Working together
www.rcis.ro

Revista de Cercetare si Interventie sociala

ISSN: 1583-3410 (print), ISSN: 1584-5397 (electronic)

THE EFFECT OF GENDER AND GENDER-DEPENDENT FACTORS ON THE DEFAULT RISK

Begum CIGSAR, Deniz UNAL

Revista de cercetare și intervenție socială, 2018, vol. 63, pp. 28-41

The online version of this article can be found at:
www.rcis.ro, *www.doaj.org* and *www.scopus.com*

Published by:
Expert Projects Publishing House

expert projects
publishing

On behalf of:
„Alexandru Ioan Cuza" University,
Department of Sociology and Social Work
and
HoltIS Association

REVISTA DE CERCETARE SI INTERVENTIE SOCIALA
is indexed by Clarivate Analytics (Web of Science) -
Social Sciences Citation Index
(Sociology and Social Work Domains)

The Effect of Gender and Gender-Dependent Factors on the Default Risk

Begum CIGSAR[1], Deniz UNAL[2]

Abstract

The concept of gender created by the society referencing to biological sex, and the rules, sanctions, anticipations, officials put on it, is a question that crowns injustice towards women today. This problem has caused and sustained great injustices and losses not only in daily life but also in economical area. In this study, it was tried to draw attention to the fact that the study that we are currently doing are shared so that the society should be shaken as soon as possible and away from the "gender" perception. The purpose of this study is to identify data mining classification algorithms that can be used to predict default risks using data on demographic and socioeconomic characteristics of individuals, to avoid possible payment difficulties and to reduce the problems that may arise when lending. Also going into default risks for women and men are examined and so indeed it was found that women are more sensitive to their repayments. From this point of view, variables affecting the going into default rate of women were examined.

Keywords: Big Data, gender, default risk, WEKA, data mining, logistic regression.

Introduction

We can think of the Big Data as the accumulation of knowledge transferred from generation to generation. Since humankind has existed, s/he has developed various ways to convey the information s/he has. Storing the acquired information has become an important need to transfer to future generations. For this purpose, the information stored with the developing technology has also created data repository. Raw Data does not make sense or use it alone. However, it is the basis for information and knowledge. That is, it needs to be related, grouped, interpreted, explained and analyzed (Ozen, 2017).

It can be said that there is no area where Big Data does not touch. The interest in the analysis of this data is also increasing because Big Data is used from the

[1] Cukurova University, Adana. TURKEY. E-mail: bcigsar@student.cu.edu.tr

[2] Cukurova University, Adana. TURKEY. E-mail: dunal@cu.edu.tr *(Corresponding Author)*

business world to a wide area such as artificial intelligence, informatics and management. The power of a Big Data also provides a more detailed view of some areas. These areas are; sectors such as banks, agriculture, health, social media management, retail and insurance companies and politics. With the emergence of Big Data concepts, the banking sector has become one of the sectors most affected by this issue and has to focus on the concept of Big Data. It is very important to determine the customer profiles as they are in the insurance, to provide appropriate services to these customers, to determine the appropriate pricing, to determine the risk of leaving the risk groups (IABE, 2015). The challenge of handling Big Data sets and the search for useful information from the ever-growing pool of data has led scientists to put forward the terms of data analysis and data mining.

One of the most important benefits of data mining is that almost every area can be used (Gorunescu, 2011: 361). With data mining, companies can overcome customer losses, identify which products are more likely to sell, which customers tend to buy products, avoid credit card fraud, and identify risk groups. These are the problems that almost every firm wants to solve. The areas of use of data mining that can provide solutions to these and other problems; marketing sector, engineering, health sector, banking, insurance, retail etc. are the areas. The areas of use of data mining that can provide solutions to such problems; marketing sector, engineering, health sector, banking, insurance, retail etc.

