Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 8, Number 2, pages 81–86, July 2019

Analysis of Supermarket Customer Behavior by Using RFM and Huff's Gravity Model

Yepeng Cheng
Department of Information Engineering
Graduate School of Engineering, Hiroshima University
Higashihiroshima, Japan
d185088@hiroshima-u.ac.jp

Yasuhiko Morimoto
Department of Information Engineering
Graduate School of Engineering, Hiroshima University
Higashihiroshima, Japan
morimo@hiroshima-u.ac.jp

***Abstract*—This paper discusses how to evaluate supermarket competition through distance data between addresses of customers of target supermarket chain and all supermarkets in a city and ID-POS data of target supermarket chain. We also detect the degree of influence on specific customers who are affected by the competition of other nearby supermarkets based on logistic regression analysis. In this paper, we consider how to transform the distance data positive correlation to customer affected degree to make the analysis correctly since the closer the customers' address to other supermarkets, the more likely they are influenced by these supermarkets, resulting in a decreased purchasing in target supermarket. In addition, we consider how to solve multicollinearity problems for training data. In numerical experiments, we comprehensively evaluate the influence of the other nearby supermarkets on customers purchasing of target supermarket from the viewpoint of partial regression coefficients and accuracy.**

***Index Terms*—POS, RFM, Supermarket, Huff's Gravity Model**

I. Introduction

In today's supermarket business, supermarkets are not only facing more complex market competition, but also focus on the notion of customer loyalty and profitability to increase supermarket competitiveness. Customer relationship management (CRM) is a successful strategy for capturing customer preference and demands since customer value can provide correct information to build more targeted and personalized marketing [14]. The traditional analysis methods based on CRM are focus on customers' ID-POS data for target supermarket. In generally speaking, ID-POS data mainly contains customer ID, customer daily consumption record, customer shopping date., etc. The traditional analysis method is to use ID-POS data to build quantitative model such as RFM model to conduct clustering analysis on customers and use logistic regression and other methods to investigate customers' shopping preference for various supermarket products [18]. However, few studies have explored the impact of nearby competitive supermarkets on target supermarket customers. It's an important subject to consider how to use customer ID-POS data of target supermarket to evaluate the influence of the surrounding competitive supermarkets on their customers.

That is to say, it's necessary to detect the range of trade area and the attractiveness of supermarkets. In economics, Reilly's law of retail gravitation is a heuristic developed by William J. Reilly in 1931 [1]. According to Reilly's law customers are willing to travel longer distances to larger retail centers given the higher attraction they present to customers. In Reilly's formulation, the attractiveness of the retail center becomes the analogy for size (mass) in the physical law of gravity. But there is a obvious flaw that it assumes consumers will make an alternative choice for retail stores, which is not consistent with real consumer behavior. In fact, it is possible for consumers to choose two geographically close retail stores for shopping. Reilly's law studied the trade area from the macro level, but not from the micro level for consumers. The Huff's gravity model overcame this theoretical defect later.

The Huff's gravity model was proposed by David L. Huff in 1964 [2]. It is an established theory in spatial analysis. It is based on the principle that the probability of a given consumer visiting and purchasing at a given site is a function of the distance to that site, its attractiveness, and the distance and attractiveness of competing sites. It is widely used in retail store location selection according to retail store attractiveness. It can estimate the probability that consumers will go shopping in each retail store. However, the accuracy of shopping store preference of consumers are not precise enough. Even if Nakanishi et al. [3] considered the other facters except retail store area and distance facters to improve the Huff's model, which is called the multiplicative competitive interaction (MCI) model, it still can't completely solve the problem for predicting shopping store preference of consumers accurately.

In fact, Tanaka et al. [18] proposed RFM and logistic regression analysis based approach to detect the customers' preference for each product in chain supermarkets. The RFM analysis is a traditional method for goodness of customer classification [4]. The logistic regression analysis is widely used in parametric impact analysis. The coefficients of logistic regression is mathematically considered as the parameters in Odds ratio. Since Odds ratio can reflect when some parameters varied, how's the target parameter affected. As a result, they built a good customer analysis model with a high prediction accuracy. However, they didn't consider the shopping store preference of customers, which is also called the influence of nearby supermarket competition since the closer customers to other supermarkets, the more likely they are influenced by these supermarkets, resulting in a decreased purchasing in target supermarket.

