Available online at www.dergipark.org.tr/en

INTERNATIONAL ADVANCED RESEARCHES
and
ENGINEERING JOURNAL

Journal homepage: www.dergipark.org.tr/en/pub/iarej

e-ISSN: 2618-575X

International Open Access

Volume 06
Issue 02

August, 2022

Research Article

Predicting acceptance of the bank loan offers by using support vector machines

Mehmet Furkan Akça [a,*] **and** ***Onur Sevli*** [b]

[a]*Burdur Mehmet Akif Ersoy University, Graduate School of Natural and Applied Sciences MSc Student in Computer Engineering, Burdur, 15030, Turkey*
[b]*Burdur Mehmet Akif Ersoy University, Faculty of Engineering and Architecture, Department of Computer Engineering, Burdur, 15030, Turkey*

ARTICLE INFO

Article history:
Received 16 January 2022
Accepted 13 June 2022
Published 15 August 2022

Keywords:
Bank loan approval
Comparison with kernels
Machine learning
Support vector machine

ABSTRACT

Loans are one of the main profit sources in banking system. Banks try to select reliable customers and offer them personal loans, but customers can sometimes reject bank loan offers. Prediction of this problem is an extra work for banks, but if they can predict which customers will accept personal loan offers, they can make a better profit. Therefore, at this point, the aim of this study is to predict acceptance of the bank loan offers using the Support Vector Machine (SVM) algorithm. In this context, SVM was used to predict results with four kernels of SVM, with a grid search algorithm for better prediction and cross validation for much more reliable results. Research findings show that the best results were obtained with a poly kernel as 97.2% accuracy and the lowest success rate with a sigmoid kernel as 83.3% accuracy. Some precision and recall values are lower than normal ones, like 0.108 and 0.008 due to unbalanced dataset, like for 1 true value, there are 9 negative values (9.6% true value). This study recommends the use of SVC in banking system while predicting acceptance of bank loan offers.

1. Introduction

Primary business of a bank is lending. The main source of profit is the interest on the loan[1, 2]. On the one hand, banks decide whether the borrower is defaulter or non-defaulter before giving the loan to customers [3]. On the other hand, they offer personal loans to some customers who are reliable, but generally, customers reject personal loans like in our samples in dataset [4]. Due to this problem, the prediction of which customer will accept the personal loan is an important task for the banking system.

For several problems, the banking industry requires more accurate predictive modeling system[5]. Bank workers can make those models with manually, but this process takes long time and lots of man-hours. At this point, machine learning (ML) techniques are extremely beneficial to predict outcomes when dealing with huge amounts of data[5]. So, those models can be applied to banking system via using ML techniques. After that predictive model, if we can predict which customers will accept personal loan offers of banks using machine learning, the loan approval process will be automated, so banks can save lots of man-hours and improve customer service [6]. In this study, Support Vector Machine (SVM) algorithm will be used to predict which customer will accept personal loan offer of banks because of classification problem.

SVM was first time officially used by Boser et al (1992) in the article titled “A Training Algorithm for Optimal Margin Classifiers”[7]. SVM is a modern non-linear, non-parametric classification algorithm that has a lot of promise. It is appropriate for binary classification applications and includes features of non-parametric applied statistics, neural networks, and ML[8]. The structure of SVM has several computational advantages, including special direction at a finite sample and a lack of correlation between algorithm complexity and sample dimension[9]. In the situation of non-regularity in the data (i.e. not evenly distribution of the data or having an uncertain distribution), SVM can be a valuable tool for insolvency analysis[8]. SVM algorithms solve non-convex and more general optimization issues as well as convex problems (e.g., linear programming, quadratic programming, second order cone programming; integer programming, semi-infinite programming)[10].

Li *et al* applied SVM to credit assessment by using real life credit card data (245 bad records and 755 good

