International Journal of Multimedia and Ubiquitous Engineering
Vol. 9, No. 10 (2014), pp. 209-222
http://dx.doi.org/10.14257/ijmue.2014.9.10.20

Research on Forecast Model and Application of Customer Loyalty under the Background of Big Data

Yihua Zhang, Yuan Wang*, Chunfang He and TingTing Yang

School of Business Administration, Jimei University, Xiamen 361021, China
*ywward@jmu.edu.cn, *wangyuan@jmu.edu.cn, hcf051@163.com, 1363762128@qq.com*

Abstract

As competition of Chinese securities industry becomes increasingly fierce, customer loyalty management has become the core elements of enterprise competition. Combining cross industry standard process for data mining (CRISP-DM) process model, flow visualization, this article adopts the mining method of large data to find the key factors to describe customer loyalty, and establishes customer loyalty prediction model of two dimensions, according to the model can effectively predict customer loyalty degree, support enterprise decision makers develop corresponding retention strategies.

Keywords: *Big Data, Customer Loyalty, Data Mining*

1. Introduction

At present, China's securities industry is facing an increasingly competitive environment. On one hand, the business scope of financial institutions related to capital market such as banks, insurance, trust continues to expand; on the other hand, with the improvement of China's opening degree, the Chinese market with wide development prospect will become the focus for many foreign financial investment corporations to compete, which further exacerbated the competition of domestic market. Foreign security companies have strong financial strength, abundant management experience and extensive international marketing network, which make the domestic security companies face more intense competition in terms of talents, large customers, and financial innovation, etc. The two factors show that the security companies should focus on the management and maintenance of their customers, especially the management and maintenance of company's large customers and customers with high loyalty. Therefore, the study of customer's loyalty degree has been increasingly emphasized.

In the customer relationship management of security companies, how to use date mining technique to excavate loyal customers to improve the decision support function of customer relationship management has become an important issue which the current security industry faces. Larry's study (2001) found that the cost required to maintain old customers is far less than the cost of attracting new customers, and the longer the customer relationship is, the more sales opportunities and corporate profits are. Thus, the customers of security companies, especially the customers with high loyalty degree, are the key to the companies' profitability size, and are one of the company's core competitiveness. Wouter Buckinx and Dirk Van den Poel(2004) focused on the treatment of a company s most loyal customers in a non-contractual setting and built a model in order to predict partial defection by loyal clients using three classification techniques: Logistic regression, automatic relevance determination (ARD) Neural Networks and Random Forests [1]. Lee, Eun Whan(2012) proposed a data mining application in customer relationship management (CRM) for hospital inpatients and modeled

ISSN: 1975-0080 IJMUE

the patterns of the loyal customers' medical services usage via a decision tree [2]. Wei, Jo-Ting etc. (2013) combined self-organizing maps (SOM) and K-means methods to apply in RFM (recency, frequency, and monetary) model for a hair salon in Taiwan to segment customers and develop marketing strategies and identified four types of customers in this case, including loyal customers, potential customers, new customers and lost customers and developed unique marketing techniques help identify four types of customers in this case [3]. Xiao Shengling (2011) uses the clustering algorithm to study the supermarket's customer group, to propose the loyalty factor indicators to measure customer's loyalty, to build loyalty and profitable customer segmentation model, and to help corporate to accurately identify different types of customer group [4]. Wang Wenxian (2012) builds the multi-level personal customer loyalty measurement indicator and evaluation model through RFM model scoring method to classify the customers into four groups: friends, barnacle, butterfly and stranger. The result shows that the personal customer loyalty evaluation model can effectively distinguish the level of different customer's loyalty, revealing the distribution of customer's loyalty [5]. Jia Ruiyue (2013) uses the two-stage mode of factor analysis and logistic regression to analyze the factors such as service quality and price affecting the customer loyalty of Chinese commercial banks, which shows that the customers with bank conversion experience often have higher loyalty, for the bank conversion is the process for customers to optimize the services and prices, and is the initiative improvement process of their satisfaction [6].

The above research results indicate that the study of customer loyalty at home and abroad becomes mature and also utilizes the research results to practical business management, but from which we can see that currently there are still many deficiencies, for example, the research on the customer loyalty in the security company focuses on the classification and evaluation of customer loyalty. From the business perspective, if you can use a more accurate model to predict customer loyalty, it will save a lot of costs for the corporate. Therefore, it is necessary to establish customer loyalty prediction model for the security industry, so that to determine the customer loyalty as well as take effective customer management measures for the corporate.