In this paper, we propose a new supermarket competition analysis method. At first, we transform the address of customers of target supermarket to latitude and longitude to compute the distance between customers' home and each supermarket in a city. After that, we convert the distance data into probability data through Huff's gravity model to make the distance data positive correlation to customer affected degree. Transformed distance data can be combined with logistic regression analysis to detect the influence of nearby supermarket competition on the customers of target supermarket. We'll conduct the experiments from the viewpoint of partial regression coefficients and accuracy.

II. Customer Analysis

Our research on supermarket competition analysis which is based on extended RFM analysis. RFM analysis contains three indicators, Recency, Frequency, and Monetary [4]. Recency denotes how recently did the customer purchase. Frequency means how often do they purchase. Monetary shows how much do they spend. RFM analysis can reflect the degree of customer goodness and widely used in customer classification. In Section 2-1, we will explain the RFM analysis in details. In addition, distance data based supermarket competition analysis will be introduced in section 2-2. In Section 2-3, we explain related research and the position of this research, which we have covered in the previous section.

A. RFM analysis

The RFM model is first proposed by Hughes in 1994, and it is a model that differentiates important customers from large transaction data [4]. Chen and colleagues propose an extended model of RFM analysis for the challenge prediction problem of customers in the logistics industry [17]. Later, the research that combines machine learning methods has also been reported. Tanaka and collegues considered the RFM and logistic regression analysis to detect the specific customers' preference for different kinds of supermarket items [18]. They set the month elapsed from customer ' s last shopping record to the data statistic day, the frequency of customer come to store and the purchase amount of customer has spent in a time interval for R, F and M values, respectively. This may cause the linear correlation between RFM values since the F value rised, the M value also increases. Another problem is that the RFM values are in different digital dimensions. In order to solve these problems and make the analysis results more accurate, we consider to convert the RFM values into the form of customer RFM rank according to the clustering method proposed by Wu and Lin [9]. Although PCA and normalizition analysis can also play a role in data dimensionality reduction, they have the defects of poor interpretability. We set R as days elapsed from last sales record to data statistics day, F value similar to Tanaka's research and M value as the average of one time purchase amount in a time interval. We order R value by ascending and F, M value by descending and rank them 5 levels according to top rank 20%, 40%, 60%, 80%, respectively. The example of RFM rank is shown in the following table.

TABLE I
THE EXAMPLE OF RFM RANK.

Customer	*R*	*F*	*M*	*R rank*	*F rank*	*M rank*
C1	55	3	1926	3	2	3
C2	354	1	150	1	1	1
C3	7	54	2111	4	5	3
C4	332	1	3900	1	1	5
C5	339	7	932	1	3	1
C6	57	11	1517	3	4	2

B. Supermarket competition analysis

In the field of supermarket competition analysis, Madhav N. Segal et al. [5] demonstrated the usefulness of combining retail market segmentation with competitive analysis as a very effective method to understand the dynamics of retail markets and to analyse strategic options for supermarket chains. Rajiv Lal et al. [6] investigated the factors contributing to every day low pricing's success by analyzing the competition between supermarkets through a game theoretic analysis of a market consisting of both time constrained consumers and cherry pickers. Their analysis focus on price and competitive strategy. None of their studies can well reflect the impact of competitive retail stores on target store customers. In this research, we have address information and 2 year shopping data of member customers of target chain supermarkets A1 and A2. We converted the address information of customers to longitude and latitude to compute the distance to each supermarket in Higashihiroshima city in Japan by Euclidean distance. We found the tendency that the closer the residences of customers to target supermarket, the higher the shopping quota is, and vice versa. Fig 1 shows the tendency of distance and per month consumption of customers of target supermarket A1. For the sake of space, here we only show the relevant situation of target supermarket A1. Then we considered whether the closer the customers to competitive supermarkets, the more likely they are influenced by these supermarkets. Fig 2 shows the example for the comparison of the tendency of distance and per month consumption of customers of target supermarket A1 and the tendency of distance to 4 competitive supermarkets less than 3km trade area and per month consumption of customers of target supermarket A1 by polynomial regression analysis. It is obvious that the customers of target supermarket A1 close to the competitive supermarkets are affected by those supermarkets. According to the research of Tanaka et al. [18], they discovered logistic regression coefficients can effectively reflect specific customers' preference for different kinds of products in chain supermarkets by investigating ID-POS Data. ID-POS data analytics provides store-level inventory analytics in the form of easy-to-grasp data visualizations that enable inventory managers and sales & marketing teams alike to identify trends and improve product sales. Therefore, our supermarket competition analysis is to analyze the degree of influence of nearby competitive supermarkets on customers of

