

# A Research Paper on Credit Card Approval, Using Machine Learning models.

Sawan Nihatkar<sup>1</sup>, Aditya Shahari<sup>2</sup>, Tushar Patil<sup>3</sup>, Atharva Kulkarni<sup>4</sup>, Ms. Suvarna Potdukhe<sup>5</sup>

BE Student, Department of Information Technology<sup>1,2,3,4</sup>

Professor, Dept. of IT Engineering<sup>5</sup>

RMD Sinhgad School of Engineering, Pune, India

**Abstract:** This study focuses on the use of machine learning (ML) approaches to forecast consumer eligibility for a credit card in order to prevent any future credit risk that might influence the bank's financial stability and credit performance. A credit card is a credit facility provided to a consumer by banks and financial organisations all over the world. Banks and financial institutions face credit risk as a result of the credit facility. Repayments are rarely guaranteed, and the loan frequently becomes a non-performing credit facility (NPL). To reduce credit risk, banks analyse applicants' creditworthiness and eligibility before issuing a credit facility. Traditional credit scoring techniques are used to make the judgement, and credit worthiness is not always correct. Using predictive models, this initiative intends to assist banks and financial institutions in identifying and interacting with creditworthy consumers. To create models, we employed Stochastic Gradient Descent (SGD) and Logistic Regression. SGD Classifier, Logistic Regression, SVC, Decision Tree Classifier, Random Forest Classifier, Gaussian NB, K Neighbours Classifier, Gradient Boosting Classifier, Linear Discriminant Analysis, Bagging Classifier, MLP Classifier, Ada Boost Classifier, Extra trees Classifier. Both linear SVM and nonlinear SVM models were utilised to determine the optimal SVM approach. Statistical approaches are used for feature selection in filter-based feature selection methods. For training and test data, model accuracy was assessed using Mean Absolute Error, Confusion Matrix, and Area Under Curve (AUC). We tested three classifiers and discovered that Nonlinear SVM outperformed the others.

**Keywords:** SGD Classifier, Logistic Regression, SVC, Decision Tree Classifier, Random Forest Classifier, Gaussian NB, K Neighbours Classifier, Gradient Boosting Classifier, Linear Discriminant Analysis, Bagging Classifier, MLP Classifier, Ada Boost Classifier, Extra trees Classifier

## REFERENCES

- [1]. Agarwal, A., Rana, A., Gupta, K., Verma, N. (2020). A Comparative Study and enhancement of classification techniques using Principal Component Analysis for credit card dataset. In 2020 International Conference on Intelligent Engineering and Management (ICIEM) [Conference paper]. Presented at the 2020 International Conference on Intelligent Engineering and Management (ICIEM), IEEE, London, United Kingdom, pp. 443–448.  
[Link](<https://doi.org/10.1109/ICIEM48762.2020.9160230>)
- [2]. Antonakis, A.C., Sfakianakis, M.E. (2009). Assessing naïve Bayes as a method for screening credit applicants. *Journal of Applied Statistics*, 36, 537–545.  
[Link](<https://doi.org/10.1080/02664760802554263>)
- [3]. Banasik, J., Crook, J., Thomas, L. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, 50, 6.
- [4]. Bhatore, S., Mohan, L., Reddy, Y.R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. *Journal of Banking and Financial Technology*, 4, 111–138.  
[Link](<https://doi.org/10.1007/s42786-020-00020-3>)
- [5]. Blagus, R., Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14, 106.
- [6]. [Link](<https://doi.org/10.1186/1471-2105-14-106>)

- [7]. Chornous, G., Nikolskyi, I. (2018). Business-Oriented Feature Selection for Hybrid Classification Model of Credit Scoring. In 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP) [Conference paper]. Presented at the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), IEEE, Lviv, pp. 397–401.  
[Link](<https://doi.org/10.1109/DSMP.2018.8478534>)
- [8]. Data Science Process Alliance. (n.d.). What is CRISP DM? [Website]. URL: <https://www.datascience-pm.com/crisp-dm-2/> (accessed 11.20.20).
- [9]. Elreedy, D., Atiya, A.F. (2019). A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. *Information Sciences*, 505, 32–64.
- [10]. Link](<https://doi.org/10.1016/j.ins.2019.07.070>)
- [11]. Fernando, J. (2021). Delinquent Account Credit Card [Website]. Loan Basis. URL: <https://www.investopedia.com/terms/d/delinquent-account-credit-card.asp> (accessed 4.28.21).
- [12]. Galarnyk, M. (2018). Understanding Boxplots [Website]. URL: <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51> (accessed 1.20.21).
- [13]. Hussein, A.S., Li, T., Yohannese, C.W., Bashir, K. (2019). A-SMOTE: A New Preprocessing Approach for Highly Imbalanced Datasets by Improving SMOTE. *International Journal of Computational Intelligence Systems*, 12(2), 11.
- [14]. Jafar Hamid, A., Ahmed, T.M. (2016). Developing Prediction Model of Loan Risk in Banks Using Data Mining. *MLAIJ*, 3, 1–9.  
[Link](<https://doi.org/10.5121/mlaij.2016.3101>)
- [15]. Karthiban, R., Ambika, M., Kannammal, K.E. (2019). A Review on Machine Learning Classification Technique for Bank Loan Approval. In 2019 International Conference on Computer Communication and Informatics (ICCCI) [Conference paper]. Presented at the 2019 International Conference on Computer Communication and Informatics (ICCCI), IEEE, Coimbatore, Tamil Nadu, India, pp. 1–6.  
[Link](<https://doi.org/10.1109/ICCCI.2019.8822014>)
- [16]. Kumar, B. (2021). 10 Techniques to deal with Imbalanced Classes in Machine Learning [Website]. URL: <https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/>
- [17]. Kumar Gupta, D., Goyal, S. (2018). Credit Risk Prediction Using Artificial Neural Network Algorithm. *IJMECS*, 10, 9–16.  
[Link](<https://doi.org/10.5815/ijmeecs.2018.05.02>)
- [18]. Lee, T., Chen, I. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28, 743–752.  
[Link](<https://doi.org/10.1016/j.eswa.2004.12.031>)
- [19]. Leo, M., Sharma, S., Maddulety, K. (2019). Machine Learning in Banking Risk Management: A Literature Review. *Risks*, 7, 29.  
[Link](<https://doi.org/10.3390/risks7010029>)
- [20]. Madyatmadja, E.D., Aryuni, M. (2005). Comparative Study of Data Mining Model for Credit Card Application Scoring in Bank. *Journal of Theoretical and Applied Information Technology*, 59, 6.
- [21]. Marqués, A.I., García, V., Sánchez, J.S. (2012). Two-level classifier ensembles for credit risk assessment. *Expert Systems with Applications*, 39, 10916–10922.  
[Link](<https://doi.org/10.1016/j.eswa.2012.03.033>)
- [22]. Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J., Ryu, K. (2019). An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. *Sustainability*, 11, 699.  
[Link](<https://doi.org/10.3390/su11030699>)
- [23]. Nellur, S. (2020). The Neural Network at its Simplest [Website]. URL: <https://medium.com/analytics-vidhya/neural-networks-in-nutshell-7d1cc3ae6443> (accessed 2.25.21).

- [24]. Oreski, S., Oreski, D., Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications*, 39, 12605–12617. [Link](<https://doi.org/10.1016/j.eswa.2012.05.023>)
- [25]. Payments and Settlements Department. (2020). *Payments Bulletin - Third Quarter (Quarterly Report)* [Report]. *Payment Bulletin*. Central Bank of Sri Lanka