Use of data mining in banking

Value creation for the customer is the key to a successful business, no matter what the job description is. Providing customer satisfaction enables companies to increase their profitability in the long turn. If a product is developed with a customer base, or if the company develops its service approach in this way, it will increase customer satisfaction and loyalty. However, analyzes and market surveys to increase customer satisfaction will lead to an increase in the cost of the company. Taking these into consideration, a new understanding can be brought to the banking sector by analyzing data mining. With this approach, user opinions or complaints can be transformed into valuable information. The banking sector is a customer-focused establishment. For this reason, customer feedback is very important in the banking sector as well as in other customer focused firms. Therefore, analyzing the feedback we receive from our customers using data mining methods is very valuable for the banking sector. According to the results of these analyzes, promotions will provide special services to prevent customer losses and increase customer satisfaction. Using data mining in customer feedback will also allow the company to reduce the cost of doing this. Correct analysis of Big Data leads owners of it about correct investments/methods and makes them progress in earnings and status. Also, it provides to estimate potential risk or opportunities. But, companies having this huge data collection or Big Data may have to struggle for hidden costs and complexities in process of gathering of data, its analysis and evaluation.

If we think that especially the money governs the science, technology, and industry, or humanity in short, in an inevitable way, Big Data has more serious importance in companies where money is the main focus. Big Data, modeling and analyzing Big Data are not the new concepts and termsfor the actuaries, bankers or insurers obliged to understand the customers' behaviors, managing the money better or managing the system, reducing the possible risks or transferring the risks. Big Data analysis has great importance especially in determining the default risk and taking precaution in banking and insurance business, stating risks groups, rising customers' contentment according to customers' specific needs, identifing credit card fraud. Data mining is a data analysis technique that is based on statistical applications, which finds out valuable data from large quantities of data that are not related to previously identified data items.

The aim of data mining is to extract information by analyzing Big Data, making scientific and consistent deductions for the future (Okhovvat, 2015). Combined data from many databases are stored in data warehouses. Data mining methods (Inmon, 1991: 428) have emerged to analyze the data in the data warehouse and to access the information. Classification of Big Data clusters with data mining software using algorithms in data mining methods is the current research topic (Han *et al.*, 2012: 740). Data mining classification algorithms are widely used; medicine, biology, genetics. In addition to these areas, data mining algorithms are used in the banking sector in determining the credit risk (Sousa & Figueiredo, 2014). With these algorithms being used, possible risks can be estimated and the damage of the companies can be minimized.

Credit Risk Scoring

In the past decade, there are dramatic increases in credit purchases and credit card users. However, it is often seen that these users are delaying their payments or even sometimes are not able to make payments. This situation poses considerable risks for the banks. Some analyzes are made to predict these risks. These analyzes, made manually or intuitively, have left their place to computer assisted statistical methods by the development of technology. Data mining is one of these analyzes. Banks that are customer-focused are also benefiting from data mining in order to obtain information from this Big pool of data. Banks can build customer profile pools by evaluating demographic information from customers and even customer feedback using data mining techniques. With this pool, banks can better understand customers, avoid customer losses, identify risk groups and avoid potential risks. In addition, loans that meet the customer profile can be presented, and they can easily deal with the potential risks. In this way, the bank's financial position can be stabilized.

It is important to predict the risks before default occurs. Thus, the identification of the people who can not pay can be provided and measures can be taken without risk. In addition to their experience, credit services can make better decisions by evaluating the data they obtain through data mining applications. In banking, credit is also called as “risk”. The reason for such an identification is the possibility that each loan will not be paid or not paid on the due date. For this reason, dues amount at a specific date for a credit used by an individual or company can be described as risk (Selimler, 2015). Loan issuers face many risks, such as the failure of borrowers to delay payment or no payment, the change in market interest rates, and the loss of securities and stock market investments made by the institution. For example, three banks bankrupted because of credit debt in 1990 in Turkey. As a result, the country was involved in huge financial crisis and with the state guarantee given to bank deposits the failure of other banks was prevented. Subsequently, in 2001, the greatest banking crisis of the Republican history was experienced, and 25 banks were liquidated in this process. On this date, the ratio of banks’ troubled loans to total loans reached 30%. In the world, the crises caused by the loans are as follows: With Mexico’s failure to pay debts in US triggered the underdeveloped country’s debt crisis and in 1983 27 countries had to restructure their debts totaling $ 297 billion in US banks. 16 of the 27 nations are Latin American countries, and the four largest of them - Mexico, Brazil, Venezuela and Argentina - account for 74% (176 billion dollars) of all underdeveloped country debt (Wellons, 1987: 225). While 37 billion of this amount cannot be paid to 8 major US banks, non-performing loans have reached 147% of the capital of these banks (Federal Financial Institutions Examination Council, 1982). As a result, many big banks in the world have been forced to bankrupt due to troubled loans and these bankruptcies have led to a worldwide global crisis. Because of these crises, credit risks have become one of the issues to be investigated in the global economy (Dzelihodzic & Donko, 2013).