* Corresponding author. Tel.: *+90-248-213-3150.*
E-mail addresses: akcamehmetfurkan@gmail.com *(M. F. Akça),*onursevli@mehmetakif.edu.tr *(O. Sevli)*
ORCID: 0000-0003-0289-9606 *(M. F. Akça),* 0000-0002-8933-8395 *(O. Sevli)*
DOI: 10.35860/iarej.1058724

records, with 14 variables) obtained from a Chinese commercial bank. They state that SVM is better than basic grade criterion used by the bank to predict accuracy in the field of credit assessment [9]. Dall'Asta Rigo applied six ML techniques (SVM, LR, MARS, RF, XGB, and Stacking) on four real-life credit scoring datasets (Home Credit, German Credit, Credit Card Default, Give Me Credit) for classification problem [11]. Xu *et al* used 4 machine learning methods (RF, XGBT, GBM, and NN) to predict factors affecting repayment of borrowers. They conclude that the RF performs well in the classification task of default [12]. Huang *et al* explain that SVM and NN achieve better prediction accuracy than traditional statistical methods in credit rating analyses for the US and Taiwan markets [13]. Kadam *et al* used the SVM and Naïve Bayes (NB) to predict loan approval. They conclude that NB fulfills of needs of bankers [14]. Bayraktar *et al* compared commonly used machine learning methods with deep learning methods (Classification Restricted Boltzmann Machine and Multilayer Artificial Neural Networks) [15]. Aphale and Shinde used various ML techniques (Neural Network, Discriminant Analysis, Naïve Bayes, K-Nearest Neighbor, Linear Regression, Ensemble Learning, and Decision Tress) to predict the creditworthiness of borrowers [3].

It is a fact that all articles in the literature focus on credit risk management, credit rating, loan repayment, decision-aid for loaners, and credit default. However, this study aims at predicting acceptance of the bank loan offers by utilizing Support Vector Machine (SVM) algorithm. It can be said that this study is the first research to use SVM in predicting acceptance of the loan offer of a bank to customers. Therefore, this study will contribute to loan and banking system due to no existing of any article or study related to same topic in the literature.

Table 1. Features and statistical characteristics of the dataset

Feature Names	Description	Mean	Standard Deviation	Min Value	Max Value	Feature Type
ID	Customer's unique ID number	2500.50	1443.52	1	5000	Categorical
Age	Customer's age	45.33	11.46	23	67	Numeric
Experience	Work experience by number of years	20.10	11.46	-3	43	Numeric
Income	0.1% Percentage of annual income	73.77	46.03	8	224	Numeric
ZIP Code	ZIP Code of where customer lives	93152.50	2121.85	9307	96651	Categorical
Family	Family Size	2.39	1.14	1	4	Numeric
CC Avg	0.1% Percentage of average credit card spending per month	1.93	1.74	0	10	Numeric
Education	Level of education (1-Undergraduate, 2-Graduate, 3-Advanced)	1.88	0.83	1	3	Categorical
Mortgage	If any house mortgage, its value.	56.49	101.71	0	635	Numeric
Personal Loan	Acceptance of personal loan offer by the customer in the last campaign season	0.09	0.29	0	1	Categorical
Securities Account	Is there any securities account with the customer?	0.10	0.30	0	1	Categorical
CD Account	Is there any certificate of deposit account with the customer?	0.06	0.23	0	1	Categorical
Online	Is this customer using internet banking?	0.59	0.49	0	1	Categorical
Credit Card	Is this customer using a credit card issued by bank?	0.29	0.45	0	1	Categorical

