

A Machine Learning Based Method for Customer Behavior Prediction

Jing LI, Shuxiao PAN, Lei HUANG, Xin ZHU

Abstract: Under the data-driven environment, market competition is increasingly fierce. Enterprises begin to pay attention to precise marketing to make costs down, improve marketing efficiency and competitiveness. E-mail marketing is widely used in enterprises due to its advantages of low cost and wide audience. This paper uses machine-learning techniques such as decision tree, cluster analysis and Naive Bayes algorithm to analyze customer characteristics and attributes with historical purchase records, and further analyzes the key factors that affect potential customers' purchase behavior by selecting models with high promotion degree through promotion graph, to realize accurate marketing. The results show that the prediction effect of decision tree is better than clustering analysis and Naive Bayesian algorithm, and has a higher promotion degree. The customers who are 45-55 years old and commute 1-2 kilometers away are more likely to make purchases if they do not have a car or have a car at home.

Keywords: cluster analysis; decision tree; lifting chart; Naive Bayes

1 INTRODUCTION

With the arrival of the area of "Internet plus" and the explosive growth of all kinds of data, the age of big data has come. The advent of the era of big data makes the development of any business more dependent on data. The production data and operation data of enterprises, and related data of customers all have the vital impact on the development of enterprises. In addition, the data mining is the core competitive ability for the development of enterprises. Nevertheless, the traditional techniques of data mining are long-time spent and inaccuracy, and many data are not mined and utilized effectively. With the continuous development of database and data mining technologies, the needs of data mining in many enterprises are met better [1].

In foreign countries, the research of the application of data mining in retail industry was carried out early. The data mining in business is mainly focused on customers and commodities. There into, most of them are applied for sales forecasting, inventory demand, retail point selection and price analysis. The analysis of shopping baskets in Wal-Mart is a classical case praised by industry and business, especially the beer and diapers of Wal-Mart.

At present, many researchers have applied data mining technology into the data analysis of electronic commerce except traditional methods of mathematical and statistical theories. Liu Weixiao [2] proposed a kind of hybrid intelligent prediction algorithms combining artificial neural network (ANN) and discrete grey prediction model (DGM(1,1)). He obtained influencing variables with high correlation degree by correlation degree analysis. And the idea of quadratic residuals was introduced after the prediction using the combination of DGM(1,1) and ANN. Furthermore, the residuals of actual sales data and the prediction results of the combination of DGM(1,1) and ANN were added to the influence variables and the second residuals prediction was made by ANN. Finally, the feasibility and accuracy of the algorithm prediction were verified by real fashion sales data. In addition, Wang Jianwei [3] proposed a product reclassification prediction model based on sales data is presented. Product cluster was extracted according to the commonness of product sales, and the prediction result was obtained by time series model. At the same time, the probability distribution of prediction results was given by using hidden Markov

prediction model. Besides, the marketing strategy of timely adjusting shelf position and dynamic predicting commodity sales trend was proposed by Zhang Qing et al. [4]. In their work, a model of data management, analysis and decision-making was established. Moreover, the model could be solved using data mining algorithm of Apriori and Vague set. Eventually, a model of data management, analysis and decision-making was established based on the model. What's more, on the basis of data mining technology, Zhou Shang et al. [5] discovered the relationship between customers' purchasing commodities from massive supermarket transaction data and constructed a multi-objective commodity pricing model considering sales profit and sales volume. As a result, the reasonable pricing of commodities was acquired by using artificial intelligence algorithm, so as to provide reasonable pricing scheme for supermarket management.

Additionally, data mining has been applied in medical, power system, education, logistics and other fields to varying degrees, and some progress has been made. Zhou Yunhui et al. [6] focused on the information mining of breast cancer therapeutic data in medical field and achieved the mining of treatment data by Bayesian network algorithms (Bayes Net) on the WEKA data mining platform. In the statistical analysis of the cost of hematological diseases developed by Lv Feng et al. [7], the number and treatment costs of different blood diseases were achieved by k-means algorithm optimized by genetic algorithm, so as to realizing the goal of reducing treatment cost. Xu Jun et al. [8] predict electricity by the algorithms of time series, multivariate linear regression and grey prediction etc. They enriched the means of electricity forecasting and improved the ability of short, medium and long-term electricity forecasting. Besides, they also visualized the forecasting results to provide reliable data support for electricity forecasting. Xu Yuan et al. [9] designed an improved fuzzy K-means clustering based on MapReduce parallel programming model, and a new method of medium and long term load forecasting was proposed on the basis of the designed K-means clustering. Furthermore, using the method of association rules in educational data mining technology, with the help of Apriori algorithm on WEKA platform, Wu Wenling et al. [10] explored the association rules between general education curriculum and basic subject curriculum. The

association rules would contribute to improve the teaching and learning effect and provide decision-making and opinions for curriculum construction in universities, and further improve the teaching quality of universities. Besides, Peng Yuqing et al. [11] used data mining technology to carry out regression analysis, cluster analysis, principal component analysis and association rules mining for a large number of existing teaching databases to extract valuable information, which would help teaching staff to arrange teaching work reasonably, strengthen the management of colleges and departments, and play a guiding role in improving the teaching performance. Guo et al. [12] used the data mining method based on ant colony algorithm to optimize the logistics distribution path, and verified the effectiveness of the algorithm, which provided the basis for decision analysis and data processing. Zhao et al. [13] proposed an improved Apriori algorithm that combines Apriori algorithm with logistics information to build a logistics decision support system, which can find the possible changes in the future logistics market. Zheng Jun et al. [14] applied clustering analysis technology to optimize the classification of goods in logistics management, then solved the problem of logistics network distribution by using data mining technology. Zhao et al. [15, 16] analyzed bi-objective problem with time windows and scheduling optimization. Zhang et al. [17][18] used weighted combination method and Fuzzy RDF Model to solve the Big Data problem. Most of the previous researches predict the accuracy from the perspective of products, but few people analyze the customer attributes and purchasing behaviors. There is still plenty of room for the sales strategy research based on product characteristics to target customers.

Previous studies mainly concentrated on the analysis and discussion of the target user's behavior with a single method, and the conclusions obtained have certain limitations. The marketing department of Adventure Works Cycles hopes to improve sales by predicting the attributes of target users and sending emails to specific customers. Therefore, the paper uses decision trees, cluster analysis and Bayesian algorithms to deeply analyze user behavior characteristics, explores the commonalities and characteristics of the attributes of these customers, and finds the algorithm model with higher degree of promotion, which is conducive to improving sales performance and company efficiency, so as to improve the scientific and effective decision-making of market departments.

2 METHODS

2.1 The Prediction Process of Purchase Behavior

For accurately predicting the characteristics of customers' purchasing behaviors from the historical data of customers who have purchased a scooter, machine learning related technology was used to analyze and predict the characteristics of customers' purchasing behaviors. Data mining tools include linear regression, time series decomposition, moving average, autoregression, exponential smoothing and gray theory. And machine learning mainly includes logistic regression, support vector machine, decision tree, neural network, Bayesian network and other methods. Data mining is a complete process, which mines previously unknown, effective and practical information

from large databases, and uses this information to make decisions or broaden knowledge. On the basis of data mining, machine learning algorithm is used to predict customer purchase behavior. The prediction process is shown in Fig. 1.

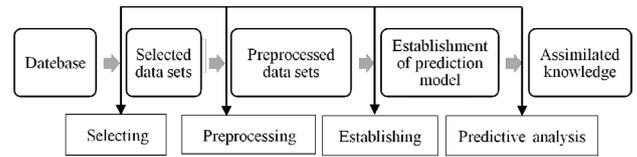


Figure 1 Prediction progress of purchase behavior of customers

2.2 Decision Trees

Decision trees are directed and acyclic tree structures which used to classify instances. A decision tree consists of a node and a directed edge. The node includes internal nodes and leaf nodes. Internal nodes are used to distinguish between different attributes or features, and leaf nodes represent different categories. For different attributes or features, leaf nodes represent different classifications. Among them, the root node has no parent node, the other nodes have and only one parent node [19], and the node without child node is called Leaf Node. Each leaf node corresponds to the value of a class identifier C , and the other internal nodes correspond to the Splitting Attribute. The core idea of decision tree is to select appropriate labels for input values, select test attributes at the decreasing speed of information entropy, process training sets with location attribute values by information gain or information gain rate, and classify unknown attribute values by estimating the probabilities of various possible results until the decision tree can train classification data effectively [20].

In information theory, entropy is a measure of uncertainty of random variables. The greater the entropy value, the higher the uncertainty of random variables, the more disorderly the data classification, and the smaller the entropy value, the better. Entropy of random variable X is defined below

$$H(X) = -\sum_{i=1}^n p_i \log p_i \quad (1)$$

Information Entropy is defined as

$$I = -\sum_{i=1}^m p_i \log_2 p_i \quad (2)$$

M means that the sample is divided into m parts. The smaller the information entropy is, the higher the information purity is, and the fewer classification categories are included. Therefore, the bigger the difference between the original information entropy and the classification effect is, the better the classification effect is.

Information Gain is a measure of the degree to which information complexity decreases under certain conditions. It is used to measure the impact of a feature on classification results

$$GAIN = Entropy(p) - \left[\sum_{i=1}^k \frac{n_i}{n} Entropy(i) \right] \quad (3)$$

In the process of building decision tree, maximizing information gain is chosen [21] as the test condition to partition the nodes. However, due to the fact that the information gain tends to take more values, the information Gain Ratio is introduced to correct the problem.

Based on information gain, information gain rate adds penalty items. Considering the number and size of branches, it is defined below

$$GainRatio(A) = \frac{Gain(A)}{Entropy(A)} \quad (4)$$

When the information gain is higher than the average level of feature [22], the feature with high information gain rate is selected.

Decision tree is a common classification and prediction algorithm in data mining. It is generated by repeatedly dividing data into homogeneous data groups. Its generalization ability is strong. It mainly includes two steps: starting from the root node, data points are divided into two groups according to similarity, and then each group is divided into two groups according to similarity, until the data points of leaf nodes are the same prediction category or further divided. The branch terminates when the homogeneity cannot be improved because it exceeds the minimum threshold. Finally, the termination criteria can be selected by cross validation. Decision tree has low requirement on data set. It can process both continuous data and categorized data. The algorithm complexity is not high, and it is easy to understand and implement.

2.3 Cluster Analysis

Cluster analysis is a process of dividing the research object into several classes based on similarity. The similarity between the same classes is high, while the similarity between the different classes is low. Clustering analysis belongs to unsupervised learning with simple logic and strong ability to process low-dimensional data. Classification mainly depends on the characteristics, nature and clustering purpose of the data itself. It aims to divide the samples in the data set into several disjoint subsets. The specific process is shown in Fig. 2.

K-Means algorithm is a classical clustering analysis method, which distributes clustering members by average distance value. Input a data set containing n objects, randomly select k objects as clusters, select the nearest cluster according to the distance between the remaining objects and the center of each cluster, and then recalculate the average value of each cluster until the function converges, so that the similarity within the cluster is high, while the similarity between clusters is low [23]. Generally, the square error criterion is used:

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (5)$$

Among them, E represents the sum of squares of errors

[24], p is the point in space, and m_i is the average value of cluster C_i .

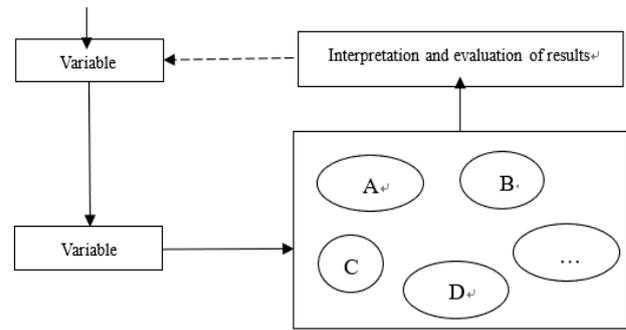


Figure 2 Processing of Cluster Analysis

2.4 Naive Bayes

Bayesian classification is a general term for a class of classification algorithms, which are based on classical Bayesian probability theory. Naive Bayes is a classification algorithm based on Bayesian theorem and independent assumption of characteristic conditions. It describes the possibility of an event on the basis of prior knowledge.

Bayesian theory expresses uncertainty by judging the probability of occurrence of one event. Based on prior knowledge and posterior knowledge, it calculates the probability of occurrence of another event according to the probability of occurrence. The theorem is expressed by formula:

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)} = \frac{P(x,c)}{P(x)} \quad (6)$$

For data set D , formula C represents a case where random events occur. X refers to the factors related to random events. $P(c|x)$ is the probability of the occurrence of case C [25] under the condition of x , which is called a posteriori probability; $P(c)$ represents the probability of occurrence of case C under the condition of x ; $P(x|c)$ denotes the probability of occurrence of case C under the condition of known event c , which is called a prior probability; similarly, $P(x|c)$ denotes the probability of occurrence of case x under the condition of known event C . Usually, x is related to many factors, and the attribute values that need to be considered can be expressed as :

$$x = (x_1, x_2, x_3, \dots, x_n) \quad (7)$$

Naive Bayes Classifier (NBC) is obtained by assuming that the possibility of each attribute taking its values is independent of each other and not related to the values of other attributes. It can be expressed as:

$$P(c|x) = \frac{P(c)P(c|x)}{P(x)} = \frac{P(c)}{P(x)} \prod_{i=1}^d P(x_i|c) \quad (8)$$

