



# A Review on Credit Prediction and Risk analysis on Loans Data Machine Learning Classification Techniques

<sup>1</sup>Mahesh Shesharao Rathod, <sup>2</sup>Dr V H Deshmukh

<sup>1</sup>Students, Master of Engineering (CSE) <sup>2</sup>Professor

<sup>1</sup>Computer Science and Engineering,

<sup>1</sup>PRMITR, Badnera, India

**Abstract:** Recently, with the advance of electronic commerce and big data technology, P2P online lending platforms have brought opportunities to businessmen, but at the same time, they are also faced with the risk of user loan default, which is related to the sustainable and healthy development of platforms. Therefore, based on ensemble classifier using classification algorithms such as Random Forest algorithm, SVM, Naive Bayes, multilayer perceptron (MLP) and K-nearest neighbors (KNN) algorithm, a credit risk prediction model can be built in view of the real-world user loan data. With the improvement within the banking sector, several individuals area unit applying for bank loans. However the bank has its restricted assets so they grant loans to restricted individuals solely, thus sorting out to whom the loan is be lent so that the borrower is a safer choice for the bank. Thus we try to cut back this risk issue behind choosing the safe person thus on save efforts and assets of lenders. This can be done by mining the large knowledge of the previous records of the individuals to whom the loan was granted before and based on those records and experiences the machine will be trained with the machine learning model that offer the foremost correct result. The objective of the proposed system is to predict whether or not assignment of loan to a person is going to be safe or not. So it will be split into four sections(i)Data assortment (ii) Comparison of machine learning models on collected knowledge (iii) coaching of system on most promising model(iv) Testing.

**Index Terms** – Loan Prediction, Ensemble Classifier, SVM, MLP, KNN, Naïve Bayes

## 2. INTRODUCTION

Distribution of the loans is that the core business is a part of nearly every bank. Most portions of the bank's assets are directly come from the profit attained from the loans distributed by the banks. The prime objective in the banking atmosphere is to invest their assets in safe hands wherever it's. these days several banks/financial corporations approve loans once a regress method of verification and validation however still there's no surety whether or not the chosen candidate will default or repay the loan. Through this method, we are able to predict whether or not that specific borrower is safe or not and also the whole method of validation of options is automatic by machine learning technique. Loan Prediction is extremely useful for workers of banks moreover as for the human conjointly. The aim is to produce fast, immediate, and straightforward decision to opt for the right candidate. It will be a special blessing to the bank. The Loan Prediction System will mechanically calculate the burden of every option participating in the loan process and on new check knowledge same options are processed with relevance to their associated weight. A point in time is often set for the person to visualize whether or not his loan is often sanctioned or not. Result against specific Loan IDs will be sent to varied departments of banks so they'll take applicable action on the application. This helps all alternatives department to apply other formalities.

## 3. Literature Review

A large number of research work has been done in the field of credit prediction and analysis, which assists banking institutions to optimize their decision taking capabilities. These models vary in terms of their deployment characteristics & real-time performance. For instance, work in [5, 6] proposes use of invoice data with different mining techniques, which assists in

estimation of simplistic transactional patterns. This work doesn't provide comprehensive insights into these transactions, due to which it has minimum applicability. To improve this performance, work in [7] proposes use of supervised mining models, which can be trained as per context-sensitive requirements. Due to use of supervised learning, the models are able to continuously optimize classification performance under different use cases. These models are further extended via the work in [8, 9] which discusses use of Logistic Regression with LightGBM (LR LGBM), and SVMs for high efficiency classification process. These models are useful for unbalanced datasets, which makes them highly applicable for real-time applications. Similar models are discussed in [10, 11, 12], which propose use of Online Integrated Credit Scoring Model (OICSM), Fusion Neural Networks (FNNs), and Cost-Sensitive Neural Network Ensemble (CS NNE) that are capable of augmenting data features for high-accuracy classification applications. These models are useful for low error operations, but require larger classification delays. To improve the speed performance, work in [13, 14, 15] proposes use of Slow and Fast Learning (SFL), LightGBM, and bioinspired optimization techniques, which assist in improving credit score evaluation speed via high efficiency feature reduction processes. Work in [16, 17, 18] further proposes use of Spy Model with Transfer Learning and Adaboost optimizations (SPY TRA), Multiple linear Clusters for Feature Selection, and use of trajectory datasets for effective classification of different datasets. These models are highly useful in analysis of multidimensional datasets, which assists in improving their deployment capabilities. Work in [19] further adds privacy to these models which assists in securely mitigating user credit requests for high performance applicability. But most of these models are highly linear in their processing capabilities, and also do not perform inter-user correlative analysis. To overcome these issues, next section proposes use of a novel model for Credit Prediction with Risk Minimization via Ensemble Learning process. The model was evaluated on multiple datasets, and its performance was compared with various state-of-the-art methods under different scenarios.