2. Background Knowledge

2.1. Customer Loyalty

Kandampully (1998) believes that customer loyalty is the commitment and guarantee of the corporate to customers on the service quality. The commitment of corporate can be transmitted through service personnel, and establishing long-term good relationship with customers will be able to win the loyalty and trust of customers. Selnes (1993) thinks that customer loyalty is the intention of repurchasing a product or service purchased by the customers before, which is the possibility scale of the future customer to repurchase the product or service. Day (1969), Niek & Basu (1994) point out that only defining the customer loyalty from the perspective of the behavior is not comprehensive and it should take customer's attitude trend into account. Dick and Basu (1994) think that real customer loyalty should be the repeat purchase behavior with a clear attitude.

Customer loyalty is formed on the basis of customers' past experience. Customer loyalty contains both emotional attitude trend and the trend consciously and behaviorally. For the corporate, customer loyalty not only needs the some aspects of emotional dependency, behaviorally loyalty or conscious loyalty, but also needs customers to show loyalty to the corporate on the above three aspects, so that it is able to bring high-profit customers to the corporate, and such customers are the target customers the corporate really need, and are the real loyal customers. Specific features are: the repurchase intention of a customer towards a product or service purchased before (repeat purchase behavior); they are not only willing to buy a product or a service, but also willing to buy multiple products or services (cross purchase); they are familiar with company's products or services, so they will often recommend the company's products or services to others according to their preferences; they have a certain immunity of wooing and temptation from their competitors and are able to endure the company's occasional mistakes and still remain their loyalty.

2.2. Data Mining

The data mining technology is concerned with the discovery and extraction of latent knowledge from a database (Chang *et al.*, 2001). Many algorithms are developed, proposed and applied: the decision tree, clustering, sequence clustering, association rule, Naïve Bayes, regression, neural network, *etc*. [7]. These techniques have become more popular and been frequently used in real-world applications, clustering and decision tree are selected to further explain data mining [8-9].

2.3. Decision Tree and C5.0

Decision tree is popular and powerful for both classification and prediction [7]. A decision tree is a classifier which conducts recursive partition over the instance space[10-11]. The attractiveness of tree-based methods is due largely to the fact that decision tree represent rules (Berry and Linoff, 2004). A decision tree is based on the methodology of tree graphs and can be considered one of the more simple inductive study methods (Quinlan, 1986, 1993; Russell and Norving, 1995). Even if the user lacks any statistical knowledge, he or she can use a decision tree to analyze specific behavior and it can be converted into rules easily. However, if it becomes too complicated or too huge for decision-making, trimming some of its leaves or branches may become necessary in order to improve its effectiveness. Of all the calculative methods, ID3, C4.5 (Quinlan, 1993; Cheng *et al.*, 1998), CART (Breiman *et al.*, 1984) and CHAID (Magidson and Vermunt, 2004) are the most well known.

C5.0 [12] is a standard algorithm for inducing classification rules in the form of decision tree. As the extension of ID3 [13-14], the default criteria of choosing splitting attributes in C5.0 is information gain ratio. Instead of using information gain as that in ID3, information gain ratio avoids the bias of selecting attributes with many values.

3. Modeling Process

3.1. Business Understanding

The loyalty forecast model is established through the existing basic customer information and transaction data to forecast customer loyalty of the enterprise and provide the enterprise with loyalty results of various levels, and effective maintenance measures are taken according to the loyalty degree to improve the customer loyalty in a targeted way, so that the customers can further use the enterprise's products or services.

3.2. Data Understanding

The original data are from 14783 pieces of information of securities companies in some quarter in 2011, including 57 fields. Through statistical analysis, the transaction data of each account opening year in some quarter in 2011, is as shown in Table 1.

Table 1. Data Quantity of Each Account Opening Year in Some Quarter between 1997 to 2011

Opening Year	1997	1998	1999	2006	2009	2010	2011
Data Quantity	5300	2424	1889	13	200	2434	2523

It can be seen from the table that the transaction data in each account opening year are not balanced, in which the data quantity in 1997 was up to 5300, in 2011 was up to 2500 too, in 2006 was only 13, in 2009 was only 200. There is a great difference in the data quantity of each year, such a large gap might have an adverse effect on the forecast of the customer loyalty model.

Before pre-processing the data, it is required to be familiar with the data, identify their quality problems, describe the data, generate data property report and understand the significance or calculation formula of each field in the original data. Important fields includes the account opening date, Customer status, the beginning market value, closing market value, the total commission, total amount of transactions, the average turnover, ending total assets, accumulate assets, amount of profit or loss, profit or loss rate, number of transactions, beginning total assets, ending total assets, turnover rate, total amount of commission, number of days of commission, times of commission and average traction amount.

3.3. Data Preparation

The main work in this stage is to screen the data available to research the customer loyalty from the obtained data, and analyze and research the properties of these data, find out the key factors influencing the customer loyalty, and process the value of these properties, so that they will meet the modeling requirements of the algorithm.

