science from the Tunghai University, Taichung, Taiwan in 2002, and the MS and Ph.D. degrees from the Department of Computer Science and Information Engineering at the National Cheng Kung University, Tainan, Taiwan in 2004 and 2009, respectively. Now, he is an Associate Professor in the Department of Artificial Intelligence, CTBC Financial Management College, Tainan, Taiwan. He has many excellent research papers about fault-tolerant computing which have been published on some famous journals, such as Theoretical Computer Science, Discrete Applied Mathematics, Information Sciences, and Computers and Mathematics with Applications. His current research interests include interconnection networks, discrete mathematics, computation theory, graph theory, algorithm analysis, machine learning and data science.

Yu-Da Lin received the MS and Ph.D. degrees in Computer Science from Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, in 2011 and 2015, respectively. He is currently an Assistant Professor with Department of Computer Science and Information Engineering, National Penghu University of Science and Technology. He has authored/coauthored more than 100 refereed publications. His main research interests include artificial intelligence, biomedical informatics, bioinformatics, and computational biology. Dr. Lin is a Senior Member of the IEEE Tainan Section, IEEE Young Professionals, and IEEE Computational Intelligence Society Membership. He is also an Associate Editor of IEEE Access and Frontiers in Genetics Journals.

Duc-Man Nguyen received the Bachelor of Information Technology (B.S) from Duy Tan University, Vietnam. He earned his Master of Science (MSc) in Computer Science from the Danang University of Technology in 2009 and the Ph.D. degree in Computer Science from Duy Tan University in 2020. He earned the certification of HP QTP, Load-Runner, Quality Center from HP Train the Trainers, Fagan Software Inspection Method Certification from ECCI Group, Certification of Software Architecture and Design, Capstone project in Software Engineering from Institute for Software Research-Carnegie Mellon University. He has published more than 17 research papers in National, International Journals and Conferences (SCIE, Scopus, Book chapter, IEEE, Springer). He has more than 19 years of experience in software development and mentoring for Capstone projects. At present, Dr. Man is working as Dean of International School, Duy Tan University. His research interests include software engineering, software testing, mobile application testing, ML/DL, teaching methodology, data analytics/data science, digital transformation and IoT.

Yu-Huei Cheng received the MS and Ph.D. degrees from Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan, in 2006 and 2010, respectively. He crosses several professional fields including biological and medical engineering, electrical and electronic engineering, and information engineering. He is currently a Distinguished Professor of Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung, Taiwan. His research interests include artificial intelligence and internet of things, bioinformatics, computational intelligence, robotics, and unmanned vehicle.