#### 4. Problem Statement

Repayment to bank is one of the important aspects of any banking firms. During the continuous process of repayment of loan by customer it was sure that the customer will pay it completely during its tenure. But in between some time frame there is a risk with bank about repayment by customer or not and this arises the main problem that it was difficult to predict the defaulter in any case.

- To show how loan approval and credit risk analysis can be estimated using machine learning techniques.
- An important aspect for handling an opportunity from an incorrect prediction and risk factor through verification and validation. Creating predictive model using machine learning techniques and algorithm and evaluate each of their performance in predicting loan approval and risk factor when a client is predicted to default but in actuality would have paid off his loan and the bank does not approve the loan by analyzing wrong prediction.

#### 5. Proposed Approach

As per the literature review, it can be observed that existing models for estimation of credit scores are highly linear, and do not perform inter-user correlative analysis, which limits their applicability. To overcome this limitation, a novel model for Credit Prediction and Risk analysis via Ensemble Learning process is proposed here. Flow of the proposed model is depicted in figure 2, wherein different input parameters including user's Age, Marital Status, Gender, Status of previous payments, Transactional amounts for previous bills, and Transactional amounts for previous bill payments are aggregated, and classified via different Machine Learning Models (MLMs). The outputs of these classifiers can be combined to form an ensemble learning model, which assists in identification of current user request status. Based on this status, a correlation model is selectively activated, which results in a credit score that assists banks & other financial firms to grant credit loans to requesting users. The credit levels are accompanied with a credit score, which assists investment firms to optimize their credit decisions.

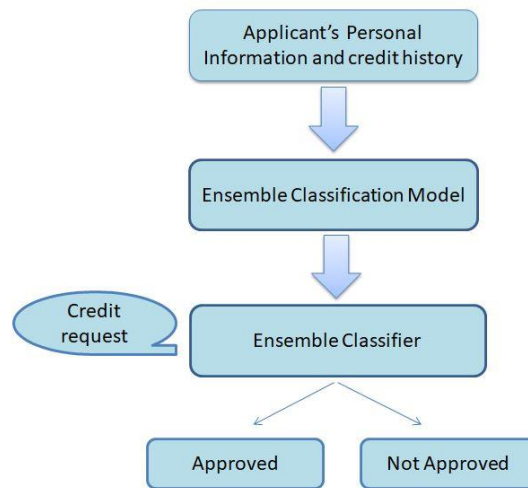


Figure 2. Overall flow of the proposed model

From the flow of model, it can be observed that input datasets are collected for user's personal information, their spending patterns, and their payment patterns. These patterns are used to train multiple ML classification for efficient classification performance. Each of these models are trained using a specific set of hyper parameters, which were selected via manual tuning process to obtain high accuracy levels.

## 6. Conclusion

This paper aimed to explore, analyze, and propose ensemble classifier using machine learning classification algorithm to identify whether the loan applicant, given certain features, can be approved its credit request and also analyze the risk associated with the applicant. This type of model could be used by Lenders for credit prediction and risk analysis future borrowers that could have the potential to default and not pay back their loan by the designated time.

The proposed Ensemble Classifier can improve the accuracy of the credit prediction and also analyze the risk. Use of such system will help financial system improve their profits.