3.3.1. Data Selecting: When the customer state value is account cancelation, it means that this customer has been lost and is valueless to the forecast, so the record with customer state of account cancelation is selected to be abandoned. Meanwhile, all data without transaction information or lack of transaction information are abandoned from the database, because these data are valueless to research the customer loyalty. At this time, the decrease of data volume is 6345 records.

3.3.2. Data Cleaning: One of the important works in the data mining is the data preprocessing. To mine the data more conveniently and effectively, it is required to provide accurate, clean, standard and tidy data. But in the actual data collection, many data are not those we want and generally there will be noise, redundancy, data inconsistency and data loss etc. The reasons for these situations might be the negligence in artificial entry or deletion by mistake or the failure of machine and equipment. There still exist dirty data and noisy data in the data initially processed by the above means. If these data are directly used without being preprocessed, the model establishment and model forecast accuracy will be influenced, so it is necessary to clean these data.

By auditing the original data, it can be found that many properties in the original data are unnecessary, because the property values of these fields are basically missing, and are not directly significant to the loyalty analysis. From the perspective of data mining, a lot of data missing will directly influence the model analysis result. Therefore, the properties without actual significance to the problem analysis can be directly eliminated. In this article, the fields with effective record less than and equal to 362 are filtered, and 33 properties are left are filtration.

By observing the original data, it is found that some customers have one, two or three transactions of onsite entrusting, online transaction entrusting and telephone entrusting. For a securities company, any entrusting form of the customer can be accepted. Generally, the more the entrusting frequency and entrusted amount are, the more the transactions facilitated and the more the commission obtained by the security company will be. Now the onsite entrusting amount, telephone entrusting amount and online transaction entrusting amount are summarized and the three kinds of entrusting amounts are added to get the total entrusting amount.

When the customer state is normal, the entrusting frequency, transaction frequency, order cancelation frequency and customer data are selected as the analysis factors, and scatter diagram is used to analyze their internal relation, and it is found that the entrusting frequency and transaction frequency are of strong linear correlation, as shown in Figure 1.

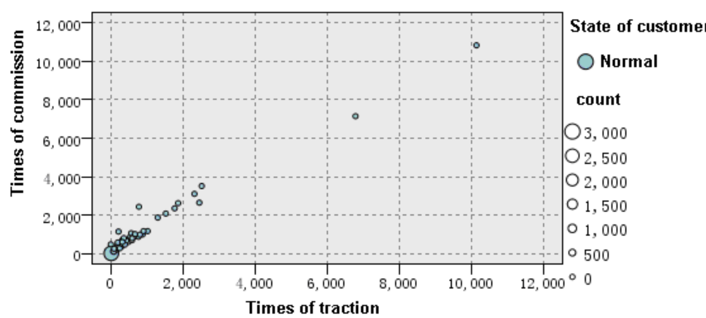


Figure1. Strong Linear Correlation of the Entrusting Frequency and Transaction Frequency

Similarly, the entrusting frequency and order cancelation frequency and transaction frequency are processed in the same way, and it is found that there is a strong linear correlation. There exists a strong linear relation among the three, so two fields can be eliminated and one reserved. There are only 2088 effective values in the order cancelation frequency, 3355 effective value for the transaction frequency and 3421 for the entrusting frequency. In consideration of small standard deviation for the transaction frequency and that transaction means that commission can be brought to the securities company, so for convenience of modeling, the field transaction frequency is selected.

3.3.3. Data Conversion: Data conversion means to convert the data or unified into a format available for mining. Here, we divide the transaction frequency into four levels, A, B, C and D, in which Level A represents that the transaction frequency is more than or equal to 0 and less than or equal to 3. Level B represents that the transaction frequency is more than or equal to 4 and less than or equal to 50. Level C represents that the transaction frequency is more than or equal to 51 and less than or equal to 150. Level D represents that the transaction frequency is more than or equal to 151.

The total entrusting amount is also divided into three levels (low, medium and high) in the same manner, in which the low level represents that the total entrusting amount is less than or equal to 10000, the medium level represents that the total entrusting amount is more than or equal to 10000 and less than or equal to 4000000, the high level represents that the total entrusting amount is more than 4000000.

There must be an output field for the modeling, because no existing fields can be used as the output field on behalf of the loyalty, a new field must be derived from other fields, here the divided transaction frequency and total entrusting amount are taken as the fields to weigh the loyalty of securities customers. The customer loyalty is reflected in behavioral loyalty and attitude loyalty of the customers, the transaction frequency indicates that customers purchase services in a securities company, the more frequent the transaction is, the more loyal the customer behavior will be; the total entrusting amount represents that the customers rely on the securities companies and are willing to be entrusted to them and reflects the customers' attitude loyalty.

