

SMART LENDER

APPLICANT CREDIBILITY PREDICTION FOR LOAN

ABSTRACT:

The enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this paper we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i) Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.

KEYWORDS:

Loan Prediction; Big data; Machine Learning; Logistic Regression; SVM; Decision Tree; Naïve Bayes;

INTRODUCTION:

As the data are increasing daily due to digitization in the banking sector, people want to apply for loans through the internet. Artificial intelligence (AI), as a typical method for information investigation, has gotten more consideration increasingly. Individuals of various businesses are utilizing AI calculations to take care of the issues dependent on their industry information. Banks are facing a significant problem in the approval of the loan. Daily there are so many applications that are challenging to manage by the bank employees, and also the chances of some mistakes are high. Most banks earn profit from the loan, but it is risky to choose deserving customers from the number of applications. One mistake can make a massive loss to a bank. Loan distribution is the primary business of almost every bank. This project aims to provide a loan [1, 8] to a deserving applicant out of all applicants. An efficient and non-biased system that reduces the bank's time employs checking every applicant on a priority basis. The bank authorities complete all other customer's other formalities on time, which positively impacts the customers. The best part is that it is efficient for both banks and applicants. This system allows jumping on particular applications that deserve to be <https://forms.gle/XLLutHr3c49szVBr6> approved on a priority basis. There are some features for the prediction like, Gender, Married, Dependents, Education, Self-employed, ApplicantIncome, CoapplicantIncome, LoanAmount, Loan_Amount_Term, Credit_History, Property Area, Loan_Status. Loans have made our life

easier, providing us the financial leverage that extends beyond our earnings. Be it Credit Card, Home Loan, Personal Loan or Auto Loan etc. loans are the credit extended to us by lenders on fulfilling certain key parameters. However, getting a loan in India can often be a tedious process for the un-initiated, but not for individuals with a good credit score. Whenever you apply for a loan, banks check your CIBIL Score and Report to evaluate your credit history and credit worthiness. The higher your score the better are the chances of your loan application getting approved.

LITERATURE SURVEY:

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. Within the credit risk area, there has been a continuous development of the Basel accords, which provides frameworks for supervisory standards and risk management techniques as a guideline for banks to manage and quantify their risks. From Basel II, two approaches are presented for quantifying the minimum capital requirement such as the standardized approach and the internal ratings

based approach (IRB) . There are different risk measures banks consider in order to estimate the potential loss they may carry in future. One of these measures is the expected loss (EL) a bank would carry in case of a defaulted customer. One of the components involved in EL estimation is the probability if a certain customer will default or not. Customers in default means that they did not meet their contractual obligations and potentially might not be able to repay their loans . Thus, there is an interest of acquiring a model that can predict defaulted customers. A technique that is widely used for estimating the probability of client default is Logistic Regression . In this thesis, a set of machine learning methods will be investigated and studied in order to test if they can challenge the traditionally applied techniques. A prediction is a statement about what someone thinks will happen in the future. People make predictions all the time. Some are very serious and are based on scientific calculations, but many are just guesses. Prediction helps us in many things to guess what will happen after some time or after a year or after ten years. Predictive analytics is a branch of advanced analytics that uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions. “Adyan Nur Alfiyatin, Hilman Taufiq and their friends work on the house price prediction. They use regression analysis and Particle Swarm Optimization (PSO) to predict house price”. One other similar work on the Mohamed El Mohadab, Belaid Bouikhalene and Said Safi to predict the rank for scientific research paper using supervised learning. Kumar Arun, Garg

