





ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)

URL: <https://dergipark.org.tr/ij3dptdi>

DATA MINING APPLICATION FOR FINANCIAL DECISION OPTIMIZATION AT RISK

Yazarlar (Authors): Enes Koçođlu , Filiz Ersöz *

Bu makaleye řu řekilde atıfta bulunabilirsiniz (To cite to this article): Koçođlu E., Ersöz F. "Data Mining Application For Financial Decision Optimization At Risk" *Int. J. of 3D Printing Tech. Dig. Ind.*, 5(2): 195-209, (2021).

DOI: 10.46519/ij3dptdi.950062

Arařtırma Makale/ Research Article

Eriřim Linki: (To link to this article): <https://dergipark.org.tr/en/pub/ij3dptdi/archive>

DATA MINING APPLICATION FOR FINANCIAL DECISION OPTIMIZATION AT RISK

Enes Koçođlu^a , Filiz Ersöz^a 

^aKarabuk University, Faculty of Engineering, Industrial Engineering Department, Karabuk, TURKEY

* Corresponding Author: fersoz@karabuk.edu.tr

(Received: 09.06.2021; Revised: 04.08.2021; Accepted: 11.08.2021)

ABSTRACT

Financial decisions can add value to the existence of businesses or individuals, as well as a wrong financial decision can cause businesses to cease to exist. Hence, financial decision or financial assumptions are vital for businesses or individuals. In financial assumptions, risk refers to the probability of losing as a result of an investment made in an asset. Measures can be taken against possible risks in the future through financial assumptions. In this study, the Logistic Regression Analysis (LR), one of the traditional methods, and the machine learning algorithm, Support Vector Machines (SVM) technique, which is one of the new approaches, are compared in the loaning process. It is aimed to determine the importance of the compared methods, the accuracy of the model, the estimation power of the model, the estimation performance of the model, the determination of the importance of the independent variables that affect the non-repayment of the loan, and the superiority of the techniques. According to the analysis results, the SVM technique is superior to the LR technique in calculating accuracy rate and prediction rate, and the LR technique is superior to the SVM technique in assumption performance calculation. The most significant variable in the SVM technique is "Lending policy", the most significant variable in the LR technique is "Interest rate", the second significant variable is "Interest rate" in the SVM technique, and "Lending Policy" as the second important variable in the LR technique. It is seen that the third most crucial variable in the two techniques is the "Income" variable. The determination of the SVM technique as the more important variable of the loan policy is deemed more suitable to the opinion of the banking expert. Detecting more realistic results of the SVM technique compared to the LR technique has shown the superiority of the SVM technique.

Keywords: Support vector machines, Logistic regression, Data mining, Decision making Under risk

1. INTRODUCTION

Uncertainties and decision-making will continue as long as life continues. A decision is any of the situations for predicting the future, choosing among alternatives or analysing the current situation with limited information support. The suitability or feasibility of the decision made is measured by statistical models [1]. The decision-making process includes three important criteria as action, result and state of nature. Action refers to the selection act made from the options. Result expresses the outcomes obtained in return for the chosen action. State of nature identify situations that the consequences of the chosen alternative may cause in the future. State of nature is the process in which risk is considered in the decision-making process. Risk can be calculated by means of probabilities [2]. The decision-making process is related to the continuity and success of the enterprises [3]. Hence, it is important for banks to make the right decisions for their survival.

Banks are responsible for allocating the deposits they collect from individual and legal persons to other individuals or legal entities in need of profit. Generally, bankers, who evaluate loan applications, make the evaluation according to subjective approaches. Such an approach produces inconsistency and inefficiency. Therefore, it is necessary to have measurable systems that can assist banks in their loan decisions. Giving loans to good customers will increase the bank's profitability. However, the bank's likely to lend to bad customers will result in the bankruptcy of the loan and a decrease in profitability

[4]. Being able to control and effectively manage credit risk is one of the main aims of banks. Hence, researchers aim to obtain meaningful results by using quantitative classification techniques to distinguish bank customers as good or bad [5].

The most significant variable to be measured in banking is the risk variable. The classification of the customers to be loaned is important for the management of the customers. Data mining used for classification enables banks to classify according to the financial behaviour of their customers [6]. Data mining is the process that enables making inferences that can be used in decision-making processes by making appropriate modelling from large amounts of data [7]. Data mining helps the decision maker by reducing time and effort in solving problems and making better decisions. Increasing the use of data mining techniques in institutions can provide meaningful and effective outcomes with the correct analysis of the system in many sectors [8]. Developments in the field of finance are significant in the development of the economy. The ability of banks to make efficient and correct decisions is an important variable of economic development.

The banking sector is one of the most essential sectors among all sectors. Banks are the most significant pillars of financial markets. Banks that show good performance with the right decisions accelerate economic development. There is evidence that shows poor performing banks inhibit economic growth, and cause poverty. The decision to lend is the most important financial risk taken by banks [9]. Risk types in the banking sector are given in Figure 1 below.

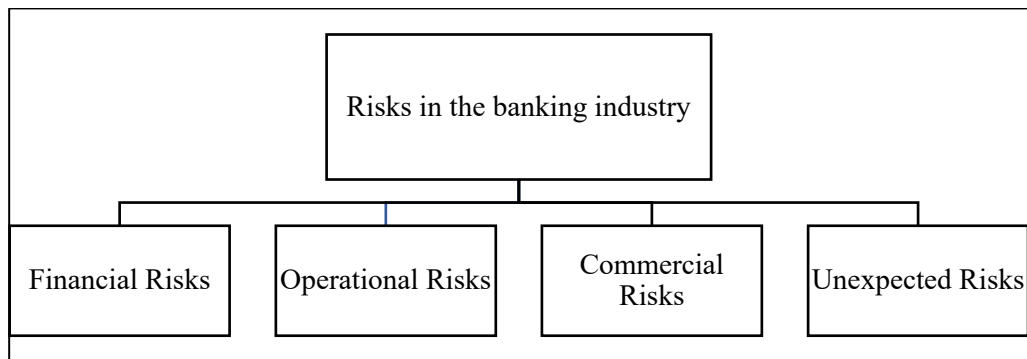


Figure 1. Types of risk in banking [10].

