Determining the factors for individual credit approval by applying logistic regression and hierarchical logistic regression

Ayse YILDIZ[1]

1 Business Administration, Ankara Haci Bayram Veli University, Ankara, Türkiye

DOI: https://doi.org/10.56293/IJMSSSR.2023.4705

IJMSSSR 2023
VOLUME 5
ISSUE 6 NOVEMBER - DECEMBER **ISSN: 2582 - 0265**

Abstract: There has been a rapid increase in applications made to lending institutions due to the inadequacy of individuals' savings to meet their growing demands and needs. However, specific criteria are expected from individuals for the approval of these applications. Various methods have been developed to accurately determine these criteria and facilitate the loan approval process efficiently and promptly. Despite being the most used method, binary logistic regression has its limitations. These include the simultaneous inclusion of all variables in the analysis, the oversight of control variables that could be determined by decision-makers, and the evaluation of each event within the group of the same variables. In an attempt to address these limitations, hierarchical logistic regression analysis was performed alongside binary logistic regression analysis. The study incorporated variables indicating individual characteristics, such as age, socio-demographic characteristics like occupation and account status, and loan attributes, including duration and installment rate. The study obtained 1000 loan applicants, and the models were evaluated based on the goodness of fit results and the explanatory power of the models. Although the study produced similar results in determining the influential variables and evaluating the models, it highlighted that hierarchical logistic regression provides more flexibility to decision-makers by allowing the evaluation of nested models formed by sub-variable groups separately. This approach can offer valuable insights to decision-makers in lending institutions, as it allows for the separate evaluation of various sub-variable groups, contributing to a comprehensive understanding of the key determinants of loan approval. By prioritizing specific variables or groups of variables, this methodology facilitates a more comprehensive analysis of the complex relationships between different factors, leading to more informed and data-driven decision-making. It is crucial, however, to ensure the appropriate application of the methodology and the continuous validation and testing of the hierarchical logistic regression models with new datasets to maintain their effectiveness and relevance in the dynamic lending landscape.

Keywords: Binary logistic regression, Hierarchical logistic regression, Individual credit approval

1. Introduction

The demands and needs of people are increasing substantially every day, leading to a notable rise in the number of applications to lending institutions, particularly banks. However, it is not feasible to approve all these applications. When a bank receives a loan application, it must assess the applicant's profile to determine whether to approve the loan. This decision carries two types of risks:

1. Approving a loan for a good credit risk ensures business for the bank. Rejecting the loan application in such cases leads to a loss for the bank.
2. Approving a loan for a bad credit risk results in a financial loss for the bank.

The primary objective of the analysis is to minimize risk and maximize profit on behalf of the bank. To minimize losses from the bank's perspective, the bank requires a decision rule for approving or rejecting loan applications. Loan managers consider an applicant's demographic and socio-demographic profiles before deciding on their loan applications (https://online.stat.psu.edu/stat857/node/215/).

Credit evaluation models are frequently utilized in this approval process. These models aim to minimize credit risk, facilitate quicker and more effective decision-making, and establish valid evaluation criteria (variables). Even a slight enhancement in achieving these goals is known to significantly reduce risk and enable a substantial amount of money to be obtained (Tsai & Chen, 2010). Hence, numerous studies, both in literature and practice, have contributed to this process.

Various methods are employed in these studies, which are continuously evolving as none of them yield all the desired results for every dataset. One of the most common methods applied in this context is discriminant analysis (Myers & Forgy, 1963; Steenackers & Goovaerts, 1989; Mylonakis & Diacogiannis, 2010). However, it is not an appropriate analysis because it does not account for the assumption of normality. This is primarily due to many variables in the credit evaluation process, such as marital status and profession, being nominal variables.