Credit risk management is a process involving the identification of potential risks, the measurement of these risks and the actual application of risk models (Van Gestel & Baesens, 2008). Financial institutions set up a model by considering possible customers’ informations such that age, gender, the city of residence, income, marital status, etc. and situations like prior credit payments. These models are designed to predict existing and future credit risks. According to the model, the acceptance or rejection of the customer’s applications is carried out (Huang, Chen, & Wang, 2006). In the credit rating analysis introduced by Fisher for the first time in 1936, only statistical discriminant and classification analyzes were used. Other credit scoring analyzes that are currently, the most widely used and data-driven; decision trees, logistic regression, k-nearest neighbors analysis, neural networks, support vector machines and so on (Sousa & Figueiredo, 2014). Analysts would like to increase credit volume, without any defaulting, so they resort to various analyzes to make a more accurate estimate of loan applications. Credit scoring is extremely important in terms of making faster decisions, reducing the costs of loan analysis, monitoring the existing accounts more closely, predicting the default risks while improving the institutions' competitiveness (Yeh & Lien 2009).

Classification of credit risk by using data mining

Data mining classification algorithms are used in many studies. it is desired to be reached by classification, is to estimate class labels unknown objects based on their probability of belonging to classes in order to construct a model in the training set (Witten, Frark, & Hall, 2011: 558). There are two main purposes of classification. These; classify the data and prepare data labels for prediction. Data mining techniques, which are important contributors to the development of information technology, can also be applied to credit scoring (Huang, Chen, & Wang, 2006). Ever since the banking has been seen, problems of the identification of risk groups and the appropriate scoring of them have also been encountered. In this context, ever since the problem has arisen, statistics science has developed various theorems to produce solutions. But under emerging technology and growing data volume, risk groups have become a bigger problem. However, while actuaries have identified risk groups, they have faced with the question of guessing group of new customers. To estimate this problem is not a situation in which basic statistical methods can succeed alone. Therefore, it is inevitable to use data mining methods because of these newly encountered questions and non-execution of Big Data analysis by manually. Almost every area of the field that uses data mining also uses statistics. However, although data mining and statistics are similar to each other, there are some fundamental differences between them. The first major difference is that statistics is a science and data mining is a method. Moreover, while data mining is concerned with the analysis of very Big Data, the statistics is carried out smaller data analyzes. When determining credit risks, various demographic and socio-economic characteristics are considered. It can be said that the most frequently analyzed variable is gender variable. The reasons for the disparity of payments between men and women are always curiosity. Although there are differences in payments between men and women in the analyze made, it is more important to research based on reasons of this differences. With more contributing women to the business life, while the economy has been revived, smaller credit ideas such as micro-credits have been actualized. Women's loyalty to their payments more than men and the timely payments of their loans reduces the risk of default. In this study, it was also determined that women were more loyal to their payments than men, and it was investigated which variables could be the effect.

Women and economy

Efforts to reduce credit risks and to anticipate possible risks should be taken into consideration especially in the context of genders. For this reason, the socio-economic and cultural structure of the family, which is the most important component of society, and the position of the woman in the family have become one of the most widespread research topics in the literacy. The position of the woman in the family and the society is known that has got changed by her achievement of economic independence, the realization of employment and contribution to the community economy. While the contribution of women's economic freedom to the society's economy is examined, the attitude of women is also important in the case of loan debts. It is therefore very important to handle the attitudes of women towards paying their loans. Because paying the debts of individuals is an important factor that opens the way to the economic freedom of the society. In developing countries, women have less taken place in the economy, which causes development to remain at a minimum level (Oren, Negiz, & Akman, 2012). For this reason, bringing women into the business world will provide economic strength. In other words, the country-wide development can only be achieved by equal employment of men and women (Karabiyik, 2012). In the studies conducted (Karabiyik; 2012), it is seen that the double-earned household model in which the woman and man work in the household has the more economic contribution. Therefore, ensuring women's employment is very important (Anthony & Horne, 2003). In particular, women's set up her own business, her competencies, and employment are realized through projects like KOSGEB (small and medium industry development organization in Turkey) like micro-credits, thus contributing to the economy. Micro-credit was given to low-income citizens who can't afford to gotten loans for their minuscule needs (Baltaci, 2011). Nowadays, micro-credit has become a source of income that especially for women's set up a business (Ozmen, 2012). Thus, along with increased female employment, economic developments can be observed and then national income increases (Anthony & Horne, 2003) (Hunt & Kasynathan, 2002). KOSGEB, micro-credit, and other supports provide social development while the creation of women's own employment through projects as well as it may facilitate the prevention of gender discrimination. It also helps raising the living standards while increasing household income levels (Ecevit, 2010).