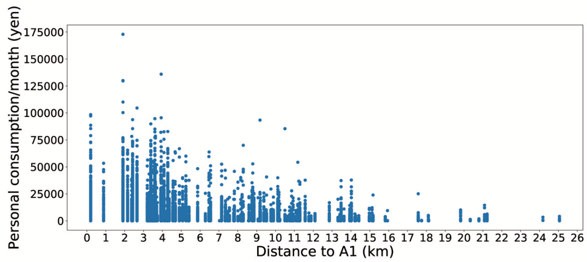

Fig. 1. The distance and sales for A1

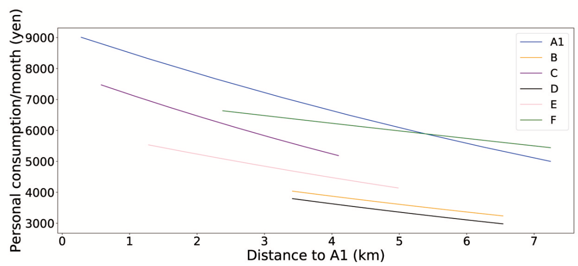

Fig. 2. The comparison of A1 with 5 competitors

target supermarket by ID-POS data and distance factors between residences of customers and competitive supermarkets.

C. Position of this research

Our research is focus on the degree of influence of supermarket competition on superior customers. It is a new approach combined with RFM analysis, distance factors and machine learning technique to detect the supermarket competition problem. Our new idea is to convert the distance factors to uniform attractiveness probability by Huff's gravity model. It can not only making distance factors positive correclation to competitive influence, but also solving multicollinearity problems. Then, supermarket managers can adopt corresponding promotion strategies according to the influence of superior customers from competitive supermarkets.

III. Analytical Approach

In section 3-1, we explain the defination of good customer. In section 3-2 and 3-3, we show how to analyze the competive influence.

A. Decyl Analysis

In the supermarket industry, it is said that a limited number of customers about 20% will produce 80 sales. According to Tanaka et al. [18], we use decyl analysis to define the good customer for target supermarket chain of this study. Decyl analysis is a method of examining the influence on the total purchase amount for each group by dividing customers into groups of ten-fold in order of purchase amount. Results are shown in Fig 3. It was found that 80.01% sales are generated in the top three groups. Therefore, good customers of supermarket chain A in this research for first one year are defined as the top three groups.

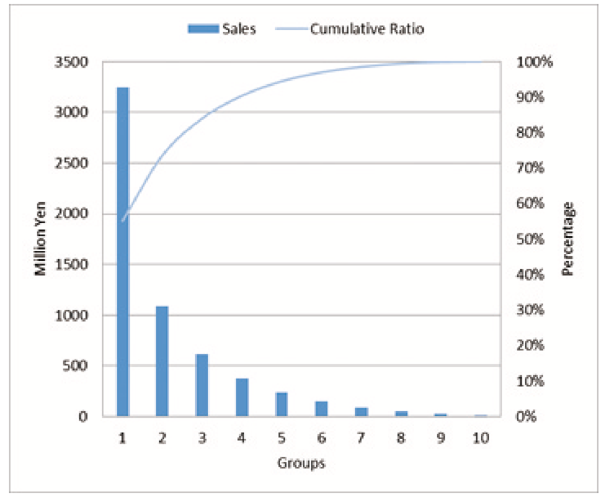

Fig. 3. The Decyl analysis for good customer

B. Logistic Regression Analysis

Logistic regression is a machine learning algorithm for binary classification. It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. The defination of logistic regression is as below. Let p_c be a probability of customer c is a good customer. ω, x and d denotes partial regression coefficients ,explanatory variables and bias, respectively.