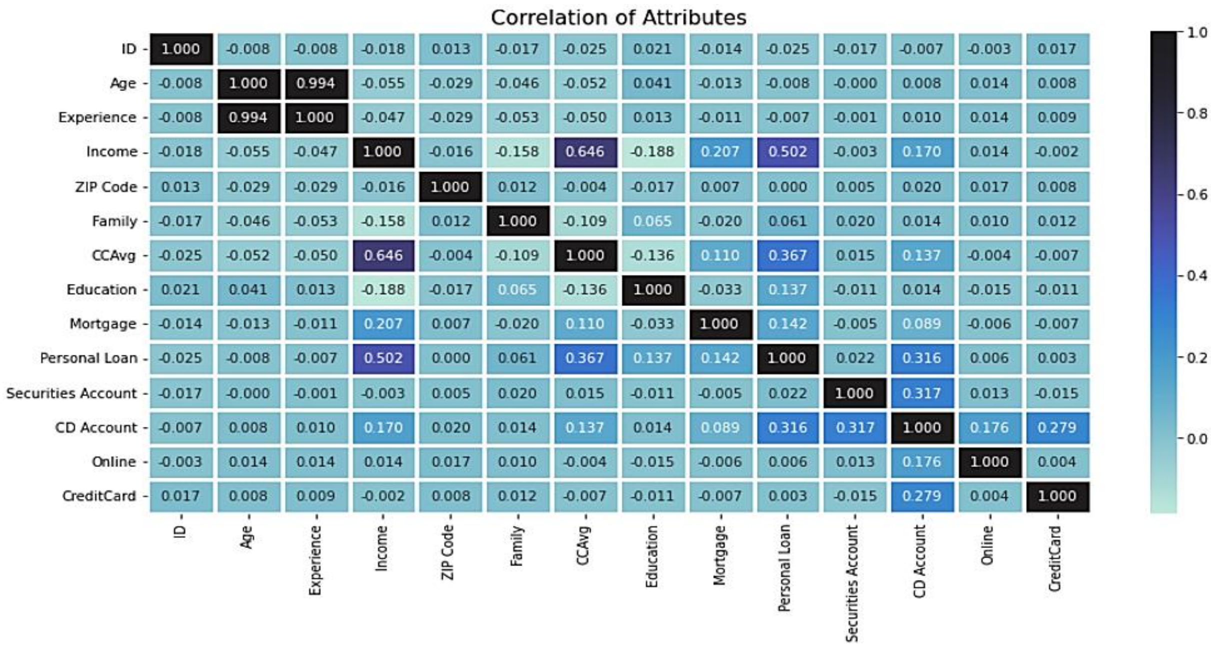

Figure 1. Correlation of attributes

2. Material and Methods

2.1 Dataset

This publicly available dataset was obtained from Kaggle and shared by Walke [4]. The dataset belongs to "Thera Bank" and contains 5000 customers' demographic information like "Age" and "Income" columns, and relationship with bank data like "Mortgage" and "Securities Account" columns. And customers' response to the last campaign, like the Personal Loan column. Among these customers, only 480 (or 9.6%) accepted this offer[4].

When the bank loan dataset was checked, selected features from dataset for this study are seen in Table 1. In addition, table 1 shows mean, standard deviation, min value, max value and feature type. There are no missing values or duplication, and also no values of string type. This information is important because some machine learning algorithms cannot work with string values, and missing values and duplication can badly affect prediction results. If a string value is existed, a label encoder could be used to solve that problem.

After these details, columns which are useless for this study must be selected. At the beginning, the correlation matrix at Figure 1 should be checked for a visualization of how the columns affect the target column, Personal Loan. As can be seen in the correlation matrix, each column has a positive or negative impact on the target value. ID column and ZIP Code column were deleted because ID is a unique value for everybody in dataset; ZIP Code is decreasing prediction accuracy. After these deletions and details, dataset can be used for ML algorithms.

2.2 Methods

Before the classification process, we must know the general process of how to apply machine learning algorithms. Figure 2 shows general process of this ML study with grid search and 5-fold cross validation. In this study, SVM algorithms were used to predict the acceptance of bank personal loan offers. Before this prediction, we must separate the Personal Loan column to another data frame because that column will be our target column. After this process, we can use train test split (TTS), but generally, TTS is not a reliable method for machine learning prediction because TTS works differently with different random state values. So, we are creating train and test data using Cross Validation.

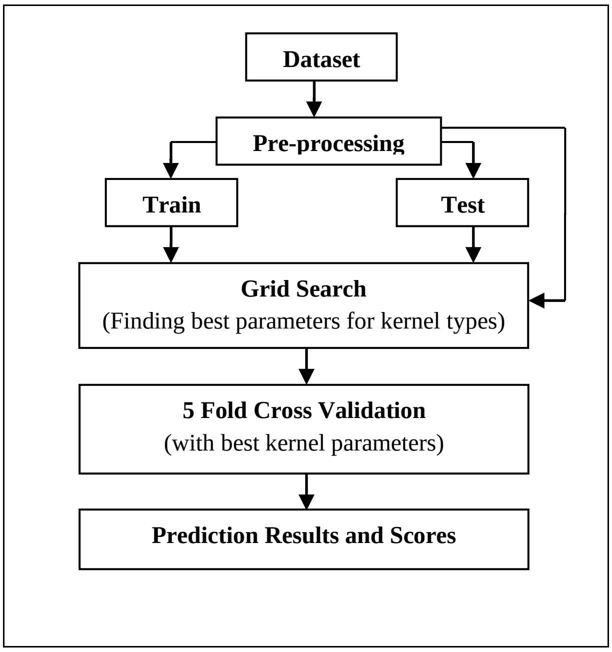

Figure 2. Our study process

2.2.1 Support Vector Classifier (SVC)

A support vector machine (SVM) is a supervised learning model (invented by Vladimir Vapnik) and used for classification and prediction of data[16]. However, SVM is generally used in classification problems. Each data item as a point in n-dimensional space is represented in the SVM method. Each feature's value corresponds to the value of a specific coordinate. Then the hyper-plane that clearly separates the two classes to complete categorization is located [17]. There are several hyperplanes from which to choose to split the two kinds of data points. The aim is to determine a plane with the largest margin, or the largest range among data points from both groups. Expanding the margin distance adds reinforcement, making following data points easier to classify categorize[18].