Thus, woman in the household who has the equal role with man enables her to play an active role in the decision-making process. Also, if the woman has her own income, the income of the household will increase, which will reduce the risk of default. Therefore, problems that bank may experience while lending can be prevented, and at the same time, it will be ensured the development of the family economy and indirectly the economy of the country. So this shows that becoming prominence of the female factor in society is an important element for economic development (Ozmen, 2012). Moreover, the idea that woman has show more awareness of repayment than men has led many researchers to work on this subject (Goodman, Zhu, & Bai, 2016; Gan et al., 2012; Horkko, 2010). In this study, going into default risk probabilities of women in proportion to men were examined. It was determined that women were, in fact, more successful in paying their debts than men. With this results obtained, the effects of reducing the risk of going into default for women were discussed only using women's information and results were shared.

Methodology

Even the small amount of information obtained by data mining analysis plays an important role in predicting the reality of society and the economy. Data mining is a data analysis technique that is based on statistical applications, which finds out valuable data from large quantities of data that are not related to previously identified data items. The need for analysis of Big Data, especially in the field of economy, society and banking , is increasingly exploiting the use of data mining techniques. The use of data mining in the banking field is particularly useful in reducing the risk of default. For example, fraud detection can be done by analysis of data mining in insurance and banking sectors. Data mining techniques are also used in areas such as the ability to offer eligible policies / credits to eligible customers or better understanding of customer requests.

In addition, data mining is very important for actuaries, bankers and insurers who want to better understand the behavior of their customers and to manage money better and control the system. Thus, they will be more successful at reducing potential risks or transferring risks. Moreover, it may be possible to make premium offers and appropriate credit proposals for the customers.Thus, they will be more successful at reducing potential risks or transferring risks and with this work all these goals will be attained. In the previous study (Cigsar & Unal, 2018) the data set of 22745 observations and 14 variable data of the household head from the 59663 units survey conducted by TUIK (Turkish Statistical Institute) in 2015 were used. The data used includes the demographics of the households in various regions, their total income, and information such as whether they can pay their debts regularly in the last 12 months. The performance of Naive Bayes, Bayesian network J48, random forest, multi-layer perceptron, logistic regression classification algorithms that give the best results from the algorithms found in WEKA 3.9 data mining software; root mean error squared, Receiver Operating Characteristic (ROC) area, accuracy, precision, and recall statistical methods were compared. Also in previous study data mining classification algorithms that was used to predict default risks using data from demographic and socioeconomic characteristics of individuals. From this results that Logistic Regression classification algorithm was identified as best algorithm with statistical methods.

Findings

In this study, only the household heads over the age of 15 were selected from 59,663 unit survey data made by TUIK in 2015. The data contains demographic and socioeconomic characteristics of individuals. WEKA 3.9 software and logistic regression classification algorithm were used for data in 20275unit data set consisting of 12variables, one of which is a class variable. Payment / non-payment cases of past credit card debts are taken as class variables. The data set structure is given in *Table 1*.