Ishan and Kaur Sanmeet work on bank loan prediction on how to bank approve a loan. They proposed a model with the help of SVM and Neural networks like machine learning algorithms. This literature review helps us carry out our work and propose a reliable bank loan prediction model. Manjeet et al (2018) there are seven types of variables that may influence consumer loan default; consumer's annual income, debt-income ratio, occupation, home ownership, work duration and whether or not consumer possesses a saving/checking account. In a work by Steenackers and Goovaerts, the key factors that may influence loan default are borrower's age, location, resident/work duration, owner of phone, monthly income, loan duration, whether or not applicant works in a public sector, house ownership and loan numbers. Another study by Ali Bangher pour on a large dataset within the period of 2001-2006 indicated that loan age was the most important factor when predicting loan default while market loan-to-value was the most effective factor for mortgage loan applications. In addition to identifying factors that may influence loaned fault, there is also a need to build robust and effective machine learning models that can help capture important patterns in credit data. The choice of model so great importance as the chosen model plays a crucial role in determining accuracy, precision and efficiency of a prediction system. Numerous models have been used for loan default prediction and although there is no one optimal model, some models definitely do better than others. In 2019, Vimala and Sharmili proposed a loan prediction model using and Support Vector Machines(SVM)methods. Naïve Bayes, an

independent speculation approach, encompasses probability theory regarding the data classification. On the other hand, SVM uses statistical learning model for classification of predictions. Dataset from UCI repository with 21 attributes was adopted to evaluate the proposed method. Experimentations concluded that, rather than individual performances of classifiers (NB and SVM), the integration of NB and SVM resulted in an efficient classification of loan prediction. In 2019, Jency, Sumathi and Shiva Sri proposed an Exploratory Data Analysis(EDA) regarding the loan prediction procedure based on the client's nature and their requirements. The major factors concentrated during the data analysis were annual income versus loan purpose, customer's trust, loan tenure versus delinquent months, loan tenure versus credit category, loan tenure versus number of years in the current job, and chances for loan repayment versus the house ownership. Finally, the outcome of the present work was to infer the constraints on the customer who are applying for the loan followed by the prediction regarding the repayment. Further, results showed that, the customers were interested more on availing short-tenure loans rather than long-tenure loans. In 2019, Supriya, Pavani, Saisushma, Vimala Kumari and Vikas presented a ML based loan prediction model. The modules in the present approach were data collection and pre-processing, applying the ML models, training followed by testing the data. During the pre-processing stage, the detection and removal of outliers and imputation removal processing were carried out. In the present method, SVM, DT, KNN and gradient boosting models were employed to predict the

possibilities of current status regarding the loan approval process. The conventional 80:20 rule was adopted to split the dataset into training and testing processes. Experimentation concluded that, DT has significantly higher loan prediction accuracy than the other models. In 2017, Goyal and Kaur presented a loan prediction model using several Machine Learning (ML) algorithms. The dataset with features, namely, gender, marital status, education, number of dependents, employment status, income, co applicant's income, loan amount, loan tenure, credit history, existing loan status, and property area, are used for determining the loan eligibility regarding the loan sanctioning process. Various ML models adopted in the present method includes, Linear model, Decision Tree (DT), Neural Network (NN), Random Forest (RF), SVM, Extreme learning machines, Model tree, Multivariate Adaptive Regression Splines, Bagged Cart Model, NB and TGA. When evaluated these models using Environment in five runs, TGA resulted in better loan forecasting performance than the other methods. In 2016, Aboobyda Jafar Hamid and Tarig Mohammed Ahmed presented a loan risk prediction model based on the data mining techniques, such as Decision Tree (J48), Naïve Bayes (NB) and BayseNet approaches. The procedure followed was training set preparation, building the model, Applying the model and finally. Evaluating the accuracy. This approach was implemented using Weka Tool and considered a dataset with eight attributes, namely, gender, job, age, credit amount, credit history, purpose, housing, and class. Evaluating these models on the dataset, experimental results concluded that, J48 based loan prediction