With the advancements in computer technology, it has been observed that machine learning-based techniques, such as artificial neural networks, decision trees, and support vector machines, are widely used. Artificial neural networks (West, 2000; Sönmez, 2015) along with support vector machines (Abdou, Pointon, & El-Masry, 2008; Aphale & Shinde, 2020; Kadam et al. 2021; Kibria & Sevkli, 2021) are used to and compared performance of these techniques to aid in this process. Some studies have been carried out combination with fuzzy logic, such as neuro fuzzy, fuzzy support vector machines, and fuzzy k-nearest neighbor algorithm, and fuzzy logistic regression techniques (Lessman et al., 2015; Tsai & Chen, 2010; Abdou, Pointon & El-Masry, 2008; Thomas et al., 2002). Although these methods have been successful in making accurate estimations, they are insufficient in providing the necessary information for decision-making. Neural networks and support vector machines have shown success in classifying the datasets used, but they struggle to manage interactions that arise with changes in the population. Moreover, the processes of these methods in handling inputs are often described as a "black box," making their results challenging to interpret. As a result, lending institutions cannot explore rejections based on the outcomes of these methods (Dong et al., 2010). Furthermore, the majority of studies employing machine learning techniques are centered on model comparisons, neglecting the crucial issue for decision-making, which is the selection of appropriate inputs rather than the models themselves. Accurate and valid models obtained with inappropriate inputs raise doubts about their reliability. Logistic regression, another extensively used method in the literature (Sarlija et al., 2004; Tabagari, 2015; Torosyan, 2017; Kwofi, Owusu-Ansah, & Boadi, 2015; Chen et al., 2018; Al-Aradi, 2014; Dong et al., 2010; Bolton, 2009), has gained preference for several reasons. Firstly, apart from the challenges of missing data and multicollinearity, it imposes minimal assumptions on the variables (Bolton, 2009), allowing for the use of continuous or nominal independent variables. Secondly, it provides outcomes that are easily and understandably interpretable. Thirdly, it could generate linear scorecards (Chen et al., 2018). However, this method also has limitations in determining the critical input variables for decision-making and identifying groups with similar characteristics.

A proposed models applying binary and hierarchical logistic regression methods have been developed for decision-makers to identify the key variables that influence credit scoring of the applicants. Consequently, the study's contribution lies in its practical implications, rather than theoretical considerations. Given this framework, the study comprises four sections. The second section provides methodology information. The section includes insights into the binary logistic regression and hierarchical logistic regression methods, along with their relevant attributes and dataset included. This theoretical foundation serves to underpin the subsequent analytical discussions. The third section delves into the analysis results, accompanied by their pertinent interpretations. This critical section likely outlines the key findings derived from the application of the binary logistic regression and hierarchical logistic regression methods to the dataset, shedding light on the factors influencing the loan approval process. Lastly, the fourth section entails general assessments and proposes potential avenues for further research. This section serves to provide a holistic perspective on the implications of the study's findings, offering valuable insights for both academic research and practical applications in the field of credit evaluation and lending practices.

2. Methodology

2.1. Related attributes with dataset

In assessing the creditworthiness of individuals, multiple variables are employed across various methods. While some variables may pertain to the financial status of the consumer, such as the possession of a checking account, credit card ownership, and the duration of the relationship with the current bank, other variables might reflect the stability of the consumer's life, such as the length of residence in the same house and the duration of employment in the same job. Moreover, there are additional variables that provide insights into the consumer's resources, including residential status, employment status, and the employment status of the spouse. Therefore, any variable that aids in the prediction process can be integrated into the model to facilitate decisions concerning the credit score of individuals. However, it is crucial to note that certain characteristics, such as race, religion, and gender are illegal to use in credit-scoring systems (Thomas et al., 2002, p. 5).