Table 1. Data Structure

Attribute Name	Description
Age	Age
Gender	Gender (1=Male; 2=Female)
Marriage	Marital status (1=married; 2=Other)
Education	The Highest Education (1=illiterate; 2= primary school; 3=secondary school; 4=high school; 5=higher degree)
Work	Working Status (1=Working; 2=Looking for a job; 3=retired; 4=Other non-active)
Health	Health (1=good; 2=medium; 3=bad)
Region	Region (1=The Mediterranean; 2=The Aegean; 3=The Marmara; 4=The Black Sea; 5=The Central Anatolia; 6=The Eastern Anatolia; 7=The Southeastern Anatolia)
House	Housing Status (1= paying rent; 2=not paying rent)
Revenue	Individual Revenue (1=low income; 2=medium income; 3=higher income)
House loan	Within the last 12 months the non-payment of house rent; interest-bearing debt repayment or the housing loan payment as planned (1=No; 2=Yes)
Bills	In the last 12 months electricity, water and gas bills have not been paid as planned (1=No; 2=Yes)
Class	In the last 12 months installments credit card and other debt payments have not been paid as planned (1=No; 2=Yes)

WEKA

The WEKA data mining implementation was developed in the University of New Zealand. It is open source software written in Java under General Public License. It contains several supervised and unsupervised methods such as classification, clustering, association, data visualization, etc. For this study WEKA 3.9 implementation and its experimenter user interface have been used for classification problem and also used for specifying the risky attributes by logistic regression algorithm.

Logistic Regression

Logistic regression measures the relationship between response variable and independent variables like linear regression. It belongs to the family of exponential classifiers. Logistic regression refers to a classifier that classifies an observationinto one of two classes (Kumar & Sahoo, 2012). Logistic Regression algorithm analysis can be used when the variables are nominal or binary (Hosmer & Lemeshow, 2000). The data is analyzed by applying the discretization process for the continuous variables as in the Bayesian group.

Chi-Squared Attribute Eval Analysis

Chi-Squared analysis was applied via the WEKA Select Attribute panel to determine the variables that explain the model the best. Chi-square is an analysis that shows the value of the selected variable depending on the class variable used when the variables are nominal. It is recommended to subtract the lowest rank variant from the model according to this analysis. However, since WEKA can not apply chi-square based on the algorithm, it is preferable for the user to exclude variables from the model. The rank results of applying the chi-square to the variables are given in *Table 2*.

Table 2. Chi-Squared Analysis Result

Rank	Ranked Attributes
935.51	Bills
324.726	House Loan
60.641	House
54.848	Work
14.678	Region
11.931	Health
7.814	Education
2.814	Marriage
0.73	Individual Revenue
0	Age

Thus, based on the Chi-Squared Analysis which results shown above age attribute was removed from the data set. After testing the significance of the variables by chi-square test, the effect of the sex variables with the variables other than the age variable was investigated. The first objective of this study is to determine whether gender is an effect on the defaulting and if so, what effect it is. For this purpose, an analysis has been carried out primarily to determine which of gender has fallen in default. The result for the first objective is that gender is an effect of default risk and that women are more risk free in this regard. In other words, it can be shown that with odds ratio (*Table 3*), women have less tendency than men going into default.

Table 3. Odds Ratio of Gender Attributes

Class	
Variable	1
Gender =2	1.1174

The second objective of this study is to investigate the effect of selected demographic and socioeconomic characteristics on the likelihood of women defaulting using data mining logistic regression classification algorithm. For this purpose, only the women's data set has been examined and the attributes that affect the stepping down of the default rate of women has been investigated. In *Table 3*, only odds ratios (1 = male, 2 = female) were taken into account and Odds Ratio values were taken into account and the default rates of men and women were shown. According to odds ratio in the table, it can be said that the probability of women's default is lower. Taking this result into account, only the women's data set has been examined and the causes affecting the lowering of the default rate of women targeted for further investigation of this group have been investigated.

Variables in the data set for women as households only include the marital status, occupation, level of education, health status, living quarters, payment and non-payment of invoiced debts, variables indicating the payment of house rent, and class variable including whether credit card debts are paid or not. In *Table 4* below, only the lowest and highest odds ratios of the variables with significant results are given. Analysis interpretations according to the highest and lowest odds ratios of the variables examined are as follows:

Table 4. Odds Ratios by Attributes

Attribute Name	Sub Group	Odds Ratio
Work	Retired Looking for a job	1.2381 0.7632
Marriage	Other (not married)	1.0274
Education	Illiterate Secondary School	1.0453 0.9666
Health	Good Medium	1.0533 0.9184
Region	Mediterranean Southeastern Anatolia	1.1063 0.91
House	Not Paying Rent	1.1008
House Loan	Yes	0.3449
Bills	Yes	0.0838
Individual Revenue	Low income Medium income	1.0316 0.9682