There have been many kernel functions in SVM. However, following 4 kernel functions are popular[19]:

- ***Linear kernel:***

$$K(xi, xj) = xi^{T} * xj \quad (1)$$

- ***Polynomial kernel:***

$$K(xi, xj) = (\gamma\, xi^{T} * xj + r)d, \gamma > 0 \quad (2)$$

- ***RBF kernel:***

$$K(xi, xj) = \exp(-\gamma \left\| xi - xj \right\| 2), \gamma > 0 \quad (3)$$

- ***Sigmoid kernel:***

$$K(xi, xj) = \tanh(\gamma\, xi^{T} * xj + r) \quad (4)$$

Where, γ, r and d are kernel parameters

In this study, above kernel types and grid search algorithm will be used to find the best parameters. After this selection, cross validation will be applied to predict results and take several metric results like accuracy, precision, recall and f1 score.

3. Experimental Study and Findings

In this study, confusion matrix, accuracy score, precision score, recall score and f1 score metrics will be used to evaluate SVM algorithm.

3.1 Evaluation Metrics

The performance of a model can be explained using evaluation metrics. The ability of evaluation metrics to differentiate between model results is a key feature[20].

3.1.1 Confusion Matrix

It is a N X N matrix in which N is the number of classes being predicted [20]. For this article confusion matrix as Table 2 will be used.

Table 2. Representation of cells in confusion matrix

	Predicted:0	Predicted:1
Actual:0	TN	FP
Actual:1	FN	TP

3.1.2 Accuracy Score

The percentage of correct guesses in the total number of predictions is called accuracy[20]. Accuracy was calculated with following equation (5).

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (5)$$

3.1.3 Precision Score

The fraction of accurately detected affirmative cases is known as precision[20]. Precision score show us the perfectness of predictive model [21]. Precision was calculated with following equation (6).

$$\text{Precision} = (\text{TP})/(\text{TP} + \text{FP}) \quad (6)$$

3.1.4 Recall Score

The fraction of real positive instances that are accurately detected is referred to as recall[20]. Recall was calculated with following equation (7).

$$\text{Recall} = (\text{TP})/(\text{FP} + \text{FN}) \quad (7)$$

3.1.5 F1 Score

For a classification problem, the F1-Score is the harmonic mean of precision and recall values[21]. F1 was calculated with following equation (8).

$$\text{F1} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right) \quad (8)$$

3.2 Results

Support vector machine algorithm with 4 kernel types, results shown in below with Table 3. Confusion matrices and other metrics are mean version of 5-fold cross validation results. All metric scores were generalized in Table 4.

According to the Table 3 and Table 4, all our kernels have a successful accuracy score, but we cannot evaluate with just one parameter, accuracy score. It can be said that scores over 80% is good scores, so our precision scores except sigmoid kernel is successful. About recall scores there is one kernel (polynomial) we can say is successful. About F1 scores polynomial and rbf kernels are successful. Reason for lower scores, we do not have a balanced dataset, so there are a few positive results, so when SVM kernels try to classify that few positive results, there is failure happening especially with sigmoid kernel.

We can clearly say that according to this paper if we have an unbalanced dataset we cannot get successful results with sigmoid kernel, we can choose polynomial kernel.

There exists no any article or study related to same topic with this study. Some studies related to bank loan approval and banking system topics are given in Table 5.

