



Prediction of Loan Defaulter Using Machine Learning

Kadam Apurva, Pawar Harshada, Phapale Shweta, Ganeshe Chinmayee

Prof. Neha Hiray

Students of Computer Engineering,

Sandip Institute of Engineering and Management, Nashik

Abstract: With the improving banking loans in recent times and the trend of loans taken, large number of people applies for loans through various sectors. But one of the major problems banking sectors face in this ever-changing economy is the increasing rate of loan defaults, and the banking authorities are finding it more difficult to correctly evaluate loan requests and implement the risks of people defaulting on loans. The purpose of this study is to provide a complete research and to make a model to predict the loan defaulter. This kind of models becomes inevitable as the issue of bad loan is very much critical in the financial sector especially in minimum level financing banks of various underdeveloped and developed countries. We are mainly focusing towards applying data mining methods for the prediction and classification of the loan defaulter. This paper does a comprehensive and comparative analysis between three algorithms (i) Random Forest, and (ii) KNN (iii) XG-Boost. All algorithms will be used on the same dataset. We compile a large dataset, different machine learning algorithms are applied to predict in sample (training sample), and to forecast out-of-sample (testing data). By observing and directing the execution of machine learning to train a model with credit default data to determine the probability (“default” vs “non-default”) of the user’s accountability.

Index Terms - Machine Learning, Loan Defaults, Random Forest, KNN, XG-Boost.

I. INTRODUCTION

Loan Prediction is very helpful for employee of banks as well as for the user also. The main purpose of the paper is to classify and analysis the nature of the loan applicants. It can provide superiority to the bank. As intended analysis of data set and constraints of the banking sector. This paper proposes three machine learning models to forecast whether an individual should be given a loan by evaluating certain attributes and therefore help the banking organization by relieve their process of selecting suitable people from a given list of contenders who applied for a loan. To develop a model to predict the loan default.

The goal of this project is to make a ML model that can predict if a person they fail to pay the loan. The model is calculated to be used as a reference tool for the client and his bank to help make decisions on issuing loans, so that the possibility can be reduced, and the profit can be maximum. In this System we will spontaneously calculate each characteristic which takes part in loan processing and on newly created data test same features are processed with respect to their associated weight .A time limit can be set for the contender to check whether his/her loan can be sanctioned or not. Loan Prediction System allows to move further to specific application so that it can be check on first concern. This Paper is simply for the managing authority of commercial company, whole process of prediction is done privately no stakeholders would be able to alter the processing.

II. MOTIVATION

The Motivation is to build the predictive model for the individual assessment of loan application to determine whether the applicant will default or not. To produce predicative model to group each borrower as defaulter or not using the data collected when the loan has been given. Determining probability of user liability. Generating an interactive UI that will take users details and give back output. The results of these case studies give insight into techniques for accurately predicting Loan Defaulters, compare the accuracy of MI algorithms have been considered for the prediction and classification of student performance respectively using two machine learning algorithms including Random forest,

KNN, XGBoost are implemented to predict the loan defaulter. This dataset for the current study was based on customer's behavior. If the project get successful then it will be great help for faculty to enhance Banking system.

III. OBJECTIVE

- Minimize the risk of borrowers defaulting the loans using created model.
- Create predicative model to classify each borrower as defaulter or not using the data collected when the loan has been given.
- Determining probability of user liability.
- Creating an interactive UI that will take users input and return an output.

IV. LITERATURE SURVEY

This section discusses in short concerning a number of the work that has already been done on making milliliter models AN algorithms to enhance the loan prediction method associate degree facilitate the banking authorities and money corporations choose an eligible candidate with terribly low credit risk.

Many researches are conducted supported data processing and knowledge analyzing within the field of monetary and banking sector. This section presents concisely a number of these techniques that area unit utilized in loans management and their finding Sudhakar et al centered on specifying the information mining applications utility, these applications area unit mistreatment many machine learning algorithms like call trees and Radial Basis Neural Networks. This study came with within which thanks to apply these applications in a very empowerment assessment field. McLeod presents neural networks properties and their fitness for the credit granting method.