The transaction frequency has four levels, and total entrusting amount has three levels, so twelve results will be generated. We divide the customer loyalty into four states, respectively zero loyalty, invisible loyalty, inertial loyalty and high loyalty, as shown in Figure 2.

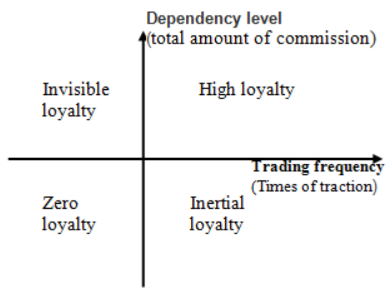


Figure 2. The Loyalty Dimension

Zero loyalty means that some customers have few transactions and little rely on the company's products or services, so they brining little commission to the broker. These customers might want to have a try, or they dare not take a great risk because they have just contacted the securities investment.

Inertial loyalty means that some customers have a high transaction frequency, but they rely little on the products and services. They transact frequently because of inertia and they are satisfied to the products or services to some extent, but do not concern the products or services of the competitors temporarily.

Invisible loyalty means that some customers have a low transaction frequency, but actually they rely on the products or services greatly, just they are limited by some factors, for example, the influence of their friends.

High loyalty means that customers do not only rely on the securities companies' products or services to a large extent, but also have very frequent transactions.

Condition for high loyalty is (level of transaction frequency='C' and division level of total entrusting amount ='high') or (level of transaction frequency='D' and division level of total entrusting amount ='high'). Condition for Zero loyalty is (level of transaction frequency='A' and division level of total entrusting amount ='low') or (level of transaction frequency='B' and division level of total entrusting amount ='low'). Condition for Inertial loyalty is (level of transaction frequency='A' and division level of total entrusting amount ='medium') or (level of transaction frequency='B' and division level of total entrusting amount ='medium') or (level of transaction frequency='A' and division level of total entrusting amount ='high') or (level of transaction frequency='B' and division level of total entrusting amount =' high '). The conditions for the invisible loyalty are the remaining situations.

The whole process for the data preprocessing is as shown in the Figure 3 and Figure 4.

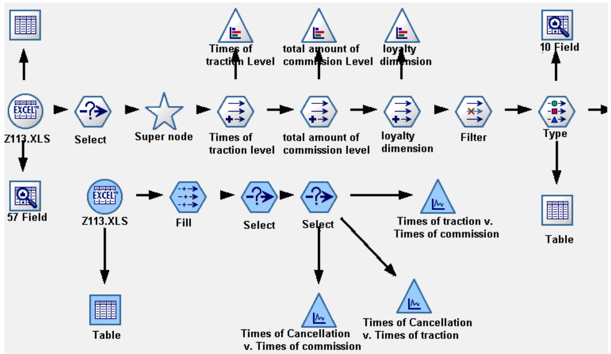


Figure 3. The Whole Process for the Data Preprocessing

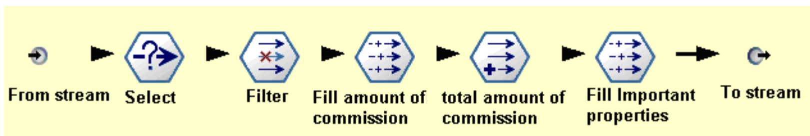


Figure 4. Super Node

3.4. Prediction Model of Customer Loyalty

We divide the preprocessed data set (6435records in total) into two parts: training set and test set, respectively accounting for 50% of the total data.

First, the training set is selected for modeling, the C5.0 rule set and decision-making tree were used to forecast the lost customers. In the forecast result of the rule set model, each node is a rule consisting of the values of various variables or the value ranges, and each data can be distributed to leaf node according to the value of different forecast variables. The input fields are the total Commission, the average turnover, total assets at the end, accumulated assets, accumulate assets、profit or loss rate、times of commission、number of days of commission and turnover rate，The input field is the loyalty dimension. Here the loyalty dimension is taken as the output field, while the loyalty dimension is determined through the level of transaction volume and the level of total entrusting amount, and the level of transaction times and level of total entrusting amount are classified through the transaction times and total entrusting amount, so in the process of modeling, it is no need for the transaction times, total entrusting amount, level of transaction times and level of total entrusting amount to participate in the modeling.

3.4.1. Interactive Verification: the sample data are divided into 10 equal data randomly on average, 10 models are established respectively, and the error is the average value of 10 models in the remaining 1/10 samples. The forecast result is the multi-model voting result. The 10 models in the decision-making tree model has an average forecast precision of 94.7, standard deviation of the forecast precision is 0.3 and there 3227 training data participating in the modeling, with an analysis accuracy up to 95.507%.