Table 3. Confusion Matrix

	Actual Value	Predicted Value 0	Predicted Value 1
Linear SVC	0	895.2	8.8
	1	40.2	55.8
Poly SVC	0	892.4	11.6
	1	16.4	79.6
Sigmoid SVC	0	832.4	71.6
	1	95.2	0.8
Rbf SVC	0	895.2	8.8
	1	23.4	72.6

Table 4. Metric results

Kernel Type	Metrics			
	Accuracy	Precision	Recall	F1
Linear	0,951	0,863	0,581	0,694
Poly	0,972	0,872	0,829	0,850
Sigmoid	0,833	0,108	0,008	0,009
Rbf	0,967	0,893	0,757	0,818

Table 5. Comparison with similar studies in the literature

Authors of the Article	Highest Score ML Technique	Accuracy
Sheikh *et al.* [21]	Logistic Regression	81.1%
Vimala and Sharmili [22]	SVM	~79%
Fati [23]	Logistic Regression	79%
Madaan *et al.* [24]	Random Forest	80%
Sreesouthry *et al.* [25]	Logistic Regression	77%
Yaurita and Rustam [26]	SVM (Rbf)	85%
Kumar *et al.* [27]	Decision Tree	95%
Ndayisenga [28]	SVM	77%

4. Conclusion

Literature review shows that machine learning algorithms have played a significant role in the prediction of acceptance of personal bank loan offers.

SVM is among the best-performing statistical-learning or machine-learning algorithms in terms of accuracy [29, 30, 31]. In this study, a support vector machine algorithm with four kernel types was used. According to the results of the analyses, the best results were obtained with a polynomial kernel (97%) and the worst results with a sigmoid kernel (83%). Some precision and recall values are very low compared to normal because our dataset is an unbalanced dataset, meaning for every true value, there are 9 negative values. When we use an unbalanced dataset, this problem can occur. But the general performance of support vector machines is satisfying, and we can say that SVM with a polynomial kernel is a good choice to predict loan results like in our study. When we compare with similar studies, there are different types of ML algorithms used. Generally, accuracy scores between 77% and 85%.

After the comparison we can say that SVM with polynomial kernel is successful for banking system classification problems because our study's accuracy and other metric scores are higher than similar study's.

Finally, if banking systems use a machine learning method to predict acceptance of bank loan offers, they can predict their future profit easily.

Declaration

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

Author Contributions

M.F. Akça developed the methodology and performed the analysis. O. Sevli supervised and improved the study. M.F. Akça and O. Sevli wrote the manuscript together.

References

1. Arun, K., G. Ishan, and K. Sanmeet, *Loan approval prediction based on machine learning approach*. IOSR J. Comput. Eng, 2016. **18**(3): p. 18-21.
2. Bhandari, M., *How to predict loan eligibility using machine learning models*. [cited 2022 02 January]; Available from: https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057.
3. Aphale, A.S., and S.R. Shinde, *Predict loan approval in banking system machine learning approach for cooperative banks loan approval*. International Journal of Engineering Research & Technology, 2020. **9**(8): 991-995
4. Walke, K. *Bank personal loan modelling*. [cited 2021 03 October]; Available from: https://www.kaggle.com/krantiswalke/bank-personal-loan-modelling.
5. Tejaswini, J., T.M. Kavya, R.D.N. Ramya, P.S. Triveni, and V.R. Maddumala, *Accurate loan approval prediction based on machine learning approach.* Journal of Engineering Sciences, 2020. **11**(4): p. 523-532.
6. Pandey, N., R. Gupta, S. Uniyal, and V. Kumar, *Loan approval prediction using machine learning algorithms approach.* International Journal of Innovative Research in Technology, 2021. **8**(1): p. 898-902.