Financial risk in the sector can be defined as a warning or probability mechanism for banks, partners, creditors, company managers who are in contact with businesses or individuals to take financial-related decisions. Deciding how much credit limit a bank will allocate to a person or company involves financial risk assumption [11]. Financial risk is an essential variable controlled by the relevant institutions of the countries. The ability of banks to manage financial risk is important for the sustainability of their operations.

Operational risk includes the risks arising from the way the bank does business, which may be experienced in the process of providing a service such as a loan or deposit. For instance, operational risk is defined as the probability of repayment of the loan after obtaining incomplete signatures at the stage of granting a loan or entering credit risk before the guarantees of the loan are taken correctly. The reason for this risk is insufficient human resources or systematic failures.

Commercial risks are riskings that bank take for greater benefits in the course of continuing their operations. For instance, even if a contraction is foreseen in the construction sector in the country, the bank's desire to operate in this sector is an example of this type of risk. Unexpected risks in the sector are those that the bank management never thought of possible occurring. For example, as the risk of epidemics is not on the agenda of banks until 2020, it can be defined as an unexpected risk.

Credit limit allocation is an example of the risk that banks have the responsibility to manage. Managing this risk is the primary duty of banks [12]. Credit risk is one of the problems faced by banks, which can be defined as the possibility of non-repayment of the loan given by the bank.

Decreasing credit risk increases the success of banks and benefits the national economy. The ability of banks to classify their customers and define them as good or bad reduces the risk of marketing to good customers and the risk of non-repayment in bad customers [4]. The banking sector is an essential institution that balances, regulates and accelerates the economy. Considering that banks' capital is less than their total asset value, banks face bankruptcy risk if a certain part of the loans cannot be collected. Hence, customers should be classified to reduce the risk of non-collection [13].

Experts emphasize that the development of e-commerce will increase global trade and this will increase the use of e-banking, and the need for appropriate techniques, models and tools for banks to compete. As competition increases, banks' appetite for risk will increase. As a matter of fact, the risk is the reason for the existence of banking, therefore it is indispensable. Since it is not possible to eliminate the risk, it must be managed. Effective new tools are needed to manage risk [14]. Striking growth in consumer loans has reduced the impact of people-oriented approaches in evaluating customers who apply for loans. A loan decision should be made by considering the increasing number of applications and all the increasing variables [15].

In the traditional credit assessment approach, banks take into consideration various quantitative and subjective factors. In this approach, a reactive decision is made instead of a predictive one. This situation indicates the necessity of developing a quantitative and affirmable estimation approach. A machine learning approach can provide much better insights than a human decision maker. A learnable, risk-oriented technique that considers the changing habits of customers without rules is required for analysis [16].

In this study, the factors affecting the bank's lending and return process are compared with the logistic LR technique and the SVM technique. In comparing the established data mining models, the accuracy rate, the predictive power of the model and the predictive performance of the model are researched.

The remainder of this article is organized as follows. Section 2 presents literature studies on financial risk or credit rating with SVM and LR technique. Section 3 presents material and methods that the researcher used. Section 4 presents the findings of the analysis, Section 5 explains the results of the study.

2. LITERATURE REVIEW

Literature studies on financial risk or credit rating with SVM and LR techniques in the world and our country are given below.

Derelioğlu and Gürgen (2011) have made credit risk analysis comparison for 512 SMEs in Turkey with the use of multi-layered sensor-based neural networks. It has been determined that SVM are the technique that gives the most accurate result with the least data among the decision trees, SVM, factor analysis and principal component analysis methods [17]. The ability of the SVM method to produce accurate results with less data is important in choosing the method. According to Golbayani et al. (2020), credit rating institutions tend to request long-term data. The long wait reduces the opportunity to work. For this reason, credit scoring studies using artificial intelligence have attracted great interest in recent years. Successful machine learning programs can provide rapid analysis of credit scores. As a result of the research, it has been determined that neural networks and support vector machine methods provide superior performance compared to other methods [18].

In a study conducted by Ming-Wei et al. (2016), to compare traditional methods with machine learning methods, a comprehensive simulation analysis is performed with data obtained from 34 financial indices over a six-year period. According to the results, the best machine learning method has confirmed that it produces more accurate predictions than the best traditional statistical methods. In the study, it is determined that the best machine learning method used for financial forecasting is SVM, and the best traditional method is the autoregressive moving averages method [19]. SVM method, which is frequently seen in the literature, shows superior success compared to traditional methods.

Nawai and Shariff (2012) analyzed the factors affecting the repayment performance of the loan using the logistic regression model in Malaysian financial institutions and because of this, it is understood that the variables of gender, religious education, distance to the bank, monthly total income, total loan debt, delayed payment of the loan instalments affects the repayment performance of the debtor. Furthermore, they have found that the pressure to be exerted by banks to repay the loan do not affect the customer's loan repayment performance [20]. In their studies, Nazari and Alidadi (2013) have aimed to reveal the classification criteria to identify good and bad customers in Iranian banks. 18 variables are identified to identify good customers among 497 customers. The most important criteria of the results, the current loan debt and the frequency of withdrawal, the history of the relationship between the customer and the bank, and the variables that have the least important among the criteria for classifying the service received variables [4].

There are complex rules and restrictions applied by banks in the evaluation of the decision to provide a loan. The branches of the bank, which directly takes the lead in lending, should create a customer-specific payment evaluation development system by classifying the loan requests of their customers correctly to eliminate the difficulties. Unver et al. (2018) has developed a "Logistic Regression (LR) Model" to analyze 100 credit records applied in 2016. In the model, age, gender, profession, marital status, education, income, debt-income ratio and working time in the bank are defined as independent variables. Among the variables, the variables of marital status, gender, and the number of cross products is statistically significant, however, the variables of occupation, income and education are found to be insignificant. Also, variables such as debt-income ratio, credit card debt and other debts are seen as significant variables that negatively affect credibility [21].