As previously mentioned, various variables can be employed in credit evaluation. These predictors can encompass the loan amount, as well as personal information such as the individual's annual income, occupation, any previous defaults, outstanding debts, and credit history Owing to privacy concerns and the unavailability of data from banks, this study was conducted using the German Credit dataset, available in the UCI machine-learning data archive The German Credit Data set comprises information on 20 variables and classifies each of the 1000 loan applicants as either a Good or a Bad credit risk. The development of a predictive model using this dataset aims to offer valuable guidance to bank managers, assisting them in determining whether to approve a loan for a prospective applicant based on their profiles (https://online.stat.psu.edu/stat857/node/215/). The dataset is detailed in Table 1 below.

Table 1 Variables with Definitions

Variable	Variable Scale	Categories
Credit Amount	Numerical value	
Age	Numerical value	
Duration	Numerical value	
Installment rate	Numerical value	
Residence length	Numerical value	
Length employed	Categorical Variable	
	1: Less than 1 year	1
	2: 1-4 years	1
	3: 4-7 years	1
	4: Greater than 7 years	1
	5: Unemployment	*0*
Account status	Categorical Variable	
	1: No debt history	1
	2: No current debt	1
	3: Payments current	1
	4: Payments delayed	1
	5*: Critical account*	*0*
Housing	Categorical Variable	
	1: Rent	1
	2: Own	1
	3: *Free*	*0*
Credit approval (dependent variable)	Categorical Variable	
	1: No	1
	0: Yes	*0*

As indicated in Table 1, the variables such as age, duration, installment rate, and residence length are identified as numeric variables, while length employed, account status, and housing, including the dependent variable for loan approval, are classified as categorical variables. Within the categorical variables, italicized indicator variables are utilized for comparisons with other levels of the variable. According to Thomas et al. (2002), certain characteristics such as marital status and gender, which are deemed illegal, have not been included in the analysis.

Preliminary data inspections revealed that continuous variables such as age and credit amount have a positive skewness, necessitating the use of methods like logistic regression that do not meet the normal distribution assumption (https://online.stat.psu.edu/stat857/node/216/).

2.2. Method

2.2.1. Binary logistic regression

Binary Logistic regression seeks to establish the non-linear relationship between the dichotomous dependent variable and independent variables, which may comprise nominal and/or continuous variables. Additionally, it endeavors to predict the probability of the occurrence of the dependent variable, which assumes values of 0 and 1 within the model. In our study, this model was selected due to the fact that the desired outcome, namely the "default status," has two potential results: 0 and 1.

Logistic regression, differing from linear regression models, is centered on a dichotomous dependent variable and probability estimation. It is grounded on two pivotal concepts, namely odds and logit. Odds, represented in Equation 1, signifies the ratio of the probability of an event occurring to the probability of it not occurring.

$$Odds = \frac{P_{(x)}}{1-P(x)} \tag{1}$$

When the odds ratio surpasses 1, the likelihood of the event increases; if it equals 1, the likelihood remains the same; and if it falls below 1, the likelihood decreases. To ensure that the odds are confined between 0 and 1, the logit, the logarithm of odds, is calculated. Logit plays a vital role in establishing a link with the linear regression model, as demonstrated in Equation 2:

$$Logit\ p(x) = ln[\frac{P}{1-p}] = \beta_1 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots \beta_p x_p \tag{2}$$

This equation reveals that a negative logit value corresponds to odds less than 1, while a positive logit value corresponds to odds greater than 1. The logistic regression equation is formulated in Equation 3, illustrating the relationship between the explanatory variables and the probability of the dependent variable:

$$P(Y = 1) = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{-(\beta_0 + \beta_1 X)}} \tag{3}$$

As apparent from the equation, due to the non-linear relationship between the dependent variable and the independent (explanatory) variables, the interpretation of the regression model's coefficients should be based on odds rather than direct parameters. The odds of the dependent variable represent the exponential function of the linear regression. Hence, the logistic coefficients offer insights into the direction of the relationship, while the exponential logistic or odds value indicates the magnitude of change in the likelihood of the dependent variable resulting from a unit change in the independent variable.