When x_i attribute is discrete, $P(x_i|c) = \frac{|D_{c,x_i}|}{|D_c|}$; If the properties of x_i are continuous, assuming $p(s_i|c) \sim$

$N(\mu_{c,i}, \sigma_{c,i}^2)$, that:

$$P(x_i | c) = \frac{1}{\sqrt{2\pi}\sigma_{c,i}} \exp\left[-\frac{(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2}\right] \quad (9)$$

Naive Bayesian algorithm has high practical value. The calculation based on prior probability effectively avoids errors caused by objective problems such as insufficient samples. It has less time and space consumption, good robustness, insensitivity to missing data and stable classification effect.

3 PREDICTION OF THE TARGET CUSTOMS

3.1 Data Preparation

Adventure Works Cycles is a large multinational manufacturer that produces metal and composite

substitutes for bicycles, which are exported to North America, Europe and Asia. This paper is based on the data of existing customers and potential new customers in Adventure Works DW database of Adventure Works Company, including the historical data of surrogate bicycle buyers and the data of expected surrogate bicycle buyers. The historical data of bicycle buyers are used to construct and test the prediction model. And select the model with good prediction effect to predict the possibility of potential new customers to purchase behavior.

The subjects of the experiment are the customers who have purchased substituted bicycles. The historical data of satisfied customers are collected and de-privacy processed. 18484 pieces of data are obtained. The specific information is shown in Tab. 1.

The basic attributes of customer commuting distance, occupation, marital status, the number of cars and the number of children in the family are shown in Tab. 2.

Table 1 Customer Information

ID	Age	Scooter Buyer	Commute Distance	Customer Key	English Occupation	Gender	Geography Key
1	51	1	1-2 Miles	11000	Professional	M	26
2	52	1	0-1 Miles	11001	Professional	M	37
3	52	1	2-5 Miles	11002	Professional	M	31
4	49	1	5-10 Miles	11003	Professional	F	11
...
18481	49	1	1-2 Miles	11004	Professional	F	19
18482	52	1	5-10 Miles	11005	Professional	M	22
18483	52	1	5-10 Miles	11006	Professional	F	8
18484	53	1	0-1 Miles	11007	Professional	M	40

Table 2 Attribute of Customer

Attribute	Commute Distance	Occupation	Marital Status	Number Cars Owned	Number Children at Home
1	0 - 1Miles	Professional	M	0	0
2	1 - 2Miles	Skilled Manua	S	1	1
3	2 - 5Miles	Management	Missing	2	2
4	5- 10Miles	Clerical	M	3	3
5	Others	Others	S	Others	Others

3.2 Data Preprocessing

Effective data discretization can reduce the complexity of the algorithm, reduce computing time and save space, and decision tree and naive Bayesian algorithm are based on discrete data. Therefore, discretization of data can improve the classification ability and anti-noise ability of the algorithm. The recorded data of customer's age and annual income are continuous, and the discretization process is shown in Tab. 3.

Table 3 Data discretization

Attribute	Age	Yearly Income
Content	Discretized	Discretized
Discretization Buket Count	4	3
Discretization Method	Clusters	Clusters
Type	Long	Double

The reasons for missing values are various, including random missing and completely non-random missing. The data missing in this paper does not depend upon any other variables and belongs to random missing. The missing values in data sets are dealt with by deleting cases.

A total of 17484 valid data were obtained by data preprocessing, of which 9004 were males and 8480 were females. The distribution of customer commuting distance was shown in Fig. 3.

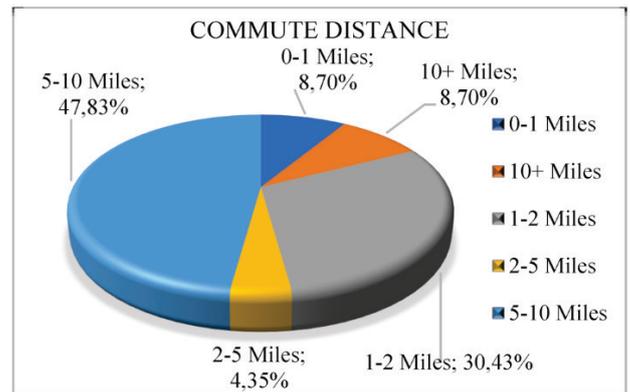


Figure 3 Commute Distance

3.3 Model Construction and Application

Mining model applies the algorithm to data creation, which can be applied to new data in order to generate a set of data, statistical information and patterns for prediction and inference of relationships. The above attributes are used to train data structure prediction model, and the tested model is used for the prediction of activity process. Define the data source and create a view of the data source, save and manage the data source in the project, and deploy it to

the Microsoft SQL Server Analysis Services database.

Data mining model usually divides data into training set and test set. Training set is used to train sample data, discover and predict the potential relationship between sample data, and test set is used to test the precision of trained model. Here, the training set is 70% and the test set is 30%.

Decision tree, clustering analysis and Naive Bayesian algorithm are used to analyze the factors affecting customer purchasing, and to explore the dependence between different factors. The specific analysis is as follows.

1) Decision Tree Analysis

Decision tree model will provide the importance and dependence of various factors affecting the purchase of surrogate bicycles. The results show that among 17484 historical customers, 8639 (49.41%) have the purchase behavior of surrogate bicycles, and 8845 (50.59%) have no purchase behavior of surrogate bicycles. Among them, customers between 44 and 52 years old and without automobiles are the most likely to buy surrogate bicycles, which is 85.92%. Further analysis shows that customers between 47 and 52 years old more tend to buy surrogate bicycles. If there are no children and the profession is a clerk, it is usually 100% sure that the customer will buy surrogate bicycles. Through the above analysis, we can see that the most important factor affecting customer buying behavior is age and the number of cars in the family.

2) Cluster analysis

Cluster analysis is grouped according to customer attributes. The classification diagram shows all the classifications in the mining model, as shown in Fig. 4. Lines between different classifications represent the magnitude of correlation between different classes, and their brightness depends on the similarity between classifications. The color of each classification represents the frequency of occurrence of variables and states in the classification.

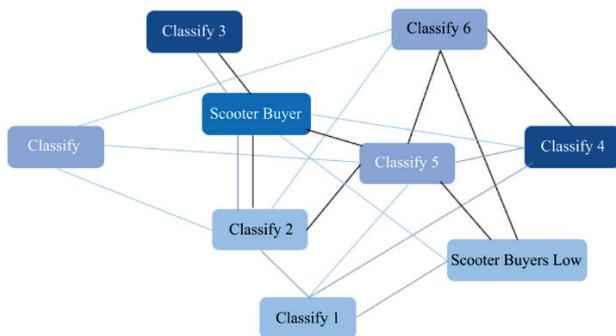


Figure 4 Classification diagram

Table 4 Attribute of the most powerful customer

Attribute	Value
Total Children	1
Number Cars Owned	1
Age	58

The results of the cluster analysis divided the historical customers into 10 categories, and the probability of purchasing the scooter with the most purchasing customer group was 68%. The characteristics of such customers are shown in Tab. 4. In most households, there is a child, a car, and the average age of 58 (36-88) customers tend to buy a

scooter.

A detailed comparison analysis of the various attributes of the different customer groups to purchase the scooter, drawing a bar chart as shown in Tab. 5.

3) Naive Bayes

The Naive Bayes shows the interaction between the purchase and input properties of the scooter. The results show that the quantity of cars and commute are the most important factors affecting customer-buying behavior followed by the number of family children, cultural level, region and marital status. Customers who have no children at home, have close commuting distances, and live in North America are more likely to purchase a scooter. People who do not have a scooter or who have a scooter are most likely to buy, and customers with more than three cars have a probability of purchasing a scooter of less than 8%.

We compare and analyze the personal attributes of customers who purchase scooters and those who do not purchase scooters. The records of consumer behavior of 8845 customers show that there are 1,460 customers without cars, 6091 customers with scooters, and those with no cars at home will buy a scooter, the probability of having a child's customer to buy a scooter is 43.92%; the probability of a customer with two cars to buy a scooter is 21.24%.

Table 5 Comparative analysis of purchasing power

Variable	Value	Buyers High	Non Buyers High
Number Cars Owned	1	[Bar]	[Bar]
Total Children	1	[Bar]	[Bar]
Number Children At Home	0	[Bar]	[Bar]
Region	Europe	[Bar]	[Bar]
Total Children	0	[Bar]	[Bar]
Number Cars Owned	2	[Bar]	[Bar]
Number Cars Owned	0	[Bar]	[Bar]

The mining model that graphically represents the mining model provides a comparison of the improvement scores for random speculation, improvement, and metrics. Fig. 6 shows an elevation map for creating a target delivery model with a target value of 1, indicating that the customer will generate purchase behavior.