3.4.2. C5.0 Data Stream: The training set and test set is separated, C5.0 model data stream is generated, as shown in Figure 5.

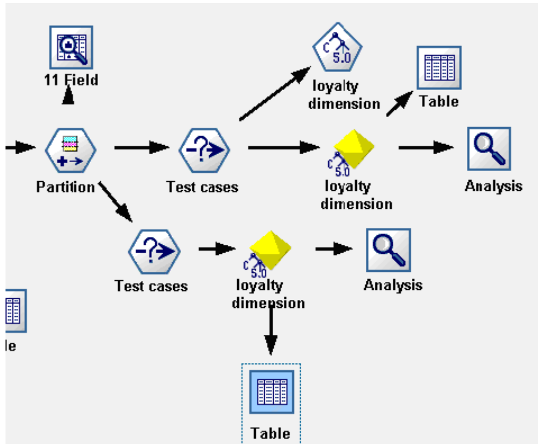


Figure 5. Modeling Process

3.4.3. Results: The analysis result of C5.0 decision-making tree is displayed in the graphic form of texts and decision-making trees, and the detailed results are explained as follows.

Class I zero loyalty: if the total commission is less than or equal to 15.860 (1784 samples), and the average transaction amount is less than or equal to 6203(1724 samples), then the customers are zero-loyalty customers, with confidence coefficient of 0.999. If the total commission is less than or equal to 15.860(1784 samples), and the average transaction amount is more than 6203(60 samples) and less than or equal to 10332.809(38 samples), then the customers are zero-loyalty customers, with confidence coefficient of 0.711.

Class 2 inertial loyalty: if the total commission is less than or equal to 15.860 (1784 samples), and the average transaction amount is more than 10332.809(22 samples), then the customers are inertial customers, with confidence coefficient of 1.0.

If the total commission is more than 15.860 (1443 samples), and the average transaction amount is less than or equal to 61(1074 samples), then the customers are inertial-loyalty customers, with confidence coefficient of 0.929. If the total commission is more than 15.860 (1443 samples) and less than 249.320(50 samples), and the average transaction amount is more than 61(369samples), then the customers are inertial-loyalty customers, with confidence coefficient of 0.62.

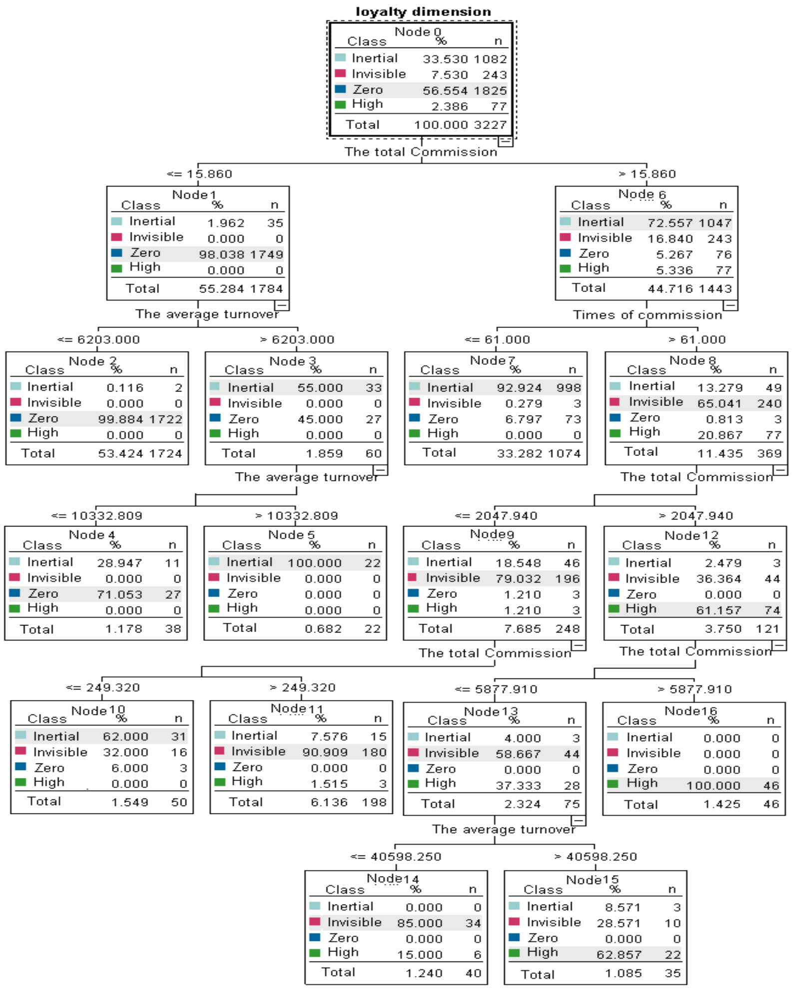


Figure 6. The Decision Tree Mode

Class 3 invisible loyalty: If the total commission is more than 249.320 (198 samples) and less than 2047.940(248 samples), then the customers are invisible -loyalty customers, with confidence coefficient of 0.909. If the total commission is more than 249.320 (198 samples) and less than 5877.910(75 samples), and the average transaction amount is less than or equal to 40598.250(40 samples), then the customers are invisible -loyalty customers, with confidence coefficient of 0.85.

Class 4 high loyalty: If the total commission is more than 2047.940(121 samples) and less than 5877.910(75 samples) and the average transaction amount is more than 40598.250(35 samples), then the customers are high -loyalty customers, with confidence coefficient of 0.629.

It can be seen from the above analysis result that most of the forecasts have a relatively high confidence coefficient.

The decision tree model is as shown in Figure 6.

4. Evaluation of the Model

To judge whether the model mined with data mining technology are reasonable and accurate and can adapt to the actual demand of the enterprises, it is surely needed to evaluate the model. Furthermore, in the process of C5.0 model establishment, the training set is based from modeling to trimming, and the evaluation accuracy on the test set directly influence the modeling success. So it is needed to use the partitioned data set to evaluate and analyze the model established.

4.1. Analysis on Model Test Result

The fields $C-loyalty dimension and $CC-loyalty dimension are newly added in the table, respectively representing the forecast value and forecast confidence coefficient of each record, and the records of some fields and data are selected, is as shown in Table 2.

Table 2. Predicted Results of Training Sets

NO	Col1	Col2	Col3	...	Col(n-2)	Col(n-1)	Col(n)
1	38322.23	148034.9	6	...	high loyalty	high loyalty	0.94
2	28983.89	12517.87	65	...	high loyalty	high loyalty	0.94
3	8645.34	14347.78	52	...	high loyalty	high loyalty	0.94
4	8188.38	41099.38	48	...	high loyalty	high loyalty	0.94
5	10	1504.5	1	...	zero loyalty	zero loyalty	0.997
6	10	2381.5	1	...	zero loyalty	zero loyalty	0.997
7	7.9	9880	1	...	zero loyalty	zero loyalty	0.667
⋮	⋮	⋮	⋮	...	⋮	⋮	⋮
3221	36.23	5009.143	4	...	inertial loyalty	inertial loyalty	0.927
3222	35.64	5337	4	...	inertial loyalty	inertial loyalty	0.927
3223	28.86	18036.4	1	...	inertial loyalty	inertial loyalty	0.927
3224	2187.79	54695.26	22	...	invisible loyalty	High loyalty	0.589
3225	4058.75	27898.67	32	...	high loyalty	invisible loyalty	0.795
3226	308.35	10146.16	25	...	invisible loyalty	invisible loyalty	0.896

*Column name means:Col1 means the total commission , Col2 means the average transaction amount, Col3 means Number of days of Commission, Col(n-2) means Original loyalty dimension, Col(n-1) means Predicted $C-loyalty dimension Col(n) means Confidence of $CC-loyalty dimension.

It can be seen from the forecast result that the confidence coefficient is relatively high generally, but there is also deviation in a few forecast results, for example, the customer No.3224 is of invisible loyalty, but the forecast result is high loyalty, with confidence coefficient of only 0. 589744; the customer No.3225 is of high loyalty, but the forecast result is invisible loyalty, with confidence coefficient of 0. 795455.

In comparison of data of customer No.1 with those of customer No.2, the forecast is high loyalty, with a large transaction volume each time and relatively large difference of entrusting days, and these customers will bring more total commissions to the company, indicating that even for customers of the same level, there will be different values for the securities companies.

In comparison of the data of No.2 and No.3226, although the average transaction volume of the two is more or less, No.2 is a high-loyalty customer, while No. 3226 is invisible-loyalty customer and the customer No.2 contributes total commission to the securities companies far more than customer No. 3226. Although the average transaction volume of the two customers is more or less, the customers with high loyalty to the companies are more valuable than those with low loyalty, and the customers of higher loyalty will bring more profits to the securities companies. Therefore, for the securities companies, it is required to strengthen the management and maintenance of high-loyalty customers and give them more encouragement, so as to make them maintain a high loyalty. The companies should take individual marketing to draw the attention of customers, enhance their reliance on the securities companies and strive to transform them into high-loyalty customers.

4.2. The Assessment Results Result

The accuracy and error rate of the forecast result of training set and test set samples are counted respectively, and as an evaluation result, the comparative analysis intuitively reflects the accuracy and predictability of the model, is as shown in Table 3 and Table 4.