Consequently, a positive change in an explanatory variable, denoted by an exponential coefficient greater than 1, would likely lead to an increase in the odds value. In other words, the model would reflect a higher likelihood of the event's realization, indicated by P(Y=1). Conversely, if the coefficient is negative, resulting in an exponential coefficient less than 1, the odds value would probably decrease, leading the model to assign a lower likelihood value for the realization of P(Y=1).

Post regression analysis, the developed model must undergo rigorous testing and validation to ensure its suitability for use in the decision-making process. To this end, various tests and measures are employed, including the Omnibus test of the model coefficients, the Hosmer-Lemeshow goodness-of-fit test, measures of explained variance such as Pseudo R^2 values, including Cox & Snell R^2 and Nagelkerke R^2, and the assessment of classification success results.

2.2.2. Hierarchical logistic regression

In conventional regression models, cases are typically treated as independent of one another. In contrast, hierarchical regression models focus on the selection and inclusion of predictors within the model. This model is particularly suitable when the data exhibits a nested structure among cases. It serves as a framework for assessing whether the variables of interest explain a statistically significant amount of variance in the dependent variable while controlling for other variables. This framework facilitates model comparisons, extending beyond a mere statistical method.

In the hierarchical model, multiple logistic regression models are constructed by progressively adding variables to the model in steps referred to as "blocks," with subsequent models always including the smaller models constructed in the preceding steps. The emphasis often lies in evaluating the improvement in R^2, which signifies the proportion of explained variance in the dependent variable upon the addition of the new variable to the model (data.library.virginia.edu).

It is crucial to underscore that stepwise logistic regression is not the preferred approach, as it selects decision variables solely based on statistical criteria, without necessarily considering the relevance of these variables to decision-makers. Given the specific objectives of this study, which aim to identify decision variables that are meaningful for decision-makers, the use of stepwise logistic regression could potentially hinder the attainment of these goals.

3. Analysis Results

3.1. Binary logistic regression analysis results

This research employed binary logistic regression analysis, with the dependent variable set to "0" denoting non-approval for a loan and "1" denoting approval for a loan. The independent variables considered in the analysis included age, duration in months, installment rate, account status, length employed, and housing. The findings of the binary logistic regression analysis conducted using these variables are presented in Table 2.

Table 2 Binary Logistic Regression Results for Credit Approval of Individuals

Variables	bj	$S(b_j)$	Wald	df	P	Odds Ratio (b_j)
Age in Years	.008	.023	.108	1	.742	1.008
Duration in months	-.032	.020	2.515	1	.113	.969
Installment rate	**-.448**	**.209**	**4.616**	**1**	**.032**	**.639**
Account status			11.040	4	.026	
Account status (1)	.551	.947	.338	1	.561	1.734
Account status (2)	**3.148**	**1.144**	**7.578**	**1**	**.006**	**23.295**
Account status (3)	-.353	.509	.482	1	.487	.702
Account status (4)	.754	.694	1.181	1	.277	2.126
Length employed			5.281	4	.260	
Length employed (1)	.038	1.609	.001	1	.981	1.038
Length employed (2)	1.276	1.264	1.0184	1	.313	3.582
Length employed (3)	1.997	1.280	2.433	1	.119	7.365
Length employed (4)	1.597	1.253	1.624	1	.203	4.937
Residence length	.094	.244	.149	1	.699	1.099

Housing			3.690	2	.158	
Housing (1)	-1.324	.877	2.282	1	.131	.266
Housing (2)	*-1.314*	*.694*	*3.582*	*1*	*.061*	*.270*
Constant	-.660	1.959	.113	1	.736	.517

Table 2 highlights two key variables that significantly influence the evaluation of an individual's credit or loan approval. Firstly, the installment rate demonstrates a negative relationship with credit approval. The odds ratio for this variable is 0.639, indicating that, with the control of other variables, a one-unit change in the installment rate results in 0.639 times decrease in the probability of credit approval. Secondly, the account status 2 category, pertaining to the account status, shows a positive correlation with credit default. The odds ratio for this category is 23, suggesting that an individual with no current debt is 23 times more likely to receive a positive response compared to an individual with a critical account status.

Following the completion of the regression analysis and the acquisition of the regression results, it was imperative to evaluate and confirm their applicability for decision-making. To achieve this goal, several critical assessments were conducted, including the Omnibus test of the model coefficients, the Hosmer-Lemeshow goodness-of-fit test, measures of explained variance such as Pseudo R^2 values e.g., Cox & Snell R^2 and Nagelkerke R^2 and the results of the classification success analysis. These comprehensive results are presented in Table 3.

Table 3 Evaluation of the Model

Tests	χ^2	df	Value	p
Omnibus test	33.465	14		.007
Hosmer & Lemeshow Test	6.729	8		.566
Cox & Snell R^2			.138	
Nagelkerke R^2			.238	
Classification Success				86.8%

Based on the results presented in Table 3, the Omnibus test of model coefficients indicates a notable improvement in model accuracy compared to the baseline model, which only includes the intercept variable (χ^2= 33.465, df = 14, $p < .05$). Moreover, the Hosmer & Lemeshow goodness-of-fit test is not statistically significant, suggesting that the model fits the data well (χ^2= 6.729, $p > .05$). The R^2 values, representing the proportion of variation in the dependent variable explained by the identified variables, are calculated as 0.138 for Cox & Snell R^2 and 0.238 for Nagelkerke R^2 respectively. These values indicate that the selected variables account for approximately 13% to 23% of the variation in credit defaults among the applicants.

Furthermore, the classification success value demonstrates that the model exhibits an impressive 86.8% success rate in correctly classifying the applicants. Taken together, these analysis results confirm the overall validity and robust classification capability of the developed model. Consequently, the obtained results can serve as a reliable basis for making informed decisions regarding the approval or rejection of credit applications.

3.2. Hierarhical logistic regression analysis results

While binary logistic regression offers valuable insights for decision-makers, it may not be the most appropriate method for determining the crucial variables that significantly impact the dependent variable. Through hierarchical logistic regression analysis, lenders can build models based on their past experiences, taking into account which variables are influential and should be considered together to enhance decision-making accuracy in loan approvals.

In this study, the first model was developed, incorporating only the "age" variable, representing the personal characteristics of the individuals. Subsequently, the second model was constructed by including "length employed," "housing," and "residence length" variables, reflecting the socio-economic status of the individuals. The third model encompassed "duration in months" and "installment rate," focusing on the credit characteristics, while the fourth model introduced the "account status" variable as an indicator of financial solvency.

The regression model coefficients and the explanatory power of these models, along with the classification success, are presented in Table 4. This comprehensive analysis enables a deeper understanding of the interplay between various variables and their impact on the decision-making process for loan approvals.

Table 4 Hierarchical Regression Analysis Results for Different Regression Models

Predictors	Regression Model 1	Regression Model 2	Regression Model 3	Regression Model 4
Age	**1.032***			
Cox & Snell R^2	.021			
Nagelkerke R^2	.037			
Class. Rate	84.4			
Age		1.017		
Length employed				
Length employed (1)		.624		
Length employed (2)		2.575		
Length employed (3)		4.414		
Length employed (4)		2.590		
Housing				
Housing (1)		.343		
Housing (2)		.396		
Residence Length		1.073		
Cox & Snell R^2		.060		
ΔR^2		.039		
Nagelkerke R^2		.104		
ΔR^2		.67		
Class. Rate		84.4		
ΔClass. Rate		0		
Age			1.012	
Length employed				
Length employed (1)			.745	
Length employed (2)			2.742	
Length employed (3)			5.583	
Length employed (4)			3.647	
Housing				
Housing (1)			.279	
Housing (2)			.353	
Residence Length			1.136	
Duration in months			.980	
Installment rate			**.689***	
Cox & Snell R^2			.085	
ΔR^2			.025	
Nagelkerke R^2			.147	
ΔR^2			.43	
Class. Rate			84.9	
ΔClass. Rate			0.5	
Age				1.008