## 5. REFERENCES

- [1] N. A. Nguyen-Thi, Q. Pham-Nhu, S. Luong-Ngoc and N. Vu-Thanh, "Using novel noise filtering techniques to ensemble classifiers for credit scoring: Using noise filtering techniques to ensemble classifiers for credit scoring," 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), 2019, pp. 1-6, doi: 10.1109/TAAI48200.2019.8959831.
- [2] A. Safiya Parvin and B. Saleena, "An Ensemble Classifier Model to Predict Credit Scoring - Comparative Analysis," 2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), 2020, pp. 27-30, doi: 10.1109/iSES50453.2020.00017.
- [3] K. B. Addi and N. Souissi, "An Ontology-Based Model for Credit Scoring Knowledge in Microfinance: Towards a Better Decision Making," 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 380-385, doi: 10.1109/IS48319.2020.9199981.
- [4] D. I. Calibo and M. A. Ballera, "Variable Selection for Credit Risk Scoring on Loan Performance Using Regression Analysis," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019, pp. 746-750, doi: 10.1109/CCOMS.2019.8821664.
- [5] Y. Sun, W. Jian, Y. Fu, H. Sun, Y. Zhu and Z. Bai, "A new perspective of credit scoring for small and medium-sized enterprises based on invoice data," 2021 2nd International Conference on Computing and Data Science (CDS), 2021, pp. 477-482, doi: 10.1109/CDS52072.2021.00088.
- [6] A. Hassan and R. Jayousi, "Financial Services Credit Scoring System Using Data Mining," 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT), 2020, pp. 1-7, doi: 10.1109/AICT50176.2020.9368572.
- [7] L. Duan, "Performance Evaluation and Practical Use of Supervised Data Mining Algorithms for Credit Card Approval," 2020 International Conference on Computing and Data Science (CDS), 2020, pp. 251-254, doi: 10.1109/CDS49703.2020.00057.
- [8] B. Liu, L. Lu, Q. Zeng and Y. Li, "Implementation of credit scoring card model based on logistic regression and lightgbm," 2021 International Conference on Control Science and Electric Power Systems (CSEPS), 2021, pp. 175-178, doi: 10.1109/CSEPS53726.2021.00042.

- [9] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar and C. V. N. M. Praneeth, "Credit Card Fraud Detection Using Machine Learning," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 967-972, doi: 10.1109/ICICCS51141.2021.9432308.
- [10] Z. Zhang, K. Niu and Y. Liu, "A Deep Learning Based Online Credit Scoring Model for P2P Lending," in IEEE Access, vol. 8, pp. 177307-177317, 2020, doi: 10.1109/ACCESS.2020.3027337.
- [11] T. Zhang and Y. Du, "Research on user Credit score Model based on Fusion Neural network," 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021, pp. 1391-1395, doi: 10.1109/IAEAC50856.2021.9390672.
- [12] W. Yotsawat, P. Wattuya and A. Srivihok, "A Novel Method for Credit Scoring Based on Cost-Sensitive Neural Network Ensemble," in IEEE Access, vol. 9, pp. 78521-78537, 2021, doi: 10.1109/ACCESS.2021.3083490.
- [13] J. P. Barddal, F. Enembreck, L. Loezer and R. Lanzuolo, "Combining Slow and Fast Learning for Improved Credit Scoring," 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 1149-1154, doi: 10.1109/SMC42975.2020.9283453.
- [14] Z. Qiu, Y. Li, P. Ni and G. Li, "Credit Risk Scoring Analysis Based on Machine Learning Models," 2019 6th International Conference on Information Science and Control Engineering (ICISCE), 2019, pp. 220-224, doi: 10.1109/ICISCE48695.2019.00052.
- [15] P. Diaconescu and V. -E. Neagoie, "Credit Scoring Using Deep Learning Driven by Optimization Algorithms," 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2020, pp. 1-6, doi: 10.1109/ECAI50035.2020.9223139.
- [16] Q. Wei, Y. Liu and K. Wu, "Transfer Learning Based Credit Scoring," 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2021, pp. 1251-1255, doi: 10.1109/CSCWD49262.2021.9437749.
- [17] L. Alfat, M. Rizkinia, R. F. Sari and D. M. Romano, "Feature Selection of Credit Score Factor Based on Smartphone Usage using MCFS," 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON), 2019, pp. 1-5, doi: 10.1109/TIMES-iCON47539.2019.9024462.
- [18] W. Wang, Y. Yang, A. Chen and Z. Pan, "A Scoring Method for Driving Safety Credit Using Trajectory Data," 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2019, pp. 558-565, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00135.
- [19] Y. Han, H. Chen, Z. Qiu, L. Luo and G. Qian, "A Complete Privacy-Preserving Credit Score System Using Blockchain and Zero Knowledge Proof," 2021 IEEE International Conference on Big Data (Big Data), 2021, pp. 3629-3636, doi: 10.1109/BigData52589.2021.9671494.