Anchal Goyal, Ranpreet Kaur(jan-feb 2019),[1]Loan Prediction exploitation assortment of Technique .The Eleven machine learning models with 9 properties area unit inbuilt the planned work to predict the credit risk of consumers WHO have applied for a loan. This paper conferred associate degree ensemble model for loan predictions mistreatment many parameters like Accuracy to match completely different coaching algorithms. The most goal of this paper is to assess the accuracy of models and to make a group of model that mixes the outputs of 3 completely different models to predict client loan amount. The necessary feature is calculated mistreatment Real Coded Genetic Algorithms. These options supports in predicting a client & credit risk. The K-fold validation predictive is employed to work out the prophetic model dependability, validity, etc. The Maximum accuracy achieved: 81.25%
Model used: Tree model for genetic algorithm.

X.Francis Jency, V.P.Sumathi, Janani Shiva Sri (2018) ,[2]This paper proposed Exploratory Data Analysis (EDA) as a method for predicting loan amounts based on the nature of the client and their needs. The major factors concentrated during the data analysis are Annual income versus loan purpose, customer trust, and loan occupancy versus delinquent months, loan tenure versus credit category, and loan tenure versus credit category, loan tenure versus the number of years in current job, and chances for loan repayment versus homeownership. Eventually, the purpose of this study was to gather the constraints that the customer faces when applying for loan, as well as to make a projection about repayment. Moreover, the results revealed that customers were more interested in current liabilities than long-term loans.

Tarig Mohammed Ahmed, Aboobyda Jafar Hamid(March 2016),[3]In this paper, three algorithms - j48, bayes Net, and naive Bayes - were used to build predictive models that can be used to predict and group loan applications introduced by customers as good or bad by analyzing customer behaviors and previous payback credit. With the help of Weka application the model was built. It was uncover that the best algorithm for loan classification is the J48 algorithm after applying classification & data mining practice algorithms such as j48, Bayes Net, and Naïve Bayes. The J48 algorithm is the best because it is highly accurate and has a low mean absolute error.
The Accuracy achieved: 78.37
Model used: j48

Myneedi Pavani, Pidikiti Supriya, Nagarapu Saisushma, [4] Pre-Processing and knowledge collection, applying machine learning models, testing, and coaching the information were the modules coated during this page. Outlier detection and removal, yet as imputation removal process, were done throughout the pre- process stage. To predict the possibilities of current standing concerning the empowerment method, SVM, DT, KNN, and gradient boosting models were utilized in this methodology. It divides the dataset into training and testing processes, which follows the rule 80:20 was used. Analysis terminated that the choice Tree has rem higher loan exceptional prediction accuracy than the opposite models.
The Accuracy achieved: 0.811
Model used: Decision

Sudhamathy G (Oct-nov2016), [5] Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, Using the R package, this proposed a risk analysis method for sanctioning a loan for customers. The steps involved in developing the model Data selection, pre-processing, feature extraction and selection, building the model, prediction, and evaluation were among. The USI repository provided the dataset for evaluation and prediction. Because the most important and time-consuming step is pre-processing, classification and clustering techniques in R were used to prepare the data for further use. The decision tree classifier was then built using the preprocessed dataset. The Precision achieved: 0.833.

Vimala and Sharmili (2019) [6] planned a loan prediction model mistreatment NB and Support Vector Machines (SVM) strategies. Naïve mathematician, associate degree freelance speculation approach, encompasses applied mathematics concerning the information classification. On the opposite hand, SVM uses applied mathematics learning model for classification of predictions. Dataset from UCI repository with twenty one attributes was adopted to judge the planned methodology. Experimentations terminated that, instead of individual performances of classifiers (NB and SVM), the combination of NB associate degree SVM resulted in an economical, and the integration of NB and SVM resulted in an efficient classification of loan prediction.

Kacheria, Shivakumar, Sawkar and Gupta (2016) [8] suggested a loan sanctioning prediction procedure based on NB approach integrated with K-Nearest Neighbor (KNN) and binning algorithms. The seven parameters considered were Income, age, profession, existing loan with its tenure, amount and approval status. The sub-processes include, Pre-processing, Classification using NB approach and Updating the dataset frequently results in appropriate improvement in the loan prediction process. Experimentation put-forth the conclusion that, integration of KNN and binning algorithm with NB resulted in improved prediction of loan sanctioning process.