As can be seen from the lift chart: the random speculation model sends the target mail to all potential customers. It could receive the response of half of the target users, which is the baseline for evaluating the lift. The peak of the ideal model is about 48%, that is, with the accuracy of the error, only need to send mail to 48% of potential customers, you can get 100% of the target customer response. When determining the target group of 48%, the actual prediction model has a degree of improvement between 60% and 75%. Decision trees, cluster analysis, and Naive Bayes algorithms all achieve higher response rates than random guess models. The response rate of decision tree is 73.43%, cluster analysis is 61.66% and Naive Bayes is 63.89%. So the decision tree model shows the biggest improvement, and the response is better than cluster analysis and Naive Bayes model.

Further, use the decision tree to further analyze the customer's scooter purchase behavior. In order to understand whether there are differences in customer purchase behaviors between different genders, the marketing department will modify the model to better meet the needs of Adventure Works Cycles target email

delivery, then choose the appropriate advertising methods and channels.

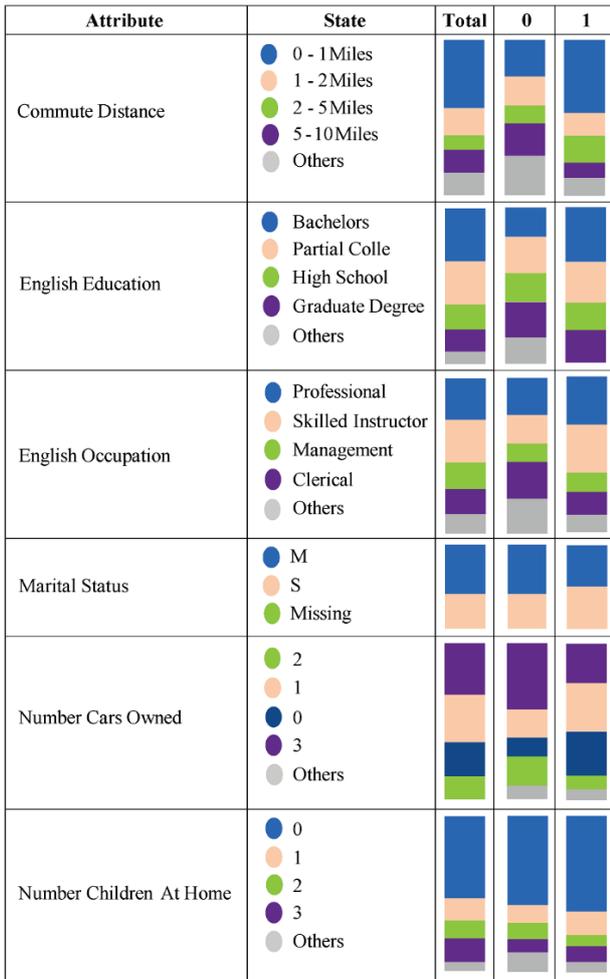


Figure 5 Different categories of relationships

The forecast results show that there are 980 customers who have a tendency to purchase. Different customers have different characteristics and different purchase possibilities. Among them, there are six potential customers who must have purchase behavior, and their age is 47 years old. In addition, those people whose age range is 45-55 years old, commuting distance is 1-2 kilometers, and do not have car at home, are more likely to purchase.

4 CONCLUSION

In this paper, data mining technology is applied to the target user prediction analysis of purchasing scooter. Cluster analysis, Decision tree analysis and Naive Bayesian algorithm are used to explore the characteristics of target groups in which purchase behavior will occur. The performance of different algorithms for customer behavior prediction and the accuracy of prediction results are compared and analyzed, so that targeted services can be provided to satisfy the individual requirement of customers and improve customer satisfaction. Breaking the customer behavior predicted by experience to traditional single method, the actual results of target user classification and prediction show that the decision tree model provides the greatest improvement, and its performance is better than cluster analysis and Naive Bayes model. The prediction result can provide useful information for the production and sales of the scooter, and provides a research idea for satisfying the individualized needs of the user, and has a bright market application prospect. Customer segmentation is mainly based on company history data, but the factors affecting the market environment are complex and diverse. This article is lacking in this aspect, which will be the focus of further research.