$$p_c = \frac{1}{1 + e^{(d + \omega_1 x_{1,c} + \omega_2 x_{2,c} + \ldots + \omega_k x_{k,c})}} \tag{1}$$

In this research, it can be used for detecting the influence of competive supermarket on superior customers of target supermarkets. The top three customer groups in the latter one year of decyles analysis are tagged as target variables of the good customer classification. For the explanatory variables, we use 3 indicators of RFM analysis and the distance factors between customer address of target supermarkets and each supermarket in Higashihiroshima city.

C. Huff's gravity model

We propose a new approach to convert the distance factors to attracitiveness probability by Huff's gravity model which is defined as below.

$$p_{ij} = \frac{\frac{s_j}{d_{ij}^{\alpha}}}{\Sigma_{j=1}^{n} \frac{s_j}{d_{ij}^{\alpha}}}, \quad \Sigma_{j=1}^{n} p_{ij} = 1 \tag{2}$$

The business circle of a store depends on its attraction to customers, while the attraction of a store to customers in a region can be measured. This attraction mainly depends on two factors, the size of the store denoted as s and the distance between store and customers denoted as d. Therefore, p_{ij} denotes the attractiveness probability of customer i will go shopping at store j and α denotes the distance decline coefficient. It is easily find that the distance will be highly correlated with the RFM rank. The new idea is that if we fix the size of every store, distance can be converted into a uniform attractiveness probability. The attractiveness probability positive correlation to the competitive influence since distance negative correlation to the attractiveness probability. The multicollinearity problem also can be solved since the sum of the attractiveness probability per customer is 1 irrespective of the high or low of the RFM rank.

IV. Experiment

The experiments are executed on Windows 8 with 2.50GHz Inter Core i7 and 8GB memory and conducted on python environment. For all cases, the store area is fixed as $1000m^2$ and distance decline coefficient is fixed as 2.

A. Experimental Data

In this research, we use ID-POS data for 2 years of one supermarket chain including A1 and A2 in Higashihiroshima city in Japan. There are 176076 customer information and 2251 categorized product information and 40977672 sales records in the ID-POS data. In addition, there are 30 supermarkets including 2 target supermarkets in Higashihiroshima city.

B. Experimental Procedure

The ID-POS data of target supermarket chain for 2 years is divided into the first one year as current customer information and the second one year as future customer information. We combine RFM rank indicators with 30 attractiveness probability to build feature quantities for 2 years customer information. Decyl analysis was also conducted on 2 years ID-POS data respectively. Top 3 groups was tagged as good customers as objective variables. We use the customer ID in the ID-POS data, 33 feature quantities as explanatory variables and objective variables to build experimental data. The experiment data is divided into 2 pieces, 75% for training data and 25% for verification data. The oversampling and undersampling problems are judged that it is unnecessary in this experiment [16]. We train the first year experiment data of each customer by logistic regression to build the model and use the constructed model to classify the good customers in second year. There are 2 stages experiment in our research. The first stage is an experiment of classification on entire supermarket chain. The second stage is an experiment of classification on individual stores of target supermarket chain. Both of the stages include two model analysis, RFM model and RFM+ model. The RFM model only includes RFM rank explanatory variables. The RFM+ model include both 3 RFM rank and 30 attractiveness probability explanatory variables. The evaluation of the model is carried out from two viewpoints of accuracy, precision, recall rate, classification accuracy using F-score, and feature understanding of good customers.

C. Experimental Analysis for supermarket chain

Table II presents the prediction results of proposed 2 models for entire customers of target supermarket chain A. To investigate the performance of each model for customer classification, we generate 2 models. 'RFM-A' shows the classification model only includes 3 RFM rank indicators and build for the customers of entire supermarket chain A. 'RFM-A+' denotes the classification model combined RFM 3 indicators with 30 attractiveness probability. The all cases show that the RFM+ model is superior to RFM model for good customer classification. The statistical significance level also can be confirmed that all cases are less than 5%. Next, by means of regression coefficient analysis, supermarket competition analysis is carried out.

Table III shows the RFM model of good customer classification for customers of entire supermarket chain A. In order

TABLE II
THE ACCURACY ANALYSIS FOR CHAIN A

	Accuracy	*Precision*	*Recall*	*F1-score*
RFM-A	91.89%	79.82%	97.47%	0.88
RFM-A+	92.05%	80.28%	97.67%	0.88

TABLE III
THE RESULTS OF RFM MODEL FOR CHAIN A

Variables	***Coefficients***	***P values***
Intercept	-20.01	0.0%
R rank	0.26	0.0%
F rank	3.47	0.0%
M rank	1.96	0.0%

TABLE IV
THE RESULTS OF RFM+ MODEL FOR CHAIN A

Variables	***Coefficients***	***P values***
Intercept	-19.83	0.0%
R rank	0.26	0.0%
F rank	3.53	0.0%
M rank	2.04	0.0%
A2	0.71	0.0%
A1	0.29	0.0%
C2	0.12	0.0%
B1	0.11	0.0%
N2	0.04	0.0%
C7	-0.01	0.0%
M	-0.09	0.0%
J1	-0.12	0.0%
C4	-0.16	0.0%
I	-0.19	0.0%
C6	-0.22	0.0%
L	-0.42	0.0%
H2	-0.51	0.0%
H1	-0.57	0.0%
F1	-0.62	0.0%
F3	-0.67	0.0%
C9	-0.70	0.0%
F2	-0.73	0.0%
G	-0.80	0.0%
C5	-0.81	0.0%
B2	-0.86	0.0%
B3	-0.95	0.0%
J2	-1.02	0.0%
E	-1.11	0.0%
N1	-1.12	0.0%
C8	-1.16	0.0%
C3	-1.31	0.0%
K	-1.71	0.0%
C1	-1.94	0.0%
D	-3.30	0.0%

to avoid the linear correlation for RFM indicators, we set R rank by ranking 5 level for the days elapsed from customers' last shopping records to the data statistic day. Similar to R rank, we also rank the frequency of customer come to store and the average of one time purchase amount in current and future one year for F and M rank, respectively. All partial regression coefficients are positive show that good customers have the attribution of high RFM rank. This is consistent with intuitive understanding. By unifying the dimensions of the three indicators, we can find that F value is more important for high-quality customers. However, only through the three indicators, we can get too little useful information, and can not well analyze the impact of supermarket competition.

Table IV presents the competitive analysis by the partial regression coefficients of RFM and attractiveness probability for customers of target supermarket chain A. There are 7 supermarket chains in Higashihiroshima city including 1 target supermarket chain A and 6 competive supermarket chains B, C, F, H, J and N. We use the chain name with number to denote each store of supermarket chains. We order the the 30 attractiveness probability coefficients by descending and find that all other competitive supermarkets give the good customers of target supermarket A negative influence except supermarket B1, C2 and N2. The statistical significance level are all less than 5%. The values of target supermarkets A1 and A2 are positive, which is consistent with the intuitive idea since the closer to them, the probability of good customer increases. The good customers of supermarket chain A is strongest affected by supermarket C7, the supermarket manager should consider corresponding competition strategy to adopt for C7.

D. Experimental Analysis for individual supermarkets

Table V and VIII presents the prediction results of proposed 2 models for A1 and A2 store respectively. We build 4 models. 'RFM-A1', 'RFM-A2' shows the classification model only includes RFM 3 indicators and build for the customers of A1 and A2 store, respectively. 'RFM-A1+', 'RFM-A2+' denotes the classification model combined RFM 3 indicators with 30 attractiveness probability. The all cases also show that the RFM+ model is superior to RFM model for good customer classification in individual supermarkets. The statistical significance level also can be confirmed that all cases are less than 5%.