A.R.Ghatge, P.P .Halkarnikar(2013)[9] develops the unreal neural network model for predict the credit risk of a bank. The Feed- forward back propagation neural network is employed to forecast the credit default. They additionally compare the results with the manual calculations of the bank conducted in year 2004, 2005 and 2006. The results offer the higher And better performance over manual calculations of bank.

Maher Alaraj, Maysam Abbod, and Ziad Hunaiti (2014) [10] planned a brand new ensemble methodology for classification of costumer loan. This ensemble methodology is predicated on neural network. They state that the planned methodology offer higher results and accuracy as compared to single classifier and the other model.

Marc Claesen, Frank American state Smet, Johan A.K. Suykens, aristocrat American state Moor(2014)[11] planned a model supported support vector machine that reduced the quality of coaching knowledge and predicts the model with high accuracy. This model is employed to avoid duplicate storage of information

Gang Wang, Jian Ma (2011) [12] planned an ensemble approach supported boosting and random topological space and therefore the model named as RS-Boosting for the risky prediction. It offers higher performance. The results shows that the planned approach offers best performance among seven alternative strategies i.e., supply multivariate analysis (LRA), call tree (DT), artificial neural network (ANN), bagging, boosting and random topological space

M. Yaghini , T. Zhiyan , and M. Fallahi (2011)[13] presents a model that's supported feed forward neural networks to spot the unhealthy costumers within the bank. They use 3 completely different ways like fast, dynamic and multiple ways. To stop the model from over fitting, cross validation is completed on the model. To judge the planned model, the results of neural network is compared with some common predictions strategies particularly call tree and supply regression. The results state that the three layer neural network supported the rear propagation learning formula with fast strategy has higher accuracy.

Hussain Ali Bekhet , Shorouq Fathi Kamel Eletter(2014)[14] planned two credit model particularly supply regression model and Radial basis operate evaluation model to support loan call for Jordanian industrial banks using data processing technique. The experimental result shows that the supply regression model performed slightly higher than the radial basis operate model in terms of accuracy.

Suresh Ramakrishnan, Maryam Mirzaei and Mahmoud Bekri (2015) [15] methodology and makes an empirical comparison. The most goal is to match ensemble classifiers. This study explores ADA Boost and material ensemble for default prediction to distinction with many classifiers together with learning supply Regression (LR), call Tree (DT), artificial Neural Networks (NN) and support vector machine (SVM) as base learner..

After going through this literature it is found that loan approval Prediction problem is very important for banking System. Machine learning algorithm are very useful in predicting outcomes even when data is very big in size. This Literature survey investigated some machine learning algorithms and applied ML on test data set of loan approvals.

V. SYSTEM ARCHITECTURE

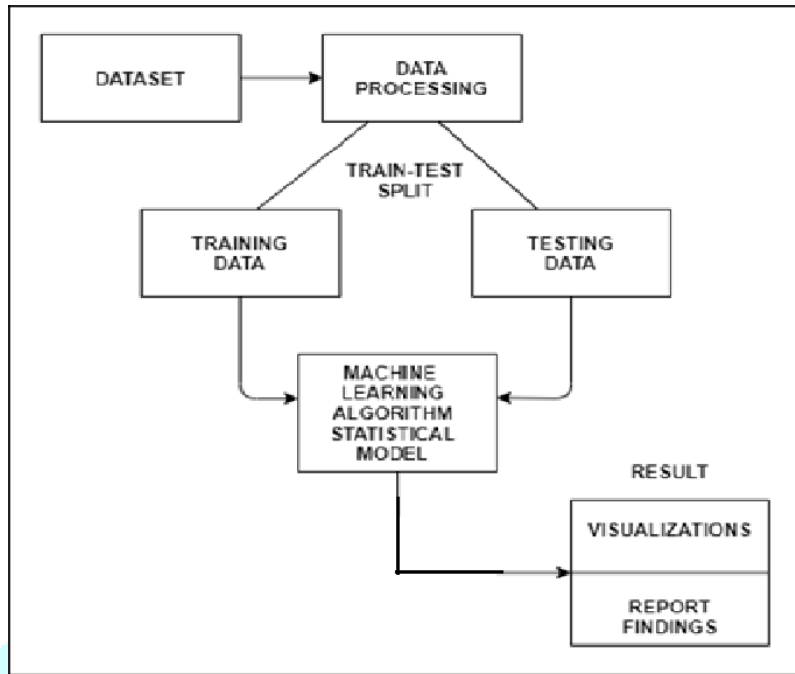


Fig.1.System Architecture

Dataset: A collection of instances is dataset.

Data processing: In this process collected data will be cleaned and required data will split into training and testing

Determine the training and testing data: Training data in this data is use to train an algorithm/model to predict the Outcomes.

Testing data is used to measure the performance such as accuracy or efficiency of the algorithm.

After training and testing data with the help of Machine Learning Algorithm models are formed.

Further we can visualize the result in graphs or histogram etc. and report is delivered to the applicant in step with Their standing.

VI. CONCLUSION

This study has proposed a comprehensive research and model development for the prediction of the default loans. This paper aimed to explore, analyses, and build a machine learning algorithm to correctly identify whether a person, given certain attributes, has a high probability to default on a loan. As the issue related to the high ratio of bad loans is very much critical in the financial sector especially in micro-financing banks of various under develop and developed countries. Although, loan lending has been proven very substantial in the stability of any country's economy in this century such a huge amount of loan defaults is also very critical.

REFERENCES

- [1] Ranpreet Kaur, Anchal Goyal, —"Accuracy Prediction for Loan Risk victimization Machine Learning Models", International Journal of engineering science Trends and Technology (IJEST), Vol. 4, Issue one pp-52-57, jan-feb 2016
- [2] V.P.Sumathi, Janani Shiva Sri, X. Francis Jency, —"An EDA for Loan Prediction supported Nature of the Clients", International Journal of Recent Technology and Engineering (IJRTE), Vol. 7, No. 48, pp. 176-179,2018.-202,2015.
- [3] Tarig Mohammad Ahmed and Aboobyda jafar Hamid, Machine Learning and Applications: a global Journal (MLAJ)—"Developing Prediction Model of Loan Risk in Banks exploitation information Mining", Vol.3, No.1, pp. 1-9, March 2016.
- [4] Nagarapu Saisushma, Namburi Vimala Kumari ,Pidikiti Supriya, Myneedi Pavani, K. Vikas, —"Loan Prediction by exploitation Machine Learning." intelligent system style and application (ISDA), 2010 tenth International conference on, pp.866-870, 2010.
- [5] Sudhamathy, —"Credit Risk Analysis and Prediction Modelling of Bank Loans with facilitate of R", International Journal of Engineering and Technology (IJET), Vol. 8, No. 5, pp. 1954-1966, Oct-Nov 2016.

- [6] S. Vimala, K.C. Sharmili, —Prediction of Loan Risk victimization NB and Support Vector Machinel, International Conference on Advancements in Computing Technologies (ICACT 2018), vol. 4, no. 2, pp. 110-113,2018.
- [7] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, K. Vikas, —Loan Prediction by exploitationofmachinelearning.
- [8] Aditi Kacheria, Nidhi Shivakumar, Shreya Sawkar,Archana Gupta,Loan enabling Prediction System, International Journal of sentimental Computing and Engineering (IJSCE), vol. 6, no. 4, pp. 50-53, 2016.
- [9] A.R.Ghatge, P.P.Halkarnikar, “Ensemble Neural Network Strategy for Predicting Credit Default Evaluation”, Volume2,Issue7,January2013.
- [10] Maher Alaraj, Maysam Abbod, and Ziad Hunaiti, “Evaluating client Loans victimization Neural Networks Ensembles”,Jan.8-9,2014
- [11]Marc Claesen, Frank Delaware Smet, Johan A.K. Suykens, Bart Delaware Moor, “A Library for Ensemble Learning victimizationSupportVectorMachines”,Jan2014.
- [12] Gang Wang, Jian Ma, “Study of company credit risk prediction supported integration boosting and random subspace”, 2011.
- [13] M. Yaghini , T. Zhiyan , and M. Fallahi, “A Prediction Model for Recognition of dangerous Credit Customers insamanBankexploitationNeuralNetworks”,2011
- [14] Hussain Ali Bekhet , Shorouq Fathi Kamel Eletter , “Credit risk assessment model for Jordanian business banks: Neuralmarkingapproach”,Apr2014
- [15] Suresh Ramakrishnan, Maryam Mirzaei and Mahmoud Bekri, “Adaboost Ensemble Classifiers for corporate Default Prediction”, jan 2015.