Table 3. Partitioned Statistics (training data)

State	Number of Customer	percentage
correct	3082	95.51%
error	145	4.49%
Total	3227	100%

For example, there are 3208 samples in the test sample set, in which 3030 samples are forecasted correctly and 178 wrongly, with an accuracy of 94.45% and error rate 5.55%, indicating that this model has a higher accuracy in forecast of the customer loyalty, and can be used to forecast the loyalty and loyalty degree of the customers within a certain period in the future.

Table 4. Partitioned Statistics (Testing Data)

State	Number of Customer	percentage
correct	3030	94.45%
error	178	5.55%
Total	3208	100%

Table 5 is the coincidence matrix of $C-customer state, the line in the table represents the actual value, column represents the forecast value, and the content in each cell is the record number of the model. The variable actually is the customers of inertial loyalty, and there are 1034 records in which the customers forecasted to be inertial customers. For the inertial customers actually, there are 18 records in which the customers are forecasted to be invisible-loyalty customers, 20 records in which the customers are forecasted to be of zero loyalty and 3 high loyalty, indicating that the model has a relatively high forecast accuracy.

Table 5. The Coincidence Matrix of $C- the Loyalty Dimension

Testing	inertial loyalty	invisible loyalty	zero loyalty	high loyalty
inertial loyalty	1051	15	13	3
invisible loyalty	19	214	0	10
zero loyalty	76	0	1749	0
high loyalty	0	9	0	68

The confidence coefficient value of the $CC-loyalty shows the evaluation result of the forecast confidence coefficient o the forecast test set, and the result reflects that this model is tested to have a higher accurate in the training set. In the repeated test, the confidence coefficient of the model is also relatively high, and the forecast ability is relatively excellent, so this model is relatively reliable.

According to the above analysis, this model has good forecast ability, and the model is forecasted through deployment and according to the analysis process in the rule set and decision-making tree. Classification results are as shown in Table 7.

Table 6. Classification Table of Customer Loyalty

classification	zero loyalty	inertial loyalty	invisible loyalty	high loyalty
sample size	1762	1146	238	81
percentage	54.6%	35.5%	7.4%	2.5%

In the first class of zero loyalty customers, the sample size is 1762, and the Class IV customers are the most, accounting for 54.6% of the total number, but such customers have a relatively small transaction frequency, with less transaction volume, they bring very few commissions to the company, and are not so significant for the customers, because although they have a great number, they do not understand the securities industry too much, have less experience, infrequent transaction, bring relatively few profits to the companies and the time and cost consumed to them will account for a large proportion. The companies should consider them generally, but are unnecessary to spend too much time and cost on them.

The second class of inertial loyalty customers, the sample size is 1146, accounting for 35.5% of the total number, have a great number and have very frequent transaction, but the total entrusting amount, entrusting frequency and das are fewer, indicating that the do not have reliance on the customers and there is a large gap in the commissions they bring to the companies. The transaction is frequent because of inertia. Such customers are easiest to be strived for by the competitors, so the companies should do a lot of work to retain them, for example, increase the differentiation of products or services, and for the individual investors, establish different individual financing plans for the customers to select; the companies should actively investigate and research the enterprises, and for different enterprises, pack the stock transaction, acquisition of new stocks and reacquisition of national debts according to different fund scales and investment concepts, establish feasible investment analysis report for tem, and then introduce them to select, so that they will feel that the companies serve them meticulously, so as to increase their reliance on the companies.

The third class of invisible loyalty customers, the sample size is 238, accounting for 7.4%, are fewer in number, although they do not have a frequent transaction, they have frequent entrusting, with large transaction volume each time, have strong reliance on the companies and bring very great commissions to the companies. They do not have a frequent transaction because they might be influenced by some environmental factors, for these customers, the companies should find out the factors influencing the customers' transaction frequency, for example, in any business outlet of the companies, any problem of the customers can be solved professionally, whether by the employees of the outlet or by other approaches. Of course, it requires to have a professional training on all employees of the companies, refine the business and promote the marketing knowledge and effectively solve the customers' problems. The companies should improve the customers' satisfaction, increase their transaction frequency and make efforts to transform such customers into high-loyalty customers.

The fourth class of zero loyalty customers, the sample size is 81, accounting for 2.5%, are extremely few in number, but they have more frequent entrusting, with a large transaction amount each time, and they bring the most commissions to the companies, being the main source of the companies' profit, with a great value. The companies should fully concern and award such customer, for example, the brokers should establish detailed customer visiting plans according to the customer information selected, organize professional teams to for and to provide continuous consultation services according to the specific conditions of different customers, so as to maintain their loyalty.

