

# A Review of Customer Loan Eligibility Prediction Using Machine Learning Algorithms

Divya Evney<sup>1</sup> Prof. Ajit Shrivastava<sup>2</sup> Prof. Rohit Bansal<sup>3</sup>

<sup>1</sup>M.Tech. Scholar <sup>2</sup>Professor <sup>3</sup>Assistant Professor

<sup>1,2,3</sup>Department of Computer Science Engineering

<sup>1,2,3</sup>SISTec-R, Bhopal (M.P), India

**Abstract** — In Loan status prediction is an effective tool for investment decisions in peer-to-peer (P2P) lending market. In P2P lending market, most borrowers full fill the repayment plan; however, some of them fail to pay back their loans. Therefore, a classification method can be utilized to discriminate such default borrowers. In this context, the aim of this dissertation is to propose an investment decision model in P2P lending market which consists of fully paid loans classified via the instance-based machine learning model. Customers who want any loan they apply for that loan. Company validates the customers eligibility for the loan. Company wants to automate the loan eligibility process. For validating such automatic process Gender, Marital Status Number of Dependents in their family, income of family members & finally credit score. To enhance their business in better way these types of facilities, may enhance business as well as customer satisfaction. We also used multiple Machine learning model to generate an investment portfolio based on non-default loans that are predicted to yield high returns. A comparison has been done between the actual and predicted expenses of the prediction premium and eventually, a graph has been plotted on this basis which will enlighten us to choose the best-suited algorithm. The selected algorithm will be applied for our proposed work i.e., Loan Prediction. For prediction, correctness has been measured by the Coefficient of determination. Gradient Boosting Classifier gives the best result in terms of Accuracy i.e. 0.9125 which can be used in its best possible way for the correct prediction of the Loan Prediction Guarantee for companies as well as Customers.

**Keywords:** Machine Learning, Classification, Logistic Regression, Gradient Boosting Machine & Random Forest

## I. INTRODUCTION

A recent development of machine learning techniques and data mining has led to an interest of implementing these techniques in various fields. The banking sector is no exclusion and the increasing requirements towards financial institutions to have robust risk management has led to an interest of developing current methods of risk estimation. Potentially, the implementation of machine learning techniques could lead to better quantification of the financial risks that banks are exposed to.

### A. Credit Risk in Banking

Here The various financial risks banks confront can be broadly classified as credit risk, market risk, liquidity risk and interest rate risk [4]. Author’s explains credit risk as the risk of a debtor defaulting his or her loan, which leads to losses for the lender. Authors elaborates that credit risk includes that a group of borrowers or a counterparty fails to meet its obligations, or an investment deteriorates and defaults and explains that loans are the most common source of credit risk

for banks. However, financial instruments such as bonds, swaps, options and interbank transactions all include credit risk.

### B. Challenges in Data Mining



Fig. 1: Challenges of Datamining

We all Know that Data Mining having huge Challenges due to their requirement and its behaviour. Researchers show it by pictorial Representation.

### C. Data Mining Techniques

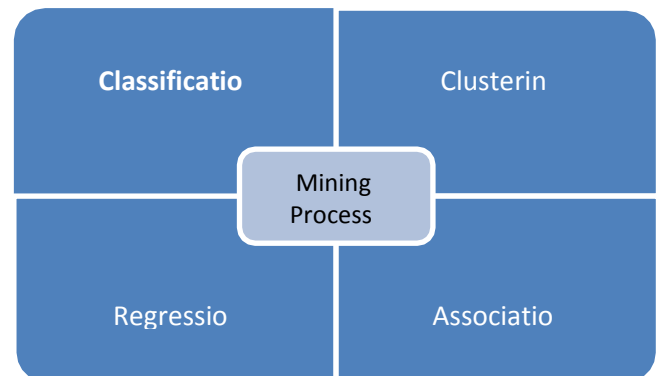


Fig. 2: Data Mining Techniques

Classification is the process of finding a model which describes and distinguishes data classes or concepts based on a class label. There are different classification algorithms some of this are Artificial Neural Network (ANN), Decision tree, Bayesian network, naïve bays etc.

Clustering is the process of analysing data objects without consulting a class label. It is process of grouping new class based on maximizing the intra class similarity and minimizing the interclass similarity. There are different clustering algorithms some of this are K nearest neighbour and k mean clustering.

Association rule learning is machine learning method which used for finding frequent patterns. Some of the association algorithm is Apriori algorithm, Eclat algorithm and FP growth algorithm.

#### D. Scope

The scope of this Dissertation is to implement and investigate how different Machine Learning Algorithm impact default prediction. The model evaluation techniques used in this project are limited to precision, sensitivity, F-score and AUC score. The reasons for choosing these metrics will be explained in more detail When we Explain Result Analysis.