approach resulted in better accuracy than the other methods. In 2016, Kacheria, Shivakumar, Sawkar and Gupta 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, Preprocessing (handling the missing values with KNN and data refinement using binning algorithm), 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. In 2016, Goyal and Kaur suggested an ensemble technique based loan prediction procedure for the customers. The sub processes in the present method includes, data collection, filtering the data, feature extraction, applying the model, and finally analysis the results. The various loan prediction procedures implemented in the present method were Random Forest (RF), SVM and Tree model with Genetic Algorithm (TGA). The parameters considered for evaluating the models were accuracy, Gini Coefficient, Area Under Curve (AUC), Receiver Operating Curve (ROC), Kolmogorov - Smirnov (KS) Chart, Minimum Cost - Weighted Error Rate, Minimum Error Rate, and K-Fold Cross Validation parameters. Experimentation outcome concluded that the integration of three methods (RF, SVM and TGA) resulted in improved loan - prediction results rather than individual method 's prediction. In 2006,

Sudhamathy suggested a risk analysis method in sanctioning a loan for the customers using R package. The various modules include data selection, pre-processing, feature extraction and selection, building the model, prediction followed by the evaluation. The dataset used for evaluation in this method was adopted from UCI repository. To fine tune the prediction accuracy, the pre-processing operation includes the following sub-processes: detection, ranking and removal of outliers, removal of imputation, and balancing of dataset by proportional bifurcation regarding testing and training process. Further, feature selection process improves the prediction accuracy. When evaluated, the DT model resulted in 94.3% prediction accuracy. The process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of knowledge discovery process. The various steps involved in extracting knowledge from raw data. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to classification technique. This approach frequently employs Decision tree based classification Algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.

TITLE 1: Improving Information Quality in Loan Approval Processes for Fair Lending and Fair Pricing

AUTHOR: M. Cary Collins

YEAR: 2013

DESCRIPTION: Bank data management on loan approval processes has great room for improvements of information quality and data problems prevention especially with regards to fair lending and fair pricing practices. They first reviewed briefly typical data collection protocols deployed at many financial institutions for loan approval and loan pricing. Federal regulations mandate portions of these data protocols. While discussing the data capture and analysis for fair lending, they illustrated some initial key steps currently needed for improving information quality to all parties involved.

TITLE 2: Loan Credibility Prediction System Based on Decision Tree Algorithm

AUTHOR: Sivasree M S, Rekha Sunny T

YEAR: 2015

DESCRIPTION: Data mining techniques are becoming very popular nowadays because of the wide availability of huge quantity of data and the need for transforming such data into knowledge. Data mining techniques are implemented in various domains such as retail industry, biological data analysis, intrusion detection, telecommunication industry and other scientific applications. Techniques of data mining

are also be used in the banking industry which help them compete in the market well equipped. In this paper, they introduced a prediction model for the bankers that will help them predict the credible customers who have applied for a loan. Decision Tree Algorithm is being applied to predict the attributes relevant for credibility. A prototype of the model has been described in this paper which can be used by the organizations for making the right decisions to approve or reject the loan request from the customers.

TITLE 3: Loan Approval Prediction based on Machine Learning Approach

AUTHOR: Kumar Arun, Garg Ishan, Kaur Sanmeet

YEAR: 2016

DESCRIPTION: With the enhancement in the banking sector, lots of people apply for bank loans but the bank has its limited assets which it grants to only limited people, so finding out to whom the loan can be granted is a typical process for the banks. So, in this paper, they tried to reduce this risk by selecting the safe person so as to save lots of bank efforts and assets. It was done by mining the previous records of the people to whom the loan was granted before and on the basis of these records the machine was trained using the machine learning model which gave the most accurate result. The main goal of this paper is to predict if loan assignment to a specific person will be safe or not. This paper has into four sections (i) Collection of data (ii) Comparing

the machine learning models on collected data (iii) Training the system on most promising model (iv) Testing the system.