Ersöz et al. (2016) analyzed individual and corporate customers who applied for a loan to a bank with data mining. The dependent variable is identified as several days with delayed loan payments, independent variables, on the other hand, is identified as loan amount, number of instalments, payments made to other banks, customer type and number of debtor banks. The aim of the study is the classification of customers as "not eligible for credit" or "eligible for credit" based on historical data. It is thought that the loan requests of the customers could be concluded faster thanks to the classification. According to the results the dependent variable, the delayed loan instalment, is positively and moderately correlated with the loan amount, negatively and moderately related to the number of instalments, and the payments made to other banks are weakly related, and the customer type is strongly negatively correlated. This determination shows that customers with large loans and payments to other banks are more likely to pay delayed instalments [22].

Plawiak et al. (2019) have stated that the factors that are considered in lending evaluations are such as profession, housing status, education, monthly income, age, marital status, region of residence, number of dependents, movements in bank accounts or history with the bank, position in which he is working or duration of existence in the company etc. [23]. Bellotti et al. (2011) have analyzed 511 banks from different countries using the SVM method to create a performance evaluation model. According to the results of the research, the DVM method can produce much better results than standard methods. It has been observed that the SVM method can be used to determine the financial status of countries and the method can be used by rating agencies, as well [24].

In the study conducted by Kara (2017), GARCH and SVM methods are used to develop an optimum profit prediction model for BIST 30 stock certificates. Possible predictions are used as variables in the model. It has been determined that the SVM method can generate higher returns in BIST 30 index compared to other methods [25]. Chen et al. (2018) has shown that the efficiency performances of Chinese banks are low by using data envelopment and SVM methods to determine the overall efficiency levels on 127 banks operating in China. It has been determined that banks' capital structure and costs are effective in the low efficiency of banks [26]. Hasan (2020) has aimed to develop a model that can predict the direction of the index after a period using the BIST 100 index. As a result of the analysis using SVM, logistic regression and random forest method, the SVM method has demonstrated superior success compared to other methods [27]. The success of the SVM method is that it enables the analysis of data belonging to many independent variables without restriction [33].

Korkmaz (2020) has shown that the SVM method is superior to the LR method in his study on the finance sector [28]. Bilik and Aydın (2019) have compared the factors affecting homeownership decisions with the DVM and LR technique. As a result of the analysis, the SVM method is superior to the LR method in terms of goodness of fit and predictive power [29]. In the literature, the techniques used in financial risk assessment and the independent variables that are used by banks in classifying their customers are analyzed. It has been observed that the use of the SVM technique is more among the technique used among machine learning techniques in classification problems and generally gives superior results compared to other methods in studies used. In his studies on credit risk in the banking sector, it has been observed that the LR technique, which is one of the traditional techniques, is used more than other traditional techniques. In this study, to make good or bad customer classification in the credit evaluation process, the advantages are analyzed by comparing the SVM and LR techniques.

3. MATERIALS AND METHODS

3.1. Research Data and Variables

The dependent variable and independent variables related to the data set constituting this study are determined according to the frequently used variables in the literature and expert opinion. The data that is used in the study is open to the public on the "LendingClub.com" website. The dataset represents 9,578 three-year credits between May 2007 and February 2010 [30]. Past bank relationship data of individual customers are used to determine the probability of non-repayment of individual loans. In this study, the logistic regression technique, one of the traditional techniques, the SVM technique, one of the innovative techniques, and the total data as training and test data are calculated as 80% Training and 20% Test data.

Table 1. Reference Variables

Variables	Variables used in the literature				
	(Finlay, 2017)	(Maldonado et al., 2017)	(Verbraken et al., 2014)	(Louzada et al., 2016)	(Jean Paul et al., 2020)
Credit policies			X		
Purpose of the loan	X	X		X	
Interest rate of the loan		X	X	X	
Instalment to be paid after loan	X				X
Documented income of the customer	X		X		
Debt divided by income	X	X	X		
Due date of ongoing loans	X	X	X	X	X
Unpaid loan balance	X	X	X	X	X

According to the studies in Table 1, there are independent variables affecting the non-repayment status of the loan. The fact that the independent variables used in this study are used in other studies in the literature is important for verifying the selection of the variables that make up the model. In this study, dependent and independent variables are defined below.

Dependent Variable (Y): Indicates the non-repayment status of the loan (the loan is not fully repaid).

Independent Variables: X_n

$$Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 \quad (1)$$

X_1 : Credit policy

X_2 : Purpose of the loan

X_3 : Interest rate of the loan

X_4 : Instalments to be paid after using the loan

- X₅**: Documented income of the customer
X₆: Debt divided by incomes
X₇: Maturity of the customer's ongoing loans
X₈: Customer's unpaid loan balance

Definitions of independent variables are given in the headings below.

Credit Policy: The credit policy is all the restrictions that banks inform their employees to determine the customers to be loaned. For instance, the decision of a bank not to give a loan to a person over the age of 65 can be seen as a decision not to give a loan to a person over the age of 75 for another bank. This difference is due to the bank's credit policy.

Purpose of the loan: The purpose of use of the loan indicates to which investment the loan will be transferred. For example, education, debt settlement, housing investment, etc. According to the statement of the customer, this criterion is taken into consideration in consumer loans. The customer's not telling the truth is the risk of this variable.

Interest Rate of Loan: The interest rate of the loan is the instrument by which the variable or fixed financing cost applied by the banks at the time of lending is calculated. The rate, which varies according to market conditions, depends on the borrowing ability of the person who will use the loan.

The instalment to be paid after using the loan: It refers to the total amount of instalments to be paid by the customer after the loan decision is assumed to be positive.

Documented income of the customer: After deducting the monthly fixed expenses of the customer and possible expenses from the total income, the remaining income must be able to cover the instalment amount to be formed in the loan. Documentation of the client's fixed income is important to predict the loan's ability to pay off.

The ratio resulting from dividing debts by incomes: The closer the loan income ratio to 1, the higher the risk of non-repayment of the loan. This rate is used to determine the borrowing measure, considering the fixed expenses and possible expenses in life.

Maturity of the customer's ongoing loans: The combination of the terms of the instalments to be paid by the customer with the instalment terms of the current debts is used to determine the criteria of the expenses that can be paid from the total income of the customer.

Customer unpaid loan balance: It is the total of outstanding loans at the same bank or other banks at the stage of the loan request of the customer.