Table VI and IX shows that the partial regression coefficients of RFM model for 2 individual supermarkets and supermarket chain A are similar to each other. Table VII and X presents the competitive analysis by the partial regression coefficients of RFM+ model for customers of A1 and A2. The statistical significance level are also all less than 5%. Both cases of supermarket A1 and A2 have a high positive value for good customer classification in their cases, respectively. It is consistent with the intuitive idea since the closer to corresponding individual supermarket, the probability of good customer increases. We also find that even among individual supermarkets of supermarket chain have competition with each other. In case of supermarket A1, A2 has a negative value, vice versa. Similar to the case of supermarket chain, we descend 30 attractiveness probability coefficients for A1 and A2. From the results, except for C6, L, A1 and N1 give a positive impact, the good customers of supermarket A1 is strongest affected by supermarket B1. Except for A2 and C2 give a positive influence, the good customers of supermarket A2 is strongest affected by supermarket I. The related supermarket manager of A1 and A2 should consider corresponding competition strategy to adopt for their strongest competitors.

TABLE V
THE ACCURACY ANALYSIS FOR A1

	Accuracy	*Precision*	*Recall*	*F1-score*
RFM-A1	92.33%	79.71%	95.34%	0.87
RFM-A1+	92.63%	80.51%	95.38%	0.87

TABLE VI
THE RESULTS OF RFM MODEL FOR A1

Variables	***Coefficients***	***P values***
Intercept	-22.44	0.0%
R rank	0.32	0.0%
F rank	4.04	0.0%
M rank	1.77	0.0%

TABLE VII
THE RESULTS OF RFM+ MODEL FOR A1

Variables	***Coefficients***	***P values***
Intercept	-21.56	0.0%
R rank	0.33	0.0%
F rank	3.94	0.0%
M rank	1.83	0.0%
C6	0.69	0.0%
L	0.29	0.0%
A1	0.25	0.0%
N1	0.13	0.0%
B1	-0.13	0.0%
F2	-0.14	0.0%
M	-0.23	0.0%
H1	-0.32	0.0%
I	-0.42	0.0%
C1	-0.53	0.0%
F1	-0.59	0.0%
C3	-0.65	0.0%
C2	-0.68	0.0%
C9	-0.71	0.0%
B3	-0.76	0.0%
J1	-0.78	0.0%
B2	-0.80	0.0%
E	-0.85	0.0%
C8	-0.91	0.0%
C5	-0.96	0.0%
N2	-0.99	0.0%
H2	-1.01	0.0%
F3	-1.02	0.0%
C4	-1.06	0.0%
C7	-1.11	0.0%
G	-1.25	0.0%
A2	-1.64	0.0%
J2	-1.69	0.0%
D	-1.71	0.0%
K	-1.97	0.0%

TABLE VIII
THE ACCURACY ANALYSIS FOR A2

	Accuracy	*Precision*	*Recall*	*F1-score*
RFM-A2	92.44%	82.69%	95.81%	0.89
RFM-A2+	92.70%	83.75%	95.99%	0.89

TABLE IX
THE RESULTS OF RFM MODEL FOR A2

Variables	*Coefficients*	*P values*
Intercept	-20.37	0.0%
R rank	0.20	0.0%
F rank	3.48	0.0%
M rank	2.23	0.0%

TABLE X
THE RESULTS OF RFM+ MODEL FOR A2

Variables	*Coefficients*	*P values*
Intercept	-19.68	0.0%
R rank	0.20	0.0%
F rank	3.43	0.0%
M rank	2.26	0.0%
A2	1.06	0.0%
C2	0.57	0.0%
I	-0.14	0.0%
B3	-0.21	0.0%
F1	-0.30	0.0%
C4	-0.34	0.0%
J1	-0.36	0.0%
C7	-0.41	0.0%
B1	-0.51	0.0%
C5	-0.55	0.0%
H1	-0.56	0.0%
C9	-0.59	0.0%
M	-0.66	0.0%
L	-0.68	0.0%
B2	-0.71	0.0%
G	-0.76	0.0%
C6	-0.77	0.0%
N2	-0.79	0.0%
J2	-0.81	0.0%
H2	-0.85	0.0%
C1	-0.89	0.0%
E	-0.94	0.0%
C8	-0.98	0.0%
F3	-1.01	0.0%
N1	-1.12	0.0%
D	-1.14	0.0%
F2	-1.17	0.0%
A1	-1.23	0.0%
C3	-1.40	0.0%
K	-1.43	0.0%

V. CONCLUSION

In this paper, we propose a mothed to build the good customer classification model for customers of individual supermarkets and supermarket chain. In the experiments, we estimate the RFM and RFM+ model by logistic regression coefficient analysis. Both of 2 models have high accuracy for good customer classification. RFM+ model is superior to RFM model from the viewpoint of accuracy and coefficient diversity. It can help the supermarket managers to grasp the influence of competitive supermarkets.

REFERENCES

[1] Reilly WJ, The law of retail gravitation, New York: Knickerbocker Press, 1931.

[2] Huff, David L, " Defining and Estimating a Trade Area, " Journal of Marketing, Volume 28, 1964, pp. 34-38.

[3] Nakanishi, M. and Cooper, L.G, "Parameter estimation for a multiplicative competitive interaction model-least squares approach," Journal of Marketing Research, 11, 1974, pp. 303-311.

[4] Hughes, A, Strategic database marketing. Chicago: Probus Publishing Company, 1994.

[5] Madhav N. Segal, Ralph W. Giacobbe, "Market Segmentation and Competitive Analysis for Supermarket Retailing", International Journal of Retail & Distribution Management, Vol. 22 Issue: 1, 1994, pp.38-48

[6] Rajiv Lal and Ram Rao, Supermarket Competition: The Case of Every Day Low Pricing, Marketing Science, Vol. 16, No. 1, 1997, pp. 60-80

[7] Hisao Iizuka, Daisuke Yonemura, Hideki Toyoda Behavior Analysis by Customer Rank, Operations Research 2, 2003, pp. 94-99

[8] Rust and Lemon and Zeithaml, Return on Marketing: Using Customer Equity to Focus Marketing Strategy, Journal of Marketing 68 1, 2004, pp. 109-127.

[9] Wu, J., & Lin, Z, Research on customer segmentation model by clustering. In Proceedings of the 7th ACM ICEC international conference on electronic commerce, 2005.

[10] M. Haenlein and A.M. Kaplan and D. Schoder, Valuing the Real Option of Abandoning Unprofitable Custorners When Calculating Customer Lifetime Value, Journal of Marketing 70, 2006, pp. 5-20.

[11] H.M. Chuang and C.C. Shen, A study on the application of data mining techniques to enhance customer lifetime value based on the department store industry, The seventh international conference on machine learning and cybernetics, 2006, pp. 168-173.

[12] V. Ravi, Advances in Banking Technology and Management: Impacts of ICT and CRM. Information science reference, Hershey, New York, Yurchak Printing Inc, 2008.

[13] Y. Kishimoto, T. Takahashi, M. Takahashi, T. Yamada, K. Tsuda, T. Terano, Analysis of customer behaviors and sales promotion measures by agent simulation, Information Processing Society Research Report Intelligence and Complex Systems, No. 16, 2009, pp. 87-92

[14] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh, Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study, Procedia Computer Science 3, 2011, pp. 5763

[15] Hui-Chu Chang, Hsiao-Ping Tsai, Group RFM analysis as a novel framework to discover better customer consumption behavior, Expert Systems with Applications 38, 2011, pp. 14499-14513

[16] A I Marqus,V Garca,J S Snchez, On the suitability of resampling techniques for the class imbalance problem in credit scoring, Journal of the Operational Research Society, Volume 64, Issue 7, 2013, pp. 1060-1070, 304.

[17] Kuanchin Chen, Ya-Han Hu, Yi-Cheng Hsieh, Predicting customer churn from valuable B2B customers in the logistics industry: a case study, Information Systems and e-Business Management, Volume 13, Issue 3, 2015, pp 475-494.

[18] T. Tanaka, T. Hamaguchi, T. Saigo, K. Tsuda, Classifying and Understanding Prospective Customers via Heterogeneity of Supermarket Stores, International Conference on Knowledge Based and Intelligent Information and Engineering Systems, 2017, pp. 956-964