EXISTING SYSTEM:

Anomaly detection relies on individuals' behaviour profiling and works by detecting any deviation from the norm. When it is used for online banking fraud detection, it suffers from three disadvantages. First, for an individual, the historical behaviour data are often too limited for profiling his/her behaviour pattern. Second, because of the heterogeneous nature of transaction data, there is no uniform treatment to various attribute values, which will become a potential barrier for development of the model and for further usage. Third, the transaction data are highly skewed, and it becomes a challenge for utilizing the label information effectively. Anomaly detection often suffers from poor generalization ability and a very high false alarm rate. We argue that individuals' limited historical data for behaviour profiling and fraud data's highly skewed nature could account for this defect. Since it is straightforward to use information from other similar individuals, similarity measurement itself becomes a great challenge due to heterogeneous nature of attribute values.

A. Advantages

1. Performance and accuracy of the algorithms can be calculated and compared.
2. Class imbalance can be dealt with machine learning approaches.

B. Disadvantages

1. They had proposed a mathematical model and machine learning algorithms were not used.
2. Class Imbalance problem was not addressed and the proper measure were not taken.

C. Proposed System

In our proposed system, we combine datasets from different sources to form a generalized dataset and use four machine learning algorithms such as Random forest, Logistic regression, Decision tree and Naive bayes algorithm on the same dataset .The dataset we collected for predicting given data is split into training set and test set in the ratio of 7:3. The data model which was created using Machine learning algorithms are applied on training set and based on maximum test result from the four algorithms, the test set prediction is done using the algorithm that has maximum performance. After that, we deploy the model using Flask Framework.

TECHNOLOGIES USED:

Python,Python Web Frame Works,Python For Data Visualization,Data Preprocessing Techniques,Machine Learning,IBM Cloud,IBM Watson Studio,Python-Flask.

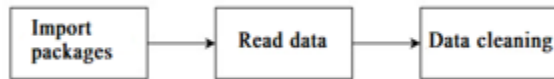
LIST OF MODULES:

- A. Data Pre-processing
- B. Data Analysis of Visualization
- C. Comparing Algorithms
- D. Deployment Using Flask

(1) DATA PRE-PROCESSING:

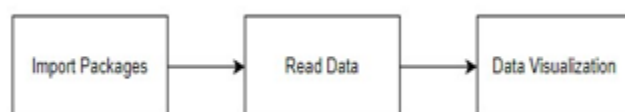
Data Pre-processing is a technique used for converting the raw data into a clean data set. Whenever a data is gathered from different sources, it is collected in raw format which is not feasible for the analysis of the model. For achieving better results, the model in Machine Learning method of the data has to be in a proper manner. Some of the specific Machine Learning model needs information to be in a particular format, for example, Random Forest algorithm don't support null values. Therefore, to proceed further with random forest algorithm, null values need to be managed from the original raw dataset. In data pre-processing, we carry out data cleaning task with the help of Python's Pandas library. Data cleaning is a process of removing missing, incomplete or duplicate data. The steps and techniques used for data cleaning will vary from each dataset. When combining multiple datasets, there are so many possibilities for the data to be duplicated or mislabeled or incomplete. The main objective of

data cleaning is to detect and remove errors to increase the value of data in analytics and decision making to get accurate outputs.



(2) DATA ANALYSIS OF VISUALIZATION:

Data visualization is a technique helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupted data, outliers, and much more. It can be used to express and demonstrate main relationships between plots and charts that are more visceral. Sometimes data does not make sense until we can look at in a visual form, such as with charts, graphs and plots. Being able to quickly visualize the data samples is an important skill both in applied statistics and in applied machine learning. It can also be used to remove outliers to get better and accurate results. It is implemented by using Python's Matplotlib library.



(3) COMPARING ALGORITHMS:

Before comparing algorithms, we build a Machine Learning Model using Python's Scikit-Learn libraries. In this library package, we must do pre-processing, linear model with logistic regression method, cross validating by K-Fold method, ensemble with random forest method and tree with decision tree classifier. Additionally, we

split the train set and test set in order to predict the result by comparing accuracy.

To find the algