International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:10, No:6, 2016

Determination of the Bank's Customer Risk Profile: Data Mining Applications

Taner Ersoz, Filiz Ersoz, Seyma Ozbilge

Abstract**—**In this study, the clients who applied to a bank branch for loan were analyzed through data mining. The study was composed of the information such as amounts of loans received by personal and SME clients working with the bank branch, installment numbers, number of delays in loan installments, payments available in other banks and number of banks to which they are in debt between 2010 and 2013. The client risk profile was examined through Classification and Regression Tree (CART) analysis, one of the decision tree classification methods. At the end of the study, 5 different types of customers have been determined on the decision tree. The classification of these types of customers has been created with the rating of those posing a risk for the bank branch and the customers have been classified according to the risk ratings.

Keywords**—**Client classification, loan suitability, risk rating, CART analysis, decision tree.

I. Introduction

UP until recently, the financial sector either failed to meet the demands and needs of the customers or turned a blind eye to these demands and needs. They have started to seek new strategies. Among these new strategies have been such issues as the customer-focused culture, organizational structure, customer profitability, customer value, continuity of customer relationships, and improvement of the customer information systems, customer differentiation and customer loyalty [1]. The Customer Relationship Management (CRM) is aimed at retaining the existing customers and increasing their potential rather than finding new customers. This is a business that understands and knows the customers, treats them in a special manner, regards them as their copartners and designs products and services for the customers not vice-versa [2]. Due to these differences in the banking sector, types of customers are determined with a view to minimize the delays in the payments by the customers.

The simplest and most basic classification technique for the bank customers is the demographic classification [3]. Due to the difficulty of the updating of the demographic data in the long run, the banks need to recognize their customers and analyze their behaviors better. Classification of the customers' demanding credit is of great significance for the banks. This classification is aimed at determining in clear terms whether the customer will be granted credit or not in the initial stage. This requires the banks to establish a strategy to classify the behavior of the customers. With this method, the use of the information technologies for the analysis of the customers' transactions has become widespread and in particular, the data mining has gained more importance [4].

In recent years, the data mining analyses have been undertaken as a result of the development of the information systems and technologies. Large quantities of data serve to shed light on many issues such as minimizing the business risks in the banking sector, estimation of the risk rating, reduction of costs, stability analysis, determination of the daily amount of cash to be distributed to the ATMs, accurate and efficient credit decisions and increasing customer and employee satisfaction [5].

It is a must in the banking sector to determine the main financial categories relating to the early warning system for crises and establish the financial indicators of these categories. This determination is essential for the *risk factor* that is the most vital factor in the banking sector. The objective is to reduce the risk in the banking sector with the accurate use of the relevant data.

Reference [6] examined to reduce the risk in the banking sector by using information about the credit card holders. Reference [7] centered the study on the data mining. To this end, they studied life probabilities, hazards probabilities and regression models for a data set pertaining to the credit card holders. The research found that the age, income and marital status were significant risk factors that affect the card holders' decisions to stop using the cards. Reference [8] examined the credit risk transfer activities and systemic risk. Reference [9] found that less credit is granted to those with customer relationships for a year or less than a year or to other financial counterparts. Reference [10] applied on 173 firms in industry and service sectors in Istanbul Stock Exchange-100 Index used and their annual financial indicators from 2004 to 2006. It was determined the most important factors of firms that operate in industry and service sectors. Reference [11] tested discriminant analysis in conjunction with the artificial neural network model in the early detection of the banking failures and obtained from both of the models highly predictable results for a period of one or two years prior to the failures. Reference [12] examined the similarities and differences between the models used in the determination of the financial failures and fraudulent financial statements. 52 variants were used in this research carried out in Taiwan. These variants were analyzed with the logistic regression, decision trees and

Taner Ersoz is with the Karabuk University, Faculty of Business Administration, Department of Actuarial and Risk Management, Balıklarkayası Mevkii 78050, Karabuk, Turkey (e-mail: tanerersoz@karabuk.edu.tr).

Filiz Ersoz and Seyma Ozbilge are with the Karabuk University, Engineering Faculty, Department of Industrial Engineering, Balıklarkayası Mevkii 78050, Karabuk, Turkey (e-mail: fersoz@karabuk.edu.tr, seymaozb@gmail.com).

artificial neural network methods. The findings of the research suggested that a number of variants were effective in the determination of the financial failures and fraudulent financial statements. It was calculated that the accuracy classification rate of the logistic regression was 99%, the artificial neural network, 91% and decision tree method, 95%. Reference [13] examined the important factors affecting the capital structures of the firms doing business under the industrial and service sectors of the Istanbul Stock Exchange (ISE) through the CART analysis. 38 different financial indicators chosen for the research suggested that the most crucial determinants of the capital structures of the enterprises are the liquidity, asset use efficiency and business risk.

In this research, the bank customers have been considered as individual or SME (small and medium-sized enterprises) customers and one of the decision tree models, CART analysis has been used. Based on the data pertaining to the bank customers, the estimated delay period relating to the monthly credit installment payments and risk rating for the bank branch have been determined.

II. Materials and Methods

In this research, we have studied 187 individual and corporate customers based on the data pertaining to the registered customers of a bank branch in the period 2010-2013 and a total of five variants have been taken into consideration including *the amount of the credit granted, the number of the installments, the number of delayed days for the credit payments, the payments to the other banks and the number of the banks to which one is indebted.*

The objective is that based on the past data, the customers will be classified as "those eligible for credit" or "those not eligible for credit". This way, when customers request credit, the eligibility to be granted a credit will be determined in advance based on the classified data and the customers will be replied in a short period of time. In this research, the CART Analysis which is one of the most used classification methods of the Decision Tree has been preferred for the data mining.

The decision tree approach is a method used for the roughly calculation of the objective functions and through which the learning function is demonstrated via a decision tree. A decision tree is a tree-shaped descriptive and predictive model [14]. This model helps the decision maker what factors should be taken into consideration while making a decision and determine what kind of a relation each one of these factors has with the different outputs of a decision [15].

CART algorithm is a statistical procedure widely used for the production of classification and regression models based on the tree structure. CART tree model includes a hierarchy of univariate binary decisions. CART ensures the best selection for splitting the data into two groups at the root node and uses different splitting criteria. All these splitting criteria split the class labels in each subset as homogeneously as possible.

III. Research Findings

A correlation analysis has been carried out to demonstrate the relationships between the variants studied under the research. The result of the analysis has been shown in Table I.

TABLE I
Variables Correlation Relationship

Variables	Delayed Credit Installments (Day)	Correlation Relationship
Amount of the Credit (TL)	0,260	Medium
Number of the Installments (month)	-0,203	Medium
Payments to Other Banks (TL)	0,142	Medium
Number of the Banks to which one is Indebted (number)	0,063	Weak
Type of the Customer	-0,312	Strong

When Table I is analyzed, it has been found that the delayed credit installment has a positive and medium level relationship with the amount of the credit, a negative and medium level relationship with the number of the installments, a weak relationship with the payment to the other banks and a strong and negative level relationship with the type of the customer.

In the research, CART Analysis, a classification model, has been applied and while the dependent (target) variant has been determined as *the delayed days for the credit payments*, the independent variants have been determined as the amount of the credit, the number of installments, payments to the other banks, the number of the banks to which one is indebted and type of the customer. According to the decision tree in Fig. 1, 187 customers pay their credit installments within 24 days. It has been found that the most significant variant for the delayed credit installments is the "the amount of the credit". The decision tree has been created based on the reference point of 67750 TL as the credit amount. As can be seen in Fig. 1, 147 customers out of 187 customers, 78.61%, has received 67750 TL or less and the estimated delayed period of credit is 21 days. The estimated delayed period of credit for 75 SME customers is 25 days. The other most significant variant for the delay of the credit is "the number of the installments". 20 SME customers the number of whose installments is 3.5 or less make credit payment within 32 days. 55 SME customers the number of whose installments is 3.5 or more make credit payment within 23 days.

The estimated payment period for 72 "individual" customers is 16 days. It has been found that the delayed credit installment payment is less for the individual customers than SEMs.

While 40 of the 187 customers receive 67750 TL or more, the estimated delayed payment period is 24 days. The delayed credit payment period is 35 days for 30 customers the number of whose installments is 9.5 or less. The delayed credit payment period is 18 days for 8 customers the number of whose installments is 9.5 or more.

The tables formed on the basis of the target variant "delayed period of credit installments" as a result of the CART Analysis have been given below. Each one of these tables is a table where the customers of a bank branch are classified according to the risk rating. According to the "node" data in the decision

International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:10, No:6, 2016

tree, the tables have been formed based on the number of customers at each node, data relating to the delayed installment payment and reference code pertaining to the customers.

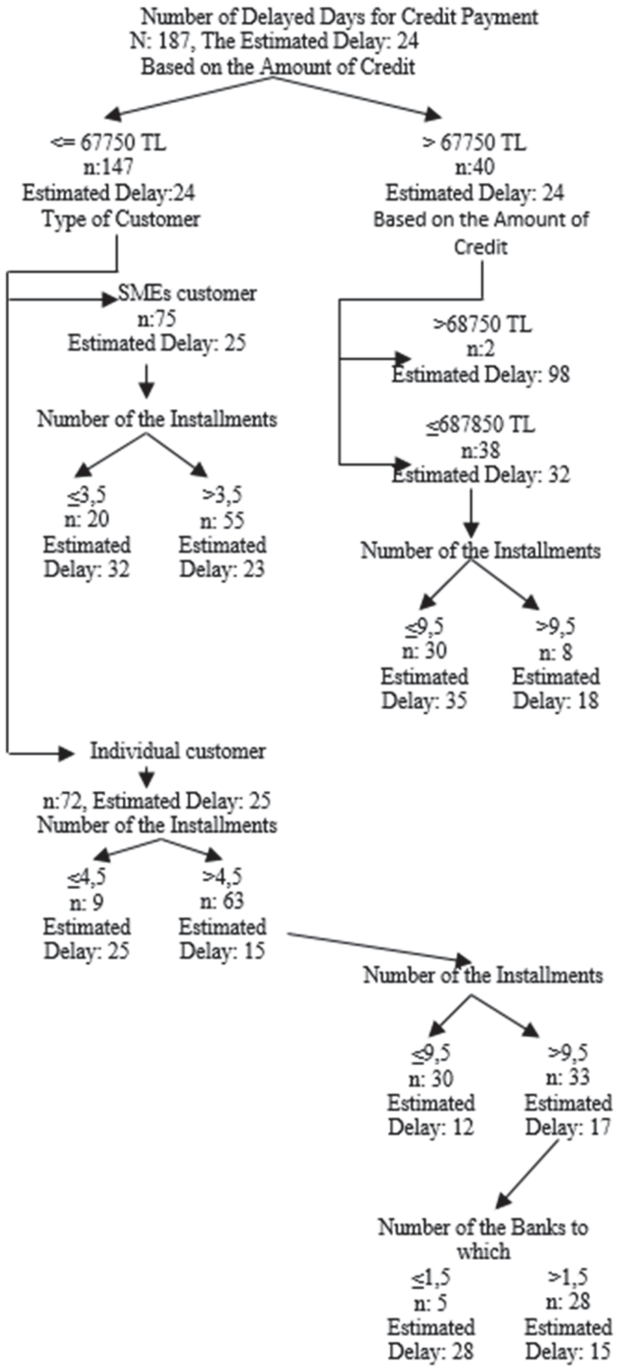

Fig. 1 CART Decision Tree

Table II shows customers who make most delayed payments and information about the delayed periods. According to the information obtained from the bank branch, these customer groups delaying their credit payment for a period of 50 days or more will in no way or under any circumstances be granted credit. No exception must be applied because these customers slow down the cash flow in the bank branch and cause the disruption of some plans. Besides, they reduce the competition grade of the bank branch. The personnel dealing with them get partial low grades and cause their year-end grades to be low. Hence, these customers defined with the Customer Reference Code are *1st Degree Risky Customers* for the bank branch.

TABLE II
THE BANK BRANCH'S 1ST DEGREE RISKY CUSTOMERS WHO DELAY THEIR CREDIT INSTALLMENT PAYMENTS FOR A PERIOD OF 50-98 DAYS

Node	Number of Customers	Average Delayed Period of Credit Payment (Day)	Customer Reference Code
5	2	97	H7, H8
18	3	51	I7, O10, P6
25	4	59	J8, K2, N7, O9
37	3	78	E8, H9, O8
38	9	50	A10, D3, F2, F3, G10, I4, J1, L5, R4

The Bank branch's 2nd degree risky customers who delay their credit installment payments for a period of 25-49 days have been shown in Table III. According to the information obtained from the bank branch, these customers can be granted credit again if they meet certain conditions. The customers shown in Node 23 can be granted credit provided that the requested credit amount is less than the previous credit and the number of installments is the same as the previous.

The customers shown in Node 26 can be granted credit provided that the requested the number of credit installment is at least two times the previous credit. As for the customers shown in Node 35, their credit requests must not be accepted unless being sure of their repayment abilities because they delay their payments despite of the fact that they meet all the requirements.

The customers shown in Node 39 can be granted credit provided that the requested credit amount is less than the previous credit and the number of installment is more than the previous one. Besides, these customers influence the competition grade of the bank branch. The personnel dealing with them get partial low grades and cause their year-end grades to be low. Hence, these customers defined with the Customer Reference Code are *2nd Degree Risky Customers* for the bank branch.

TABLE III
THE BANK BRANCH'S 2ND DEGREE RISKY CUSTOMERS WHO DELAY THEIR CREDIT INSTALLMENT PAYMENTS FOR A PERIOD OF 25-49 DAYS

Node	Number of Customers	Average Delayed Period of Credit Payment (Day)	Customer Reference Code
23	3	30	B9, M9, R7
26	7	38	C10, E9, F1, J10, J7, J9, P5
35	5	28	C4, C9, O2, P9, R1
39	9	29	B1, D6, D8, F4, H6, K9, M4, O3, O7

The Bank Branch's 3rd Degree Risky Customers Who Delay their Credit Installment Payments for a period of 15-24 days have been shown in Table IV.

TABLE IV
THE BANK BRANCH'S 3RD DEGREE RISKY CUSTOMERS WHO DELAY THEIR CREDIT INSTALLMENT PAYMENTS FOR A PERIOD OF 15-24 DAYS

Node	Number of Customers	Average Delayed Period of Credit Payment (Day)	Customer Reference Code
16	50	24	A2, A7, A8, B2, B7, B8, B10, C1, C2, C6, D2, D4, D7, F6, F9, F10, G1, G4, G7, G8, H1, H2, H4, H10, I6, I9, J3, K4, K10, L9, L10, M3, M10, N3, N4, N6, O1, O6, P5, P7, R6, S1, S3, S4, S5, S6, S9, T1, T2, T5
33	11	21	D9, E4, E5, G6, L4, L6, L7, M2, P1, P2, T3
36	28	15	A3, A5, A9, B4, B5, B6, C5, D5, H3, K5, K6, K7, K8, L2, L3, N2, N8, N10, O4, O5, P8, P10, R2, R3, R9, S2, S8, T4

Table IV shows the customers delaying their credit installment payment for a period of 15-24 days and data relating to the average delayed installment payment period for these customers. According to the information obtained from the bank branch, these customers delaying their credit installment payment for a period of 15-24 can be granted credit when certain conditions are met. However, they will be shown more tolerance than the 2nd Degree customers. By increasing the number of installments for the customers in Node 16 and Node 33, their credit request for as much as the previous amount can be accepted.

The credit requests of the customers in Node 36 must be accepted unless there is an exceptional case because they meet all the required conditions. They delay their payments partially. The credit requests of these customers must be generally accepted; however, the delayed credit payments must be prevented and the customers must be warned of this situation.

The Bank Branch's 4th Degree Risky Customers Who Delay their Credit Installment Payments for a period of 10-18 days have been shown in Table V.

TABLE V
THE BANK BRANCH'S 4TH DEGREE RISKY CUSTOMERS WHO DELAY THEIR CREDIT INSTALLMENT PAYMENTS FOR A PERIOD OF 10-18 DAYS

Node	Number of Customers	Average Delayed Period of Credit Payment (Day)	Customer Reference Code
27	4	21	E2, E6, G9, J4
28	5	9	C3, E10, G2, J5, S10
29	2	13	A1, D10
31	3	18	A6, J2, K3
40	9	12	E1, E7, G3, G5, I5, I8, M6, N5, T7
42	3	14	L1, R5, R8

Table V shows the customers delaying their credit installment payment for a period of 10-18 days and data relating to the average delayed installment payment period for these customers. According to the information obtained from the bank branch, these customers delaying their credit installment payment for a period of 10-18 can be granted credit when certain conditions are met. However, they will be shown more tolerance than the 3rd Degree customers. By increasing the number of installments for the customers in Node 27, Node 28 and Node 31, their credit request for as much as the previous amount can be accepted.

The credit requests of the customers in Node 29 must be accepted unless there is an exceptional case. The customers in Node 40 are those who can make regular payment provided that they request less credit than the previous one and the number of the installments is more than the previous one.

The requests of the customers in Node 42 must be accepted provided that the number of the installments is the same and the requested credit amount is less than the previous one because they meet all the required conditions. The credit requests of these customers must be generally accepted; however, the delayed credit payments must be prevented and the relevant variants must be studied in detail.

The Bank Branch's Least Risky Customers Who Delay their Credit Installment Payments for a period of less than 8 days have been shown in Table VI.

TABLE VI
THE BANK BRANCH'S LEAST RISKY CUSTOMERS WHO DELAY THEIR CREDIT INSTALLMENT PAYMENTS FOR A PERIOD OF LESS THAN 8 DAYS

Node	Number of Customers	Average Delayed Period of Credit Payment (Day)	Customer Reference Code
30	3	4	C7, C8, M1
32	3	6	A6, J2, K3
34	19	7	B3, F7, F8, H5, I1, I10, I2, I3, J6, K1, L8, M5, M7, M8, N1, N9, P3, S7, T6
41	2	6	A4, R10

Table VI shows the customers who delay their credit payments for a negligible period of time and the average number of delayed credit payment days. According to the information obtained from the bank branch, these customers can be granted credit under any circumstances. These customers are the group most tolerated. They are the customers enhancing the overall grade of the bank branch. They can be granted any reasonable amount of credit with the exception of some special cases.

IV. DISCUSSION AND CONCLUSIONS

In the study, the CART Analysis which is decision tree model based on the determined variants has been applied to the customers of a bank branch and 5 different types of customers have been determined on the decision tree. The classification of these types of customers has been created with the rating of those posing a risk for the bank branch and the customers have been classified according to the risk ratings, assessing the conclusions reached with the "Correlation Analysis" relating to the content of this classification.

The most significant factor in the classification of the customers of the bank branch is to determine the appropriate profiles of the customers and ensure the branch create a strategic plan. The objective of this strategic plan is to get a higher competition grade than other branches of the bank. Through this plan, the grade of the branch is determined and critical decisions relating to the branch are taken. And the grades of the personnel dealing with these customers are calculated and based on these grades, how much bonuses the personnel will be entitled at the end of the year and the changes in positions are determined.

The bank branches have annual targets and the personnel have individual aims based on these targets. How many targets are reached is determined in accordance with a grading system created by the center. Thus, customers with a high risk rating pose problems for the banks. They make it difficult for the bank branches to reach their targets. In order for the banks to reach their targets, the customers with a high risk rating must be diligently approached and monitored or their request for credit must be turned down.

According to the conclusions of the research, those customers with a less number of installment payment and those who are required to make payments to other banks and those the amount of whose credit is more and especially SMEs customers are more inclined to make delayed installment payments. This research is aimed at helping bank personnel make decisions faster and improve their foresight. Thus, the bank personnel can concentrate more on their other works, win other customers or the customers will be more satisfied when the requests of the customers are replied in a faster manner.

References

[1] F. Bryan, and S. Merlin, "CRM in Financial Services: A Practical Guide to Making Customer Relationship Management Work", Kogan Page Limited, Milford, Ct, USA, 2002.

[2] Peppard, J., 2000, "Customer Relationship Management (CRM) in Financial Services", European Management Journal, Vol. 18, No. 3, pp. 312–327.

[3] M.A. Evcim, "Segmentasyon ve Müşteri Prototipleri", CRM Institute Turkey, CRM Atölye Çalışmaları, (CD), 2004.

[4] T. Oktay, "Kurumsal Destek Sistemleri", Yöneticinin E-iş Rehberi, Mayıs, (6), 4-9, 2002.

[5] B.M Thuarisingham, "Web Data Mining and Applications in Business Intelligence and Counter Terrorism", CRC Press LLC, Boca Raton, FL, USA. 2003.

[6] O. Akgobek, and F. Cakiri, Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri, Harran Üniversitesi, Şanlıurfa 11-13 Şubat, 2009.

[7] N. Ata, E. Özkök, and U. Karabey, "Survival Data Mining: An Application to Credit Card Holders" Sigma Mühendislik ve Fen Bilimleri Dergisi, Cilt 26, No 1,33-42, 2008.

[8] R. Nijskens, and W. Wagner, "Credit Risk Transfer Activities and Systemic Risk: How Banks Became Less Risky Individually, But Posed Greater Risks to The Financial System at The Same Time", Journal of Banking and Finance, 35, 1391–1398, 2011.

[9] R.A. Cole, "The Importance of Relationships to The Availability of Credit", Journal of Banking and Finance, 22, 959-977, 1998.

[10] A.S. Albayrak, and Ş.K Yılmaz, "Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama", S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt 14, No 1, 31-52, 2009.

[11] U. Altunöz, "Bankaların Finansal Başarısızlıklarının Diskriminant Analizi ve Yapay Sinir Ağları Çerçevesinde Tahmini", Sakarya Üniversitesi İktisat Dergisi, 2(4), 1-21, 2013.

[12] L F. M.iou, "Fraudulent Financial Reporting Detection and Business Failure Prediction Models: A Comparison", Managerial Auditing Journal, 23(7), 650-662, 2008.

[13] A.S. Albayrak, and R. Akbulut, "Sermaye Yapısını Belirleyen Faktörler: İMKB Sanayi ve Hizmet Sektörlerinde İşlem Gören İşletmeler Üzerine Bir İnceleme", Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 22, 2008.

[14] Y. Ma, "Data Warehousing, OLAP and Data Mining: An Integrated Strategy for Use at FAA", Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, http://www.citeseer.ist.psu.edu/ma98data.html, 1998.

[15] C. Bounsaythip, and R. R. Esa, "Overview of Data Mining for Customer Behavior Modeling", VTT Information Technology Research Report, Version:1, s. 1-53. 2001.