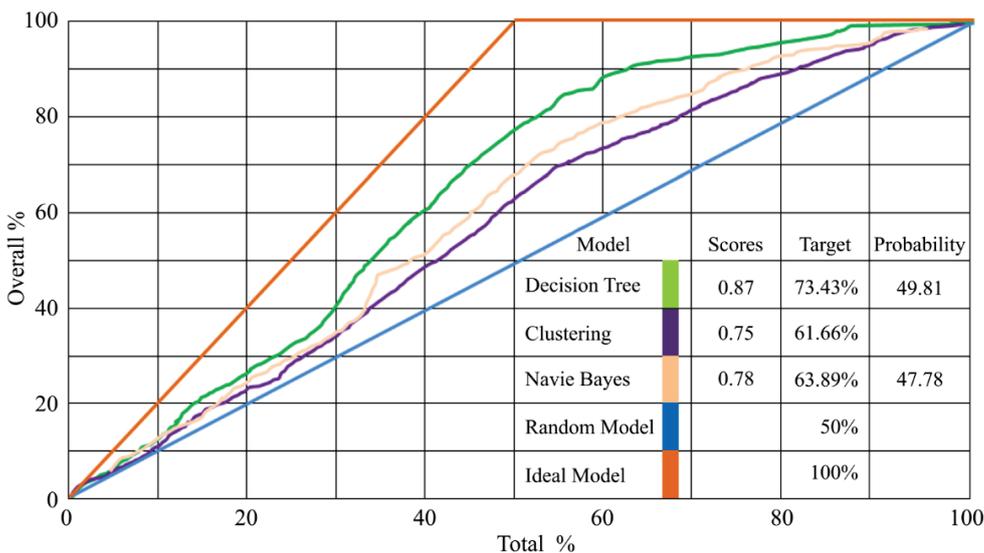


Figure 6 Lifting chart

Acknowledgments

This paper was completed under the guide of professor Zhu. The author is grateful for academic guidance and support of our group member. This study was supported by the National Science Foundation of China "Youth Fund"

project (71103014), the Beijing Municipal Philosophy Office (14JGC095), and the Science and Technology Project of the Beijing Traffic Commission (B17M00080). Thanks to my classmates and friends for their comments and suggestions. At the same time, I appreciate the scholars who gave me support and help in the references.

5 REFERENCES

- [1] Xie Qi & Zhang ZhenXing. (2013). Design of Association Rules Mining Model Based on Apriori Algorithm and OLAP. *Journal of Computer Applications*, S1, 4-5-9.
- [2] Liu Weixiao. (2016). Hybrid intelligent model for fashion sales forecasting based on discrete grey forecasting model and artificial neural network. *Journal of Computer Applications*, 36(12), 3378-3384.
- [3] Wang JianWei. Online Sales Volume Prediction Based on Items Clustering. *Computer Systems & Applications*, 25(10).
- [4] Zhang Qing, Gao GuangYin, & Jia Bo. (2016). The Application of Data Mining Technology in the Supermarket Marketing System. *Software Engineering*, 19(5), 35-38.
- [5] Zhou Shang & Huang ZhangShu. (2013). Commodity Pricing Research in Supermarket Based on Data Mining Technology. *Science Technology and Industry*, 13(2):62-64.
- [6] Zhou YunHui & Wang Jiao. (2013). Application Research on Data Mining Technology in Medical Field. *Mechanical Engineering & Automation*, 2013(4), 14-15.
- [7] Lv Feng, Yang Hong, Pu Yi, et al. (2017). Application of Genetic Algorithm Data Mining Technology in Medical Big Data. *Electronic Technology & Software Engineering*, 2017(5), 203-203.
- [8] Xu Jun, Xu WenHui, Zeng Xin, et al. (2018). Innovation and Application of Power Forecasting Method Based on Big Data Mining. *Power Systems and Big Data*, 2018(10), 203.
- [9] Xu Yuan, Cheng QianShan, Li Yang, et al. (2017). Mid-long Term Load Forecasting of Power System Based on Big Data Clustering. *Proceedings of the CSU-EPSA*, 2017(8). <https://doi.org/10.1063/1.4982406>
- [10] Wu WenLing & Gu QianYi. (2018). Application Research of Educational Data Mining Technology in College Curriculum Construction. *China Information Technology Education*, 2018(2).
- [11] Peng YuQing, Zhang HongMei, He Hua, et al. (2014). Data Mining Technology and Its Application in Teaching. *Journal of Hebei University of Science and Technology*, 22(4).
- [12] Guo JianHong & Zhang Xi. (2012). Application of ACA-based Data Mining in TPL. *Logistics Technology*, 31(19), 130-132.
- [13] Zhao ChunLing & Ning HongYun. (2007). Improvement in Apriori algorithm and using in logistics information mining. *Journal of Tianjin University of Technology*, 23(1), 30-33.
- [14] Zheng Jun, Jin Yi, Yan JiDuo, et al. (2013). Application of Data Mining Technology in Logistics Management. *Journal of Guiyang College Natural Sciences*, 8(2), 32-34.
- [15] Zhao, P. X., Luo, W. H., & Han, X. (2019). Time-dependent and bi-objective vehicle routing problem with time windows. *Advances in Production Engineering & Management*, 14(2), 201-212. <https://doi.org/10.14743/apem2019.2.322>
- [16] Zhao, P. X., Gao, W. Q., Han, X., & Luo, W. H. (2019). Bi-objective collaborative scheduling optimization of airport ferry vehicle and tractor. *International Journal of Simulation Modelling*, 18(2), 355-365. [https://doi.org/10.2507/IJSIMM18\(2\)CO9](https://doi.org/10.2507/IJSIMM18(2)CO9)
- [17] Zhang, D., Sui, J., & Gong, Y. (2017). Large scale software test data generation based on collective constraint and weighted combination method. *Tehnicki vjesnik*, 24(4), 1041-1050. <https://doi.org/10.17559/TV-20170319045945>
- [18] Zhang, D. (2017). High-speed Train Control System Big Data Analysis Based on Fuzzy RDF Model and Uncertain Reasoning. *International Journal of Computers, Communications & Control*, 12(4), 577-591. <https://doi.org/10.15837/ijccc.2017.4.2914>
- [19] Zhang Zenglian & Cao Min. (2011). Research of credit risk of commercial bank's personal loan based on CHAID decision tree. *Proceedings of the International Conference on Artificial Intelligence*. <https://doi.org/10.1109/AIMSEC.2011.6009703>
- [20] Wu Sen, Wu LingYu, Long Yu, et al. (2007). Improved Classification Algorithm by Minsup and Minconf Based on ID3. *Proceedings of the International Conference on Management Science & Engineering*. <https://doi.org/10.1109/ICMSE.2006.313896>
- [21] Mihaescu, C. M. & Burdescu, D. D. (2006). Testing Attribute Selection Algorithms for Classification Performance on Real Data. *Proceedings of the International IEEE Conference Intelligent Systems*.
- [22] Slomp, J., Krushinsky, D., & Caprihan, R. (2011). Periodic Virtual Cell Manufacturing (P-VCM) - Concept, Design and Operation. *2011 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 337-341. <https://doi.org/10.1109/IEEM.2011.6117934>
- [23] Zhang Yongqiang, Zhao Xudong, Sun Yi, et al. (2014). Waveform classification based on wavelet transform and K-means clustering. *Proceedings of the International Conference on Machine Learning & Cybernetics*. <https://doi.org/10.1109/ICMLC.2013.6890845>
- [24] Yang Shanlin, Li Yongsen, Zhang Shuchu, et al. (2005). Study on visualization of spatial data mining and knowledge discovery. *Proceedings of Spie the International Society for Optical Engineering*, 5985:59853Y-59853Y-6. <https://doi.org/10.1117/12.658218>
- [25] Tetsushi, M., Takuya, K., & Toru, F. (2010). A New method to reduce the probability of detection errors for a digital watermark using complementary decoding algorithms and minimum weight codewords of linear code. *Journal of Symbolic Logic*, 13(1), 61-64.

Contact information:

Jing LI

Department of School of Economics and Management,
Beijing Jiaotong University,
No. 3 Shangyuancun, Hai Dian District, Beijing, China, 100044
jingli@bjtu.edu.cn

Shuxiao PAN

Department of School of Economics and Management,
Beijing Jiaotong University,
No. 3 Shangyuancun, Hai Dian District, Beijing, China, 100044

Lei HUANG

Department of School of Economics and Management,
Beijing Jiaotong University,
No. 3 Shangyuancun, Hai Dian District, Beijing, China, 100044

Xin ZHU

(Corresponding author)
Management School, Beijing Union University,
No. 97 Beisihuan East Road, Chao Yang District, Beijing, China, 100101
glzhuxin@buu.edu.cn