3.2. Techniques Used in Research

3.2.1. Support Vector Machines (SVM)

The support vector machine SVM technique is data mining technology developed in 1995. This technique is a technique aimed at the statistical learning theory principle and minimum structural risk. The technique enables the development of a learning model by taking samples into account. This technique plays an important role in regression estimation, estimation time order, stock trend and pattern classification [35].

According to the studies in the literature, it has been proven that the SVM technique has a superior performance compared to other decision-making techniques [36]. After the studies of Lerner and Vapnik in 1963, Chervonenkis and Vapnik in 1964, Vapnik and Cortes have introduced the core (Kernel) indications in 1995, and the SVM technique began to be used as a superior technique in the analysis of non-separable data 32 years after the beginning of the approach [37].

SVM basically works as a linear separator between data to make two different multidimensional classifications. SVM is used to classify nonlinear data in large data sets without restriction. The aim of this technique is to maximize the distance between classes and to minimize the distance between hyperplane points in the classified region [38].

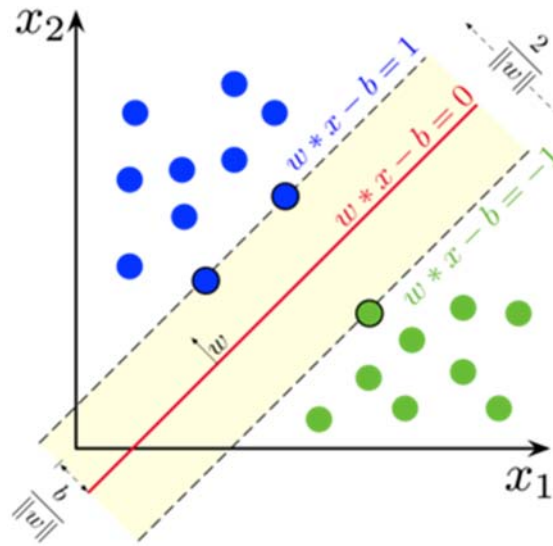


Figure 2. SVM Prospect [39]

SVM technique works with the aim of creating a hyperplane for classification. The balance that enables the separation of data by means of the hyperplane expressed [40].

$$w^T x + b = 0 \tag{2}$$

The balance that transforms the hyperplane to the optimum surface expressed

$$w_0^T + b_0 = 0 \tag{3}$$

Mathematical representation of distances on the optimal surface of X expressed;

$$x = x_p + r \frac{w_0}{\|w_0\|} \tag{4}$$

In this formula x_p , first x value is the optimal reciprocal of on the hyperplane. r is the optimum distance that is intended to be $g(x_p) = 0$

$$g(x) = w_0^T + b_0 = r\|w_0\| \text{ or } r = \frac{g(x)}{\|w_0\|} \tag{5}$$

$$r = \frac{g(x^{(s)})}{\|w_0\|} \begin{cases} \frac{1}{\|w_0\|} \text{ if } d^{(s)} = +1 \\ -\frac{1}{\|w_0\|} \text{ if } d^{(s)} = -1 \end{cases} \tag{6}$$

The optimum hyperplane defined by formula (1) is the optimum (w_0) It is possible to make the optimum classification between positive or negative samples with this algebraic process.

3.3. Logistic Regression (LR)

Logistic regression is the technique developed by statistician David Cox in 1958. The LR technique is used to estimate the probability of one or more independent variables explaining the dependent variable. This technique allows determining how an independent variable affects a dependent variable at a certain percentage. The difference is this technique; It can be used when the dependent variable has binary results such as good or bad [41].

LR, in cases where the independent variables take continuous or discontinuous values, is an important feature of the technique in that it allows the dependent variable to be categorical. The technique is highly preferred over traditional techniques due to its advantages such as the number of function parameters and its ability to produce durable results considering the assumption state [42].

Basically, the linear logistic regression function [43].

X = Independent Variable, Y = Dependent Variable

$$Y \text{ (Predicted)} = b_0 + b_1X$$

b_0 ve b_1 : Regression coefficients

The logistic regression model is as follows [44].

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots \dots \beta_mx_m \tag{7}$$

Descriptors of the model [45]:

π = Probability of the event,

β_i = Regression coefficient,

x_i = Independent variable (explanatory),

β_0 = Reference group and $x_1 \dots \dots x_m$ = observations showing the reference level.

3.4. ROC Curve

ROC (Receiver Operating Characteristic Curves) Curve is a graph that can visualize the performance of the binary classifier technique developed by Zweig and Campbell [46]. As the distance of the curve from the main diagonal increases in the ROC curve, the model performance also increases [33]. The better the model classifies, the closer the ROC curve gets to the upper left corner of the graph [47].

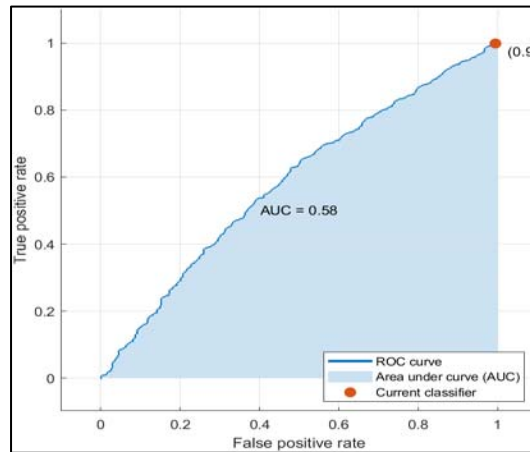


Figure 3. ROC Curve

The ROC curve shows "True positive (TP; y-axis)" vertically and "False positive" (x-axis) values horizontally. "True positive rate (TPR)" and "False positive rate (FPR)" are calculated as follows [48];

$$TPR = \frac{TP}{TP+FN}, FPR = \frac{FP}{TN+FP}$$

The ROC curve is drawn by combining TP and FP values. The ranges of values related to the ROC curve on the curve are as follows [49].

(0,0): Negative classification of all data,

(1,1): Positive classification of all data,

(0,1): Classification of all data in the right place (the most ideal result)

(1,0): Classification of all data in the wrong place (unintended result).