Length employed	
Length employed (1)	1.038
Length employed (2)	3.582
Length employed (3)	7.365
Length employed (4)	4,937
Housing	
Housing (1)	.266
Housing (2)	.270
Residence Length	1099
Duration in months	.969
Installment rate	
Account status	**.639***
Account status (1)	
Account status (2)	1.734
Account status (3)	**23.295***
Account status (4)	.702
	2.216
Cox & Snell R^2	
ΔR^2	.138
Nagelkerke R^2	0.45
ΔR^2	.238
Class. Rate	0.91
ΔClass. Rate	1.9
	86.8

The ΔR^2 and classification values in Table 4 provide insights into the variations in the coefficient of determination and the changes in the classification percentage across successive models. Analysis of these values reveals the significance of the "installment rate" and the second level of the "account status" variable (indicating no current debt) in the fourth model were statistically significant. Notably, in contrast to the findings of the binary logistic regression analysis, the variable "age" was found to be significant in the first model where it is the only variable included.

A particularly noteworthy observation is the substantial change observed in the fourth model, particularly with respect to the second level of the "account status" variable. This result is not surprising, given the statistical significance of this variable indicated by the binary logistic regression analysis where all variables were included. Consequently, within the context of the model under consideration, it can be inferred that this variable should be accorded primary consideration in determining the default status of the applicant.

4. Conclusion

Numerous factors significantly impact the determination of credit approvals for individuals. Therefore, the identification of these influential factors enables lenders to expedite their decision-making processes while ensuring greater effectiveness and accuracy. Given the variability observed in the literature, with various techniques being employed in this field and new methodologies continually under development, the need for reliable and robust techniques remains a critical concern. In this study, various quantitative techniques were employed to ascertain the key factors influencing credit scoring decisions. The variables "age," "installment rate," and the second level of the "account status" (indicating no current debt) emerged as reliable predictors of credit application approval. Consequently, this study contributes by proposing a more effective model for credit approval decisions, enhancing the decision-making process for lenders and financial institutions.

As an extension of this study, computer-based decision models can be developed using multi-criteria methods that will facilitate the evaluation of individuals applying based on these criteria, along with factor weights. This will lead to a more objective, faster, and more effective decision-making process. Consequently, institutions with specific credit criteria can efficiently prioritize among applicants, thereby utilizing their resources more effectively. By incorporating individual-specific variables along with certain economic data, a more accurate decision-making process can be established with a broader data set.