## II. RELATED WORK

Here Here Authors has studied Technology has boosted the existence of humankind the quality of life they live. Every day we are planning to create something new and different. We have a solution for every other problem we have machines to support our lives and make us somewhat complete in the banking sector candidate gets proofs/ backup before approval of the loan amount. The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people applying for the loan in the banking sector but Bank would have limited funds. In this case, the right prediction would be very beneficial using some classes-function algorithm. An example the logistic regression, random forest classifier, support vector machine classifier, etc. A Bank's profit and loss depend on the amount of the loans that is whether the Client or customer is paying back the loan. Recovery of loans is the most important for the banking sector. The improvement process plays an important role in the banking sector. The historical data of candidates was used to build a machine learning model using different classification algorithms. The main objective of this paper is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set. According to this research paper prediction accuracy is sweet for both datasets. In some situations, like client going through some disaster so here the algorithm cannot predict the appropriate result. This research paper can find out the client is potential and repay the loan and the accuracy is good. loan duration, loan amount, age, income are the most important factors for finding out there (whether the client would have been). 'Zip code' and 'credit history' are the foremost important factors for predicting the category of the loan Applicant [5].

Here Researcher's studied Machine learning algorithms are revolutionizing processes in all fields including; real-estate, security, bioinformatics, and the financial industry. The loan approval process is one of the most tedious tasks in the banking industry. Modern technology such as machine learning models can improve the speed, efficacy, and accuracy of loan approval processes. This paper presents six (6) machine learning algorithms (Random Forest, Gradient Boost, Decision Tree, Support Vector Machine, K-Nearest Neighbor, and Logistic Regression) for predicting loan eligibility.

The models were trained on the historical dataset 'Loan Eligible Dataset,' available on Kaggle and licensed under Database Contents License (DbCL) v1.0. The dataset

was processed and analyzed using Python programming libraries on Kaggle's Jupiter Notebook cloud environment. Our research result showed high-performance accuracy, with the Random Forest algorithm having the highest score of 95.55% and Logistic regression with the lowest score of 80%. Our Models outperformed two of the three loan prediction models found in the literature in terms of precision-recall and accuracy. Finally, Authors conclude that as more decision-makers in the financial industry seek to understand ways to improve their processes and maintain a balance between the security and reliability of their financial lending system, machine learning techniques can play a vital role in helping achieve this goal. Our ML models achieved high performance accuracy in predicting loan eligibility in this research. We used the ensemble ML methods (bagging and boosting) and other techniques like SMOTE to ensure optimal predictive models. In general, the methods and algorithms deployed in this research could be instrumental to the successes of financial regulators, corporate, and individual borrowers in their effort to improve their overall loan approval process [6].

Authors studied Nowadays, banks are gaining a lot of customers day by day and many have taken loans from them, and it is not possible to get hold of every customer who is not able to pay or won't be able to pay in future. So now banks are needing classification models to predict in future whether a specific person will be able to pay their loan on time or not. It allows them to take necessary actions in advance like reminding the customer early to pay their debt, or they will face the consequences. In this way, the bank won't suffer losses, and it will help them to maintain their reputation in terms of handling loans in the market. As the people grow the needs also grow which also increases the need to get the basic things. Vehicle is a very basic thing which has become a necessity and to fulfil that people take loans and complete the necessity. Our project will solely be focusing on vehicle loan default with the help of many classification models and choosing the one with better accuracy after a detailed comparison. Finally, Authors Conclude that the amount of people that are being capable to buy the vehicles of all classes are increasing and as more and more banks have started to increase giving the loans and with the less interest rates people are getting attracted and there is a increase in the loan exchange and which gives arise to the frauds that are happening in this area. Hence with the help of the model that has been used we are able to solve this problem and hence we conclude that the naive bayes performed the best amongst support vector machine and decision tree [7].

## III. PROBLEM IDENTIFICATION

Many As we all know that, banks face various financial risks, including different Loan risk. This research focuses on studying default risk, which is one of the credit risk components obliged in the Basel II regulation as explained. Since managing Loan risk is crucial for banks and calculating default risk is obliged, the objective for this research is to understand how loan granting is regulated, and how machine learning is utilized in loan granting.

#### IV. PROPOSED MECHANISM

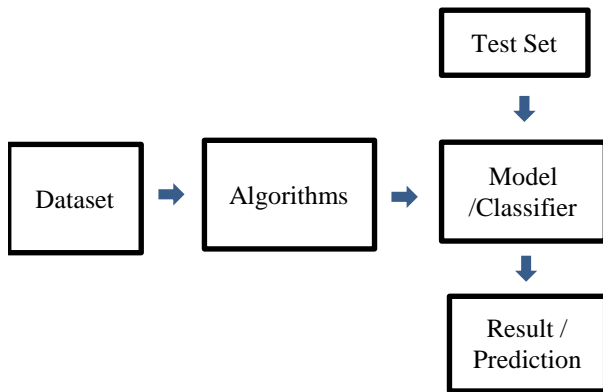


Fig. 3: Proposed Raw Sketch

In the figure 3 we have Dataset that can be created or downloaded from any authentic source. These data is passed to any specific algorithms which will create a different model. These models are verified by given data that is known as test data. Which can able to predict or give result.