7. Boser, B.E., I.M. Guyon, and V.N. Vapnik, *A training algorithm for optimal margin classifiers*, in Proceedings of the fifth annual workshop on Computational learning theory, 1992. Association for Computing Machinery: Pittsburgh, Pennsylvania, USA: p. 144–152.
8. Auria, L., and R.A. Moro, *Support vector machines (SVM) as a technique for solvency analysis.* DIW Berlin Discus. Paper, 2008. [cited 2022 02 January] Available from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1424949.
9. Li, J., J. Liu, W. Xu, and Y. Shi, *Support vector machines approach to credit assessment*, in Computational Science-ICCS 2004, Lecture Notes in Computer Science 3039, Berlin Heidelberg: Springer. p. 892-899.
10. Tian, Y., Y. Shi, and X. Liu, *Recent advances on support vector machines research.* Technological and Economic Development of Economy, 2012. **18**(1): p. 5-33.
11. Dall'Asta Rigo, E.Y., *Evaluation of stacking for Predicting credit risk scores*. MSc Thesis at TED University Graduate School Applied Data Science, 2020. p. 1-75.
12. Xu, J., Z. Lu, and Y. Xie, *Loan default prediction of Chinese P2P market: a machine learning methodology*. Scientific Reports, 2021. **11**: 1-19.
13. Huang, Z., H. Chen, C.J. Hsu, W.H. Chen, and S. Wu, *Credit rating analysis with support vector machines and neural networks: A market comparative study. Decis.* Support Syst., 2004. **37:** 543-558.
14. Kadam, A.S., S.R. Nikam, A.A. Aher, G.V. Shelke, and A.S. Chandgude, *Prediction for loan approval using machine learning algorithm*. International Research Journal of Engineering and Technology (IRJET), 2021. **8**(4): 4089-4092.
15. Bayraktar, M., M.S. Aktas, O. Kalipsiz, O. Susuz, and S. Bayraci, *Credit risk analysis with classification restricted boltzmann machine*. in Proceeding of 26th Signal Processing and Communications Applications Conference (SIU), 2018, p. 1-4.
16. IBM. [cited 2022 02 January]; Available from: https://www.ibm.com/cloud/learn/supervised-learning.
17. Ray, S. *Understanding support vector machine (SVM) algorithm from examples (along with code)*. September 13, 2017 [cited 2022 02 January]; Available from: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/.
18. Gandhi, R. *Support vector machine-Introduction to machine learning algorithms*. [cited 2022 02 January]; Available from: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.
19. Hsu, C.W., C.C. Chang, and C.J. Lin, *A practical guide to support vector classification*. Technical Report 2003, Department of Computer Science and Information Engineering, University of National Taiwan. p. 1-12.
20. Srivastava, T, *11 important model evaluation metrics for machine learning everyone should know*. August 6, 2019 [cited 2022 02 January]; Available from: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/.
21. Sheikh, M.A., A.K. Goel, and T. Kumar. *An Approach for Prediction of Loan Approval using Machine Learning Algorithm*. in Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020), IEEE Xplore Part Number: CFP20V66-ART, 2020. p. 490-494.
22. Vimala, S., and K. Sharmili. *Prediction of loan risk using naive bayes and support vector machine*. in International Conference on Advancements in Computing Technologies - ICACT 2018. International Journal on Future Revolution in Computer Science & Communication Engineering (Special Issue), 2018. 4(2): p. 110-113. Available from: http://www.ijfrcsce.org/index.php/ijfrcsce/Special_Issue/ICACT_2018_Track.
23. Fati, S.M., *Machine learning-based prediction model for loan status approval.* Journal of Hunan University Natural Sciences, 2021. **48**(10): p. 1-8.
24. Madaan, M., A. Kumar, C. Keshri, R. Jain, and P. Nagrath, *Loan default prediction using decision trees and random forest: A comparative study*. in IOP Conference Series: Materials Science and Engineering. 2021. IOP Publishing, p.1-12.
25. Sreesouthry, S., A. Ayubkhan, M.M. Rizwan, D. Lokesh, and K.P. Raj, *Loan Prediction Using Logistic Regression in Machine Learning.* Annals of the Romanian Society for Cell Biology, 2021. **25**(4): p. 2790-2794.
26. Yaurita, F., and Z. Rustam. *Application of support vector machines for reject inference in credit scoring*. in Proceedings of the 3rd International Symposium on Current Progress in Mathematics and Sciences 2017 (ISCPMS2017), AIP Conf. Proc. 2023, 020209-1-020209-6; https://doi.org/10.1063/1.5064206.
27. Kumar, R., et al., *Prediction of loan approval using machine learning.* International Journal of Advanced Science and Technology, 2019. **28**(7): p. 455 - 460.
28. Ndayisenga, T., *Bank loan approval prediction using machine learning techniques*. MSc Dissertation in Data Science in Actuarial Science at the African Center of Excellence in Data Science in University of Rwanda, 2021. p. 1-28.
29. Maroco, J., D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. de Mendonca, *Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests.* BMC research notes, 2011. **4**(299): p. 1-14.
30. Guenther, N., and M. Schonlau, *Support vector machines.* The Stata Journal, 2016. **16**(4): p. 917-937.
31. Prasad, K.G.S., P.V.S. Chidvilas, and V.V. Kumar, *Customer loan approval classification by supervised learning model*. International Journal of Recent Technology and Engineering, 2019. **8**(4): 9898-9901.