References

1. Abdou, H., Pointon, J., and El-Masry, A. (2008). Neural Nets versus Conventional Techniques in Credit Scoring in Egyptian Banking. Expert Systems with Applications, 35(3), 1275-1292.
2. Akkoç, S. (2012). An Empirical Comparison of Conventional Techniques, Neural Networks and The Three Stage Hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) Model for Credit Scoring Analysis: The Case of Turkish Credit Card Data. European Journal of Operational Research, 222(1), 168-178.
3. Al-Aradi, A. (2014). Credit Scoring via Logistic Regression.
4. Aphale, A. S., and Shinde, S. R. (2020). Predict loan approval in banking system machine learning approach for cooperative banks loan approval. International Journal of Engineering Trends and Applications (IJETA), 9(8).
5. Bolton, C. (2009). Logistic Regression and Its Applications in Credit Scoring. Master Thesis. South Africa: University Van Pretoria.
6. Chen, K., Zhu, K., Meng, Y., Yadav, A., and Khan, A. (2018). Mixed Credit Scoring Model of Logistic Regression and Evidence Weight in the Background of Big Data. In International Conference on Intelligent Systems Design and Applications (pp. 435-443). Springer, Cham.
7. Chesser, D. L. (1974). Prediction Loan Noncompliance. Journal of Commercial Bank Lending, 56(8), 28-38.
8. Chien, Y. W., and Devaney, S. A. (2001). The Effects of Credit Attitude and Socioeconomic Factors on Credit Card and Installment Debt. Journal of Consumer Affairs, 35(1), 162-179.
9. data.library.virginia.edu. (n.d.). Hierarchical Linear Regression. Retrieved 11. 11. 2020, from https://data.library.virginia.edu/hierarchical-linear-regression
10. Dong, G., Lai, K. K., and Yen, J. (2010). Credit Scorecard Based on Logistic Regression with Random Coefficients. Procedia Computer Science, 1(1), 2463-2468.
11. Kadam, A. S., Nikam, S. R., Aher, A. A., Shelke, G. V., and Chandgude, A. S. (2021). Prediction for loan approval using machine learning algorithm. International Research Journal of Engineering and Technology (IRJET), 8(04).
12. Kibria, M. G., and Sevkli, M. (2021). Application of deep learning for credit card approval: A comparison with two machine learning techniques. International Journal of Machine Learning and Computing, 11(4), 286-290.
13. Kwofi, C., Owusu-Ansah, C., and Boadi, C. (2015). Predicting the Probability of Loan-Default: An application of Binary Logistic Regression. Research Journal of Mathematics and Statistics., 7(4), 46-52.
14. Lessman, S., Baesens, B., Seow, H. V., and Thomas, L. C. (2015). Benchmarking State-of the Art Classification Algorithms for Credit Scoring: An Update of Research. European Journal of Operational Research, 247(1), 124-136.
15. Myers, J. H., and Forgy, E. W. (1963). The Development of Numerical Credit Evaluation Systems. Journal of the American Statistical Association, 58(303), 799-806.
16. Mylonakis, J., and Diacogiannis, G. (2010). Evaluating the Likelihood of Using Linear Discriminate Analysis as a Commercial Bank Card Owners Credit Scoring Model. International Business Research, 3(9).
17. Nevin, J. R., and Churchill, G. A. (2010). The Equal Credit Opportunity Act: An Evaluation. Journal of Marketing, 43(2), 95-104.
18. Sarlija, N., Vlah, K., and Rosenzweig, V. V. (2009). Logistic regression and Multicriteria Decision Making in Credit Scoring. In 10th International Symposium on Operational Research SOR'09.
19. Sohn, S. Y., Kim, D. H., and Yoon, J. H. (2016). Technology Credit Scoring Model with Fuzzy Logistic Regression. Applied Soft Computing (43), 150-158.
20. Sönmez, F. (2015). Using Artificial Neural Networks and Decision Trees for Credit Scoring: A Model Proposal. ABMYO Journal (40), 1-22.
21. Steenackers, A., and Goovaerts, M. (1989). A Credit Scoring Model for Personal Loans. Insurance: Mathematics & Economics, 8(1), 31-34.
22. Tabagari, S. (2015). Credit Scoring by Logistic Regression. Doctoral Dissertation. Tartu. and

23. Torosyan, N. (2017). Application of Binary Logistic Regression in Credit Scoring. Doctoral Dissertation. Tartu.
24. Tsai, C. F., and Chen, M. L. (2010). Credit Rating by Hybrid Machine Learning Techniques. Applied Soft Computing, 10(2), 374-380.
25. West, D. (2000). Neural Network Credit Scoring Models. Computers & Operations Research, 27(11-12), 1131-1152.
26. Whittingham, M. J., Stephens, P. A., Bradbury, R. B., and Freckleton, R. P. (2006). Why Do We Still Use Stepwise Modelling in Ecology and Behavior? Journal of Animal Ecology, 75(5), 1182-1189.