#### V. CONCLUSION

Here researches conclude We conclude that when we Implemented Number of Machine Learning Algorithms for finding best results in terms of performance. Modelling our data with Logistic Regression and Gradient Boosting. But Gradient Boosting did a little better as comparison to the Logistic Model. As Training data is less so Gradient Boosting is a good Model, as the cross-Validation Scores also came very good. We used Accuracy, Precision and Recall to evaluate Model Performance. We Finds the major Features of given Data set are Education, Marital Status, Income Group, Living Area. Most Important Factors to Predict the Guarantee of Loan is given above. Finally, we Implemented Different Machine Learning Algorithms Like Linear Regression, Support Vector Machine & Finally our Proposed Method gives results respectively i.e., 83%, 87 % and finally 91%. When we looked into our proposed methods then we can claim that with existing system our proposed methods give better results in terms of Accuracy. Apart from Accuracy we explained various Graphs from which we come for better understanding about our Implemented Models.

#### VI. FUTURE SCOPE

In The future works focus on applying some other techniques to improving the performances of these methods for up to maximum extent. Another concept that can be implemented Deep learning in place of machine learning technology. The reason behind this is best and efficient techniques using nowadays. Deep learning is also introduced nowadays which is becoming more popular for classification purpose. If we have imbalanced data, we need to apply Resampling Techniques to make it balanced. If we have skewed data then we need to apply some Data Transformation techniques. We must perform Univariate and Bivariate Analysis to understand the Better. Finally, try more and more Predictive Models, compare them using various Evaluations Metrics is a good way of finding the best models. So, we can also implement deep learning in future work also.

#### REFERENCES

- [1] JiXiang Niu, "Research on loan prediction based on interpretable machine learning ", 2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)
- [2] Naveen Kumar, D. Keerthana, M Kavitha, M Kalyani "Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector", Proceedings of the Seventh International Conference on Communication and Electronics Systems (ICCES 2022) IEEE Xplore Part Number: CFP22AWO-ART; ISBN: 978-1-6654-9634-6
- [3] Sarthak Choudhary, Chandra Shakher Tyagi, Sakshi Choudhary, Abhijeet Kumar "Loan Payment Date Prediction Model Using Machine Learning Regression Algorithms", 2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDAAI)
- [4] Praveen Tumuluru, Lakshmi Ramani Burra, M. Loukya "Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms" Proceedings of the Second International Conference on Artificial Intelligence and Smart Energy (ICAIS-2022) IEEE Xplore Part Number: CFP22OAB-ART; ISBN: 978-1-6654-0052-7
- [5] Vishal Singh, Ayushman Yadav, Rajat Awasthi "Prediction of Modernized Loan Approval System Based on Machine Learning Approach", 2021 International Conference on Intelligent Technologies (CONIT) Karnataka, India. June 25- 27, 2021
- [6] Ugochukwu.E. Orji, Chikodili.H. Ugwuishiwu, Joseph. C. N. Nguemaleu, Peace. N. Ugwuanyi "Machine Learning Models for Predicting Bank Loan Eligibility" , 978-1-6654-7978-3/22/\$31.00 ©2022 IEEE
- [7] Harshal Patel, Kancharla Nikhil Mohana Sai, Naga Sai Vishal Devulapalli "Vehicle Loan Fraud Prediction using Data Science and Machine Learning Techniques", Proceedings of the Sixth International Conference on Intelligent Computing and Control Systems (ICICCS 2022) IEEE Xplore Part Number: CFP22K74-ART; ISBN: 978-1-6654-1035-9
- [8] Spoorthi B, Shwetha S. Kumar, Anisha P Rodrigues, Roshan Fernandes, Balaji N "Comparative Analysis of Bank Loan Defaulter Prediction Using Machine Learning Techniques", 978-1-6654-1244 ©2021 IEEE
- [9] Sefik Ilkin Serengil, Salih Imece, Ugur Gurkan Tosun "A Comparative Study of Machine Learning Approaches for Non-Performing Loan Prediction", 2021 6th International Conference on Computer Science and Engineering (UBMK) | 978-1-6654-2908-5/21/\$31.00 ©2021 IEEE | DOI: 10.1109/UBMK52708.2021.9558894
- [10] Aman Soni, K.C. Prabu Shankar "Bank Loan Default Prediction Using Ensemble Machine Learning Algorithm" 2022 Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS)
- [11] Anon Mueak "Performance of Loan Repayment Status Model Using Machine Learning", 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information (ECTI-CON)

- [12] C. Prasanth, R. Praveen Kumar, A. Rangesh "Intelligent Loan Eligibility and Approval System based on Random Forest Algorithm using Machine Learning " International Conference on Innovative Data Communication Technologies and Application (ICIDCA-2023) IEEE Xplore Part Number: CFP23CR5-ART; ISBN: 979-8-3503-9720-8
- [13] Salma Khaled Shaheen, Essam ElFakharany "Predictive analytics for loan default in banking sector using machine learning techniques" ICCTA 2018, 30 October – 1 November, Alexandria, Egypt
- [14] Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar "An Approach for Prediction of Loan Approval using Machine Learning Algorithm " Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108-
- [15] Mohammad J. Hamayel, Mobarak A. Abu Mohsen, Mohammed Moreb "Improvement of personal loans granting methods in banks using machine learning methods and approaches in Palestine" 2021 International Conference on Information Technology (ICIT)