4. RESULTS

In this study, the LR technique, one of the traditional technique, and the SVM technique, which is one of the new approaches, are compared in the loaning process. It is aimed to research the ability of the compared techniques to determine the order of importance of the independent variables affecting the dependent variable, which is the accuracy rate calculation success of the model, the ability to reflect the estimation capabilities of the model, the estimation performance and the non-repayment of the loan, and the superiority of the techniques. Application IBM SPSS Modeler 18.0 data mining package program is used.

Table 2. Model Statistics

<i>LR Analysis Statistics</i>						
Model Summary						
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate		
1	.863	0.745	0.745	0.037196		
ANOVA						
	Model	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
1	Regression	32.633	7	4.662	3369.57	.000*
	Residual	11.158	8065	0.001		
<i>SVM Analysis Statistics</i>						
Model Summary						
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate		
2	.926	0.857	0.857	0.020435		
ANOVA						
	Model	Sum of Squares	df	Mean Square	F	Sig.
2	Regression	16.150	10.000	1.615	3867.346	.000*
	Residual	2.694	6450.000	0.000		

*p<0.05

In practice, firstly, the model statistics shown in Table 2 above are calculated for both techniques. The coefficient of determination for the logistic regression model (R^2) value 0.745; The coefficient of determination for the SVM model (R^2) is calculated as 0.857. This finding shows that the SVM technique has better model explanation ability compared to the LR model. Since the Significance is found to be significant in both techniques, it is seen that the model can be constructed in two techniques. LR and SVM techniques appear to be significant and significant ($p = 0.00 < 0.05$) [7]. According to these results, R^2 values show how much the independent variables used in the model explain the dependent variable.

Table 3. Model Accuracy Ratio Comparison.

<i>Analysis</i>	<i>Accuracy ratio</i>
<i>LR Analysis</i>	83.532%
<i>SVM Analysis</i>	83.718%

Accuracy rate is the rate at which the model is able to classify correctly while classifying the data. This ratio is in the "1" or "0" range. As the ratio approaches the value of "1", the classification success of the model increases [33]. The accuracy rate is indicative of the correct establishment of the classification model. According to the accuracy rates shown in Table 3, the accuracy rate of the SVM technique is higher than the accuracy rate of the LR technique. This determination reveals the success of the SVM technique.

Table 4. ROC Curve Comparison.

<i>Data</i>	<i>Training Data</i>	<i>Test Data</i>
<i>Logistics AUC Value</i>	0.645	0.631
<i>SVM AUC Value</i>	0.581	0.513

The area under the ROC curve (AUC: The Area Under the ROC Curve) is one of the important criteria showing the performance of the technique. The AUC value can be at most "1" [50]. Performance increases as the AUC Value approaches '1'. According to the comparative values shown in Table 4 above; The AUC value of the LR technique is 0.645, while the AUC value of the SVM technique is 0.581.

In classification studies, the accuracy and prediction rate of the model is expected to be high or close to each other, and it is thought that models with such outputs can make more accurate inferences in the long run [51]. As seen in Table 5, the SVM technique is found to be superior in the comparison of the prediction accuracy and success percentages of the techniques. As seen in Table 2, the SVM technique is found superior in the power of the techniques to explain the model.

Table 5. Prediction Accuracy Ratio.

<i>LR Analysis Non-Repayment of Expected Loan</i>				
<i>Observed Non-Repayment of Credit</i>	<i>Trained Data</i>	<i>Prediction Accuracy Percentage</i>	<i>Test Data</i>	<i>Prediction Accuracy Percentage</i>
<i>Right Estimation</i>	5397	83.53%	1330	82.51%
<i>False Estimation</i>	1064	16.47%	282	17.49%
<i>Total Estimation</i>	6461		1612	
<i>SVM Analysis Expected Loan Non-Repayment</i>				
<i>Observed Non-Repayment of Credit</i>	<i>Trained Data</i>	<i>Prediction Accuracy Percentage</i>	<i>Test Data</i>	<i>Prediction Accuracy Percentage</i>
<i>Right Estimation</i>	5409	83.72%	1335	82.82%
<i>False Estimation</i>	1052	16.28%	277	17.18%
<i>Total Estimation</i>	6461		1612	

In Table 5 above, it is seen that the false prediction rate of the LR technique is 17.49%, the successful prediction rate is 82.51%, the false prediction rate of the SVM technique is 17.18%, and the successful prediction rate is 82.82%. The fact that the possibility of non-repayment of the loan can be estimated

with a higher percentage of accuracy with the SVM technique shows the superiority of the SVM technique over the LR technique. The order of importance of variables in the LR technique is listed in Table 2 above, from strong to weak.

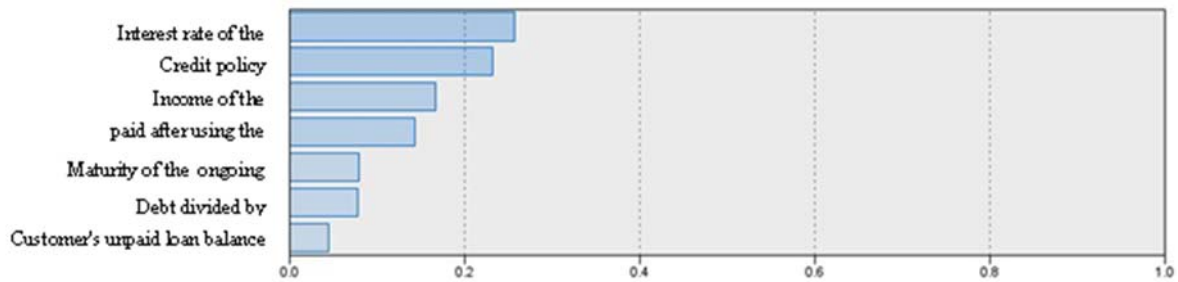


Figure 4. LR independent variable importance order

According to Figure 4, the order of importance of the variables affecting the non-repayment of the loan; interest rate, loan policy, income, instalments to be paid, current loan term, debt income ratio, available loan amount. The goal variable is found to be statistically insignificant in the LR technique. The insignificance of the objective variable is due to the weakness of its verifiability in loan applications. The customer tends to lie since s/he predicts that the probability of encountering a negative situation will increase when s/he says the purpose of using the loan is positive, and when another bank says to clear the debt. The order of importance of independent variables in SVM technique is listed in Figure 3 above from strong to weak.