52.5% of the community are married women. The model predicts that the odds of not going into the default risk are 1.0274 times higher for not married individual. Marriage status than they are for married individuals. Retired women are 10% of the community and 1.6% of the community are job seekers.Work=3 The model predicts that the odds of not going into the default risk are 1.2381 times higher for retired women than other working status. Also, it can be seen that the other inactive work status not going into default risk ratio is 1.204 times higher than other working status. There are 532 illiterate women in the community, 1501 elementary school graduates, 392 middle school graduates, 581 high school graduates and 585 higher education graduates. When looking at education levels of women, illiterate individuals' odds ratio is 1.0453. It means that illiterate persons' not going into default risk is 1.0453 times higher than other education levels. For women that education level is secondary school has more risky to going into default risk.

Health status 'good, medium, bad' were determined by individuals' survey answers. 36.1% of the women in the community had a good health status and 35.1% had a medium health status. From health attribute, women that have good health level are less risky to going into default risk 1.0533 times than other health levels. Women with moderate health status are more at risk than other health groups. Women living in the Mediterranean Region are 13.3% of the community. The women living in the Southeastern Anatolia Region are 10.2% of the community. Mediterranean Region's not going into default risk is 1.1063 times higher than other regions. Also, Southeastern Anatolia is the most risky region with 0.91 odds ratio.

People who do not have to pay rent have a very large majority, 74.9% of the community. Women that not paying rent are less risky to going into default (1.1008). Not paying rent means not having the extra debt to pay so women that not paying rent are more successful than the others in paying loan debts. Finally, when bill and house rent variables are examined, it is observed that those who regularly pay their credits, regularly pay their bills and house rents.Women that can not pay house loans and bills are also can not pay credit card debts. So if women do not pay house loans and bills they are more at risk of going into default. As predicted earlier, low-income individuals may receive lower credits or not receive credits. Therefore, repayments will be more regular if there will be a chance for them to get low-scale loans credit. The analysis result supports this finding. Looking at odds ratios, low-income individuals are 1.0316 times less likely to default than individuals with other income levels.

Discussion

The idea that women have more show awareness of repayment than men has led many researchers to work on this subject. The purpose of this study is to identify data mining classification algorithms that can be used to predict default risks using data on demographic and socioeconomic characteristics of individuals, to avoid possible payment difficulties and to reduce the problems that may arise when lending. The data used includes the demographics of the households in various regions, their total income, and information such as whether they can pay their debts regularly in the last 12 months.

The use of a Big Data set from TUIK and the analysis of these data with WEKA is the basis of this study. WEKA 3.9 software, logistic regression classification algorithm was applied to data and default rates were determined according to gender to calculate the risk of defaulting for women. First, the default rates of gender were determined based on the gender variable. The results show that women are more likely to pay their debts on time than men. That is, women are more loyal to their payments than men. After reaching the conclusion that women are more risk-free than men, a second analysis was carried out on women in more detail. As a result of this observation, the logistic regression algorithm was reapplied with the aim of investigating the reasons that would reduce the probability of default by taking only women's variables of education, marital status, region, occupation, health, bill, house rent and individual revenue levels. Logistic regression algorithm was applied again for this purpose. When we look at the results, it is observed that the household heads retired women's default risk is lower than the other women. According to the result of the analysis, non-married individuals and individuals have lower income have lower default risk than married ones and individuals having medium income respectively.

The results also show that the probability of defaulting for employees from householder women is lower than for non-working women. This situation shows that it is very important to increase women's employment. It is foreseen that the efforts to increase women's employment and the lending programs of banks and various credit institutions for women will be successful.

Conclusion

In general, it is seen that the gender discrimination of women in society or in business life is not a problem only for women. This is because the family householder women are less likely than men to fall into default, and this risk is further reduced if women work. In addition to countless studies in the literature on gender discrimination, it can be stated that women's employment and women's right in the family have a significant contribution to the country's economy, besides sociological necessity, as well as decreasing the risk of default in general.

References

Anthony, D., & Horne, C. (2003). Gender and Cooperation: Explaining Loan Repayment in Micro-Credit Groups. *Social Psychology Quarterly*, *66*(3), 293-302.

Baltaci, N.O. (2011). *Microcredit as Women's Empowerment Mechanism*. Ankara: Turkey Prime Ministry General Directorate of Women's Status

Cigsar, B., Unal, D. (2018). Comparison of Data Mining Algorithms Determining the Default Risk. Master's Thesis. Adana: Department of Statistics Çukurova University.

Dzelihodzic, A., & Donko, D. (2013). *Data Mining Techniques for Credit Risk Assessment Task*. Conference Paper. Valencia: Recent Advances in Computer Science and Applications.

Ecevit, Y. (2010). *Handbook on Gender Equality in the Labor Market*. Ankara: Pelin Ofset Tipo Matbaacilik San. ve Tic. Ltd. Sti.

Federal Financial Institutions Examination Council, (1982). *Summary of FFIEC Statements*, https://www.ffiec.gov/press_archives.htm.

Gan, C., Li, Z., Wang, W., Kao, B. (2012). Credit scoring in mortgage lending: evidence from China. *International Journal of Housing Markets and Analysis*. *5*(4), 334-350.

Goodman, L., Zhu, J., & Bai, B. (2016). *Women Are Better than Men at Paying Their Mortgages*. The United States: The Urban Institute.

Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Verlag, Berlin, Heidelberg: Springer.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Waltham: Morgan Kaufmann.

Horkko, M., (2010). *The Determinants of Default in Consumer Credit Market*. Finance Master's Thesis. Helsinki: Department of Accounting and Finance Aalto University School of Economics.

Hosmer, D.W., & Lemeshow, S. (2000). *Applied Logistic Regression*. Ohio: John Wiley & Sons Inc.

Huang, C., Chen, M., & Wang, C. (2006). Credit Scoring a Data Mining Approach Based on Support Vector Machines. *Expert Systems with Applications*, *33*, 847-856.

Hunt, J., & Kasynathan, N. (2002). Reflections on microfinance and women's empowerment. *Development Bulletin*. *57*, 71-75.

IABE, (2015). Big Data: An Actuarial perspective. Information Paper. Brussel: Institute of Actuaries in Belgium.

Inmon, W.H. (1991). *Building Data Warehouse*. New Work: Wiley Computer Publishing.

Karabiyik, I. (2012). Working Life of Women's Employment in Turkey. *Marmara University Journal of E.A.S*., *32*, 231-260.

Kumar, Y., Sahoo, G. (2012). Analysis of Bayes, Neural Network and Tree Classifier of Classification Technique in Data Mining using WEKA. *Computer Science & Information Technology, 5*, 359–369.

Okhovvat, M. (2015). Comparing various classification algorithms by WEKA. *Buletin Teknol. Tanaman, Bil, 12*,115-120.

Oren, K., Negiz, N., & Akman, E. (2012). Micro credit for women as a vehicle of struggle for poverty: A study on experiences. *Ataturk University Journal of Economics and Administrative Sciences*, *26*, 313- 338.

Ozen, U., (2017). *Management Information Systems*. Erzurum: Ataturk University Open Education Faculty.

Ozmen, F. (2012). Women's Employment And Micro Credit In Turkiye. *Suleyman Demirel University The Journal of Visionary*, *3*(6), 109-130.

Selimler, H. (2015). Analysis of Non-Performing Loans, The Evaluation Of Its Impact on The Financial Statements and Ratios Of Bank. *The Journal of Financial Researches and Studies*, *7*(12), 131-172.

Sousa, M.,M., & Figueiredo, R.S. (2014). Credit Analysis Using Data Mining: Application in the Case of a Credit Union. *Journal of* I*nformation Systems and Technology Management*, *11*(2), 379-396.

Van Gestel, T., & Baesens, B. (2008). *Credit Risk Management, Basic Consepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. Oxford: Oxford University Press.

Wellons, P.A. (1987). Passing the Buck: Banks, Government and Third World Dept. Harvard Business School Press.

Witten, I., Frank, E., & Hall, M. (2011). Data Minig Practical Machine Learning Tools and Techniques. San Francisco: Morgan Kaufmann.

Yeh, I., & Lien, C., (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, *36,* 2473-2480.