5. Conclusions

Based on existing customer transaction data related to the securities industry and in combination with multi-industry data mining standard procedure process model and thought of flow visualization, an effective judgment is made in this subject onto customer loyalty by establishing a decision-making tree prediction model of customer loyalty. Moreover, timely adjustments are made for enterprises according to market changes. Relevant measures are taken according to customer loyalty, making the enterprises to maintain customer relations more easily, so as to save considerable costs and obtain profits as well.

As one of important means used by securities traders to win customers, the security customer loyalty prediction model is also a relatively complicated model. This model is established based on a comprehensive understanding of the securities industry and basic knowledge about data mining. With continuous development of the securities industry and data mining technologies, new requirements will be also proposed for the prediction model. Therefore, it is necessary for us to continuously explore for new demands and develop new mining methods, making the prediction model more precise.

References

[1] B. Wouter and V. D. P. Dirk, "Customer base analysis: Partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting", European Journal of Operational Research, vol. 164, no.1, (**2005**), pp. 252-268.

[2] L. E. Whan, "Data mining application in customer relationship management for hospital inpatients", Healthcare Informatics Research, vol. 18, no. 3, (**2012**), pp. 178-85.

[3] J.-T. Wei, M.-C. Lee and H.-K. Chen, "Customer relationship management in the hairdressing industry: An application of data mining techniques", Expert Systems with Applications, vol. 40, no. 18, (**2013**), pp. 7513-7518.

[4] S.-L. Xiao, L.-N. Mu, W. Wang and X.-H. Gao, "Study on CustomerscGroup of Supermarket Based on Data Mining", Resource Development & Market, vol. 27, no. 08, (**2011**), pp. 683-712.

[5] W.-X. Wang, Y. Jin, Chen and D.-b., "A Study of Personal Customer Loyalty Based on RFM Model", Financial Forum, vol. 195, no. 3, (**2012**), pp.75-80.

[6] R.-Y. Jia and S. Yang, "Service Quality, Service Price and the Customer's Loyalty of Commercial Bank", Financial Forum, vol. 209, no. 5, (**2013**), pp. 60-65.

[7] Q. Kong and X. Liang, "The research on structural path between relationship benefits and relationship outcomes in customer loyalty programs: Using data mining", Advanced Materials and Engineering Materials, vol. 457-458, (**2012**), pp. 1118-1121.

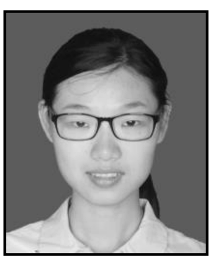
[8] H. I. Witten and E. Frank, "Data Mining: Practical machine learning tools and techniques", Morgan Kaufmann, (**2005**).

[9] M. J. Berry and G. S. Linoff, "Data mining techniques: For marketing, sales, and customer support", John Wiley & Sons, Inc., (**1997**).
[10] D. Wei and J. Wei, "A mapreduce implementation of C4.5 decision tree algorithm", International Journal of Database Theory and Application, vol. 7, no. 1, (**2013**), pp. 49-60.
[11] N. Kerdprasop and K. Kerdprasop, "A Robust Tree Induction Method Based on Heuristics and Cluster Analysis", International Journal of Database Theory and Application, vol. 5, no. 2, (**2012**), pp.17-34.
[12] J. R. Quinlan, "C4.5: programs for machine learning", Morgan Kaufmann, (**1993**).
[13] J. R. Quinlan, "Improved use of continuous attributes in C4.5", arXiv preprint cs/9603103, (**1996**).
[14] J. R. Quinlan, "Induction of decision trees", Machine Learning, vol. 1, no. 1, (**1986**), pp. 81-106.

Authors

Yihua Zhang is an associate professor at the School of Business Administration, Jimei University. He received his bachelor's degree in Economic Information Management (1994) from Southwestern University of Finance and Economics and M.S.E. in Software Engineering (2007) from Huazhong University of Science and Technology. His current research interest includes Data Mining, Big Data, System Engineering and Business Intelligence.

Yuan Wang, She received her M.S.E. in System Engineering (2005) and PhD in System Engineering (2013) from Xiamen University. Now she is full researcher of School of Business Administration, Jimei University. Her current research interest includes Data Mining, System Engineering, Business Intelligence and Theory and Technology of Decision-Making.



Chunfang He, She is an undergraduate student who major in E-commerce at Jimei University. Her current research interest includes Data Mining, Data Warehouse, System Modeling and Business Intelligence.

TingTing Yang, She is an undergraduate student who major in E-commerce at Jimei University. Her current research interest includes Data Mining, System Analysis and Design, Software Engineering and Network Programming.