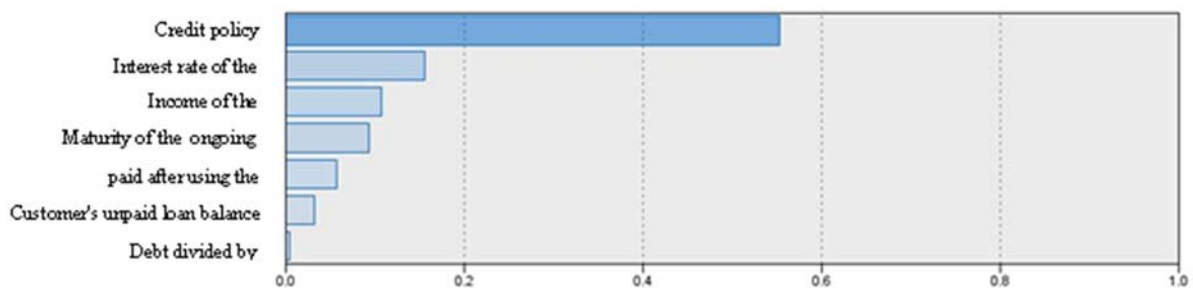


Figure 5. SVM argument importance order.

According to Figure 5, the order of importance of the variables affecting the non-repayment of the loan; loan policy, interest rate, income, available loan term, instalment amount to be paid, available loan amount, debt income ratio. The objective variable is found to be statistically insignificant in the SVM technique as in the LR technique. Difficulty in verifying the goal variable is the main reason for this determination. One of the strengths of both techniques is that the SVM and LR techniques can determine the order of significance of independent variables similar.

As a result of the analysis of the LR technique shown in Figure 2 and the SVM technique seen in Figure 3, the order of importance of the variables seems to be close to each other. While the interest rate in the LR technique is the most important variable affecting the non-repayment status of the customer, the most important variable in the SVM technique is the "loan policy". In the banking sector, when granting a loan, the financial benefit to be created by the loan to the customer is desired to be higher than the accepted cost. If the accepted cost is higher than the benefit to be gained, the customer is likely to be in financial trouble. That is why "interest rate" is significant.

Bank employees can take high interest rates and risks that should not be taken by ignoring the "credit policy". This problem arises from the performance-oriented sales pressure of the banking system and the employment of personnel. The earnings from the sale may be very insignificant compared to the loss in case the loan is not repaid. For these reasons, it is essential to emphasize the importance of credit bankruptcy rather than employing bank personnel with sales pressure.

In LR and SVM techniques, the third most important variable affecting the status of non-repayment of the loan is seen as "Income". If the benefit obtained in a decision to use a loan is negative at the beginning of the use, that loan should not be given by the bank. The purpose of the use of credit should not be to borrow more but to use the financing source that will create a lower cost than the expected benefit to obtain a benefit. For a loan to have a positive result, the "Income" variable is expected to be more than the "instalment amount to be paid" during the "term of the existing loans" and this ratio is defined as the "Debt income ratio". For these reasons, it has been evaluated that the SVM technique is superior to the LR technique in terms of more suitability according to the lending expert opinion of the banks.

5. CONCLUSION

In this study, the factors affecting the bank's lending and return process are compared with the LR technique and the SVM technique. In comparing the established data mining models, the accuracy rate, the predictive power of the model and the predictive performance of the model are researched. As a result of the analysis, it is seen that the SVM technique is superior to the LR technique in the calculation of "accuracy rate" and "successful prediction rate". According to the ROC curve values; it is seen that the LR model performance is superior to the SVM model performance, however, it is determined that it showed ideal performance in both techniques.

It has been revealed that the most important variable in the SVM technique is "Credit policy", and the most important variable in the LR technique is the "Interest rate". It has been determined that "Interest rate" is the second important variable in the SVM technique and "Credit policy" as the second important variable in the LR technique. It is seen that the third most crucial variable in the two techniques is the "Income" variable. The determination of the SVM technique as the more important variable of the loan policy is deemed more suitable to the opinion of the banking expert. The credit policy is in the form of guides prepared by banks for their personnel. Bank personnel can give a loan by ignoring the credit policies due to sales pressure and it is normal to experience problems in repayment of the loan in this situation where the general limits are neglected. If the policies are taken into consideration the decision to extend the loan, the interest rate will not be important. However, if the interest rate is taken into consideration, the credit policy is ignored. For this reason, the most important variable determined in the SVM technique is the "Credit policy" variable, it can be said according to scientific findings that it is more realistic.

According to the results of the research, the use of the SVM technique in banks' risk management, estimation analysis and testing the correct operability of the current model has been superior to the use of the LR technique. In the literature and this study, it has been observed that the prediction ability of the SVM technique is superior to traditional techniques. In this study, the "Loan interest rate" variable, which is determined as an important variable in the LR technique and SVM technique, is an important factor in the economic development of especially developing countries. The long-term forecast of interest rates is an important factor for countries to determine their long-term monetary and exchange rate policies. Hence, it is recommended to researchers to use the SVM technique in predicting future interest rates.

REFERENCES

1. Kökdemir, D., "Belirsizlik Durumlarında Karar Verme ve Problem Çözme", Doktora Tezi, [Decision Making and Problem Solving Under Uncertainty Situations][Thesis in Turkish], Ankara Üniversitesi, Ankara, 2003.
2. Koçi, E., "Riskli Ortamlarda Karar Verme", Yüksek Lisans Tezi, [Decision Making in Risky Environments] [Thesis in Turkish], Ankara Üniversitesi, Ankara, 2009.
3. Bilgütay, H., Deniz, M., "Risk Kültürünün Karar Verme Sürecine Etkisi" [The Effect of Risk Culture on the Decision - Making Process] [article in Turkish], Finans Ekonomi ve Sosyal Araştırmalar Dergisi, Vol 4, Issue 2, Pages 128 - 140, 2019.

4. Nazari, M., Alidadi, M., "Measuring Credit Risk of Bank Customers Using Artificial Neural Network", *Journal of Management Research*, Vol 5, Issue 2, 2013.
5. Maldonado, S., Bravo, C., Lopez, J., Perez, J., "Integrated Framework for Profit-Based Feature Selection and DVM Classification in Credit Scoring", *Decision Support Systems*, Vol 104, Pages 113-121, 2017.
6. Ersöz, F., "Dijitalleşme Çağında Büyük Veri ve Analitiđi: Sektörel Uygulamalar", 4th International Congress on 3D Printing Technologies and Digital Industry, 2019.
7. Ersöz, F., "Veri Madenciliđi Teknikleri ve Uygulamaları", Seçkin Yayıncılık, Ankara, 2019.
8. Ersöz, F., "Research Article Data Mining and Text Mining with Big Data: Review of Differences", *International journal of Recent Advances in Multidisciplinary Research*, Vol 6, no 1, Pages 4391- 4396, 2019.
9. Khojasteh, G., Karimzadeh, S., "Credit Risk Measurement of Trusted Customers Using Logistic Regression and Neural Networks", *Journal of System Management*, Issue 3, Pages 91-104, 2019.
10. Keramati, M. A., Shaeri, M., "Assessment of Credit Risk Management and Managerial Efficiency of Banks Using Data Envelopment Analysis (DEA) Network", *Biological Forum - An International Journal*, Vol 6, Issue 2, Pages 320-328, 2014.
11. Kapdan, F., Akta, M.G., "Durum Tabanlı Çıkarsama Yöntemi ile Finansal Risk Tahmini", *Innovations in Intelligent Systems and Applications Conference*, 2019.
12. Mandacı, P.E., "Türk Bankacılık Sektörünün Taşıdığı Riskler ve Finansal Krizi Aşmada Kullanılan Risk Ölçüm Teknikleri" [The Risks that the Turkish Banking Sector Faced and The Risk Measurement Models in Overcoming the Crisis] [article in Turkish], *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Vol 5, Issue 1, Pages 67-84, 2003.
13. Parvizi, R., Adibi, M.A., "Assessing and Validating Bank Customers Using Data Mining Algorithms for Loan Home", *International journal of Industrial Engineering and Operational Research*, Vol 2, Issue 1, 2020.
14. Asareh, B., Ghaeli, M.R., "Valuation and Assesment of Customers in Banking Industry Using Data Mining Techniques", *International Journal of Data and Network Science*, Vol 3, Issue 2, Pages 93-102, 2019.
15. Saia, R., Carta, S., Fenu, G., "A Wavelet-based Data Analysis to Credit Scoring", *ACM International Conference Proceeding Series*, 2018.
16. Kawa, D., Punyani, S., Nayak, P., Karkera, A., Jyotinagar, V., "Credit Risk Assessment from Combined Bank Records Using Federated Learning", *International Research Journal of Engineering and Technology*, Vol 6, Issue 4, Pages 1355-1358, 2019.
17. Dereliolu, G., Gürgen, F., "Knowledge Discovery Using Neural Approach for SME's Credit Risk Analysis Problem in Turkey", *Expert Systems with Applications*, Vol 38, Issue 8, Pages 9313-9318, 2011.
18. Golbayani, P., Florescu, I., Chatterjee, R., "A Comparative Study of Forecasting Corporate Credit Ratings Using Neural Networks, Support Vector Machines, and Decision Trees", *Nort American Journal of Economics and Finance*, Vol 54, 2020.
19. Hsu, M., W., Lessmann, S., Suang, M.C., Ma, T., Johnson, J.E.V., "Bridging the Divide in Financial Market Forecasting: Machine Learners vs. Financial Economists", *Expert Systems with Applications*, Vol 61, Pages 215-234, 2016.
20. Nawai, N., Shariff, M.N.M., "Factors Affecting Repayment Performance in Microfinance Programs in Malaysia", *Procedia - Social and Behavioral Sciences*, Vol 62, Pages 806-811, 2012.
21. Unver, M., Sahin, B., Ersöz, F., "An Application of Logistics Regression Model to Determining the Credit Suitability and Impacting Factors in a Special Bank Branch", Vol 12, Issue 1, Pages 1-12, 2018.

22. Ersöz, T., Ersöz, F., Özbilge, S., "Determination of the Bank's Customer Risk Profile: Data Mining Applications", Vol 10, Issue 6, Pages 2199-2203, 2016.
23. Plawiak, P., Abdar, M., Rajendra Acharya, U., "Application of New Deep Genetic Cascade Ensemble of SVM Classifiers to Predict the Australian Credit Scoring", Applied Soft Computing Journal, Vol 84, 2019.
24. Bellotti, T., Matousek, R., Stewart, C., "A Note Comparing Support Vector Machines and Ordered Choice Models' Predictions of International Banks' Ratings", Decision Support Systems, Vol 51, Issue 3, Pages 682-687, 2011.
25. Kara, M., "Yapay Zeka Modeliyle Genişletilmiş Hibrit Black - Litterman Model Önerisi, Borsa İstanbul BIST-30 Endeks Verileri ile Test Edilmesi", Doktora Tezi, [Hybrid - Litterman Model Proposal Extended with Artificial Intelligence Model, Testing with Borsa İstanbul BIST-30 Index Data] [Thesis in Turkish], Hacettepe Üniversitesi, Ankara, 2017.
26. Chen, Z., Matousek, R., Wanke, P., "Chinese Bank Efficiency During The Global Financial Crisis: A Combined Approach Using Satisficing DEA and Support Vector Machines", North American Journal of Economics and Finance, Vol 43, Pages 71-86, 2018.
27. Hasan, A., "Derin Öğrenme ve Makine Öğrenmesi Yöntemleriyle Borsa Alım Satım Davranışlarının Modellenmesi", Doktora Tezi, [Modelling Stock Exchange Trading Behaviors with Deep Learning and Machine Learning Methods] [Thesis in Turkish], Yıldız Teknik Üniversitesi, İstanbul, 2020.
28. Korkmaz, G., "Yapay Zekâ Yöntemleriyle Sınıflandırma ve Finans Sektöründe Kurumsal Müşterilere Dönük Bir Uygulama" [Classification with Artificial Intelligence Methods and An Application For Corporate Customers in the Finance Sector] [article in Turkish], Akademik Yaklaşımlar Dergisi, Vol 11, Issue 2, Pages 91-109, 2020.
29. Bilik, M., Aydın, Ü., "Konut Sahibi Olma Kararlarını Etkileyen Faktörler: Lojistik Regresyon ve Destek Vektör Makinelerinin Karşılaştırılması" [Factors Affecting The Housing Demand: A Comparison of Logistics Regression and Support Vector Machines] [article in Turkish], Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, Vol 62, Pages 184-199, 2019.
30. Demirci, İ., "Lending-loans", Retrived from <https://www.kaggle.com/izemdemirci/lendingloans>, January 2, 2021.
31. Finlay, S.M., "Predictive Models of Expenditure and Over - Indebtedness for Assessing the Affordability of New Consumer Credit Applications", Journal of the Operational Research Society, Vol 57, Issue 6, Pages 655-669, 2006.
32. Verbraeken, T., Bravo, C., Weber, R., Baesens, B., "Development and Application of Consumer Credit Scoring Models Using Profit-Based Classification Measures", European Journal of Operational Research, Vol 238, Issue 2, Pages 505-513, 2014.
33. Louzada, F., Ara, A., Fernandes, G.B., "Classification Methods Applied to Credit Scoring: Systematic Review and Overall Comparison", Surveys in Operations Research and Management Science, Vol 21, Issue 2, Pages 117-134, 2016.
34. Jean Paul, B., Loezer, L., Enembreck, F., Lanzuolo, R., "Lessons Learned From Data Stream Classification Applied to Credit Scoring", Expert Systems With Applications, Vol 162, 2020.
35. Yang, R., Yu, L., Zhao, Y., Yu, H., Xu, G., Wu, Y., Liu, Z., "Big Data Analytics For Financial Market Volatility Forecast Based on Support Vector Machine", International Journal of Information Management, Vol 50, Pages 452-462, 2020.
36. Ayhan, S., "Kaba Küme ve Destek Vektör Makineleri Kullanılarak Nitelik İndirgeme ve Sınıflandırma Problemlerinin Çözümü için Bütünleşik Bir Yaklaşım", Doktora Tezi, [An Integrated Approach for Solving Attribute Reduction and Classification Problems Using Rough Set and Support Vector Machines] [Thesis in Turkish], Eskişehir Osmangazi Üniversitesi, Eskişehir, 2013.

37. Altıntop, M.Y., "Dalgacık Dönüşümü ve Destek Vektör Makineleri ile Tahmin: Bist Üzerine Bir Uygulama", Doktora Tezi, [Estimation with Wavelet Transform and Support Vector Machines: An Application on Bust] [Thesis in Turkish], Uşak Üniversitesi, Uşak, 2016.
38. Janardhanan, P., Heena, L., Sabika, F., "Effectiveness of Support Vector Machines in Medical Data Mining", Journal of Communications Software and Systems, Vol 11, Issue 1, Pages 25-30, 2015.
39. https://en.wikipedia.org/wiki/Support-vector_machine, 18 Nisan, 2021.
40. Haykin, S., "Neural Networks, a Comprehensive Foundation", Prentice Hall, Englewood Cliffs, New Jersey, 2001.
41. Jote, G.G., "Determinants of Loan Repayment: The Case of Microfinance Institutions in Gedeo Zone, SNNPRS, Ethiopia ", Universal Journal of Accounting and Finance, Vol 6, Issue 3, Pages 108-122, 2018.
42. Cankurt, M., Miran, B., Ahmet, Ş., "Sığır Eti Tercihlerini Etkileyen Faktörlerin Belirlenmesi Üzerine Bir Araştırma: İzmir ili örneđi" [Determining of the Effective Factors on Cattle Meat Preferences: The Case of İzmir], [article in Turkish], Hayvansal Üretim, Vol 51, Issue 2, Pages 16-22, 2010.
43. Shannon, M.D., Davenport, A.M., "Using SPSS to Solve Statistical Problems: A self-Instruction Guide", Prentice Hall, ABD, 2001.
44. Kalaycı, Ş., "SPSS Uygulamalı Çok Deđişkenli İstatistik Teknikleri", Asil Yayınlar, Ankara, 2009.
45. Sperandei, S., "Understanding Logistic Regression Analysis", Biochemia Medica, Vol 24, Issue 1, Pages 8-12, 2014.
46. Zweig, M.H., Campbell, G., "Receiver-Operating Characteristic (ROC) Plots: a Fundamental Evaluation Tool in Clinical Medicine", Clinical Chemistry, Vol 39, 1993.
47. Verbakel, J. Y., Steyerberg, E. W., Uno, H., De Cock, B., Wynants, L., Collins, G. S., Van Calster, B., "ROC Curves for Clinical Prediction Models Series", Journal of Clinical Epidemiology, Vol 126, Pages 207-216, 2020.
48. Witten, I. H., Frank, E., "Credibility: Evaluating What's Been Learned, Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann Publishers, San Francisco, 2005.
49. Dondurmacı, G., "Veri Madenciliđi'nde Regresyon Ağaçları ile Sınıflandırma ve Bir Uygulama", Doktora Tezi, [Classification with Regression Trees and an Application in Data Mining] [Thesis in Turkish], Mimar Sinan Güzel Sanatlar Üniversitesi, İstanbul, 2011.
50. Yürük, M.F., Ekşi, İ.H., "Yapay Zeka Yöntemleri ile İşletmelerin Finansal Başarısızlıđının Tahmin Edilmesi: BİST İmalat Sektörü Uygulaması"[Financial Failure Prediction of Companies Using Artificial Intelligence Methods: An Application in BIST Manufacturing Sector] [article in Turkish], Mukaddime, Vol 10, Issue 1, Pages 393 - 422, 2019.
51. Tayyar, N., Tekin, S., "İMKB-100 Endeksinin Destek Vektör Makineleri ile Günlük, Haftalık ve Aylık Veriler Kullanarak Tahmin Edilmesi" [Forecasting ISE-100 Index Using Support Vector Machines with Daily, Weekly and Monthly Data] [article in Turkish], AİBÜ Sosyal Bilimler Enstitüsü Dergisi, Vol 13, Issue 1, Pages 189 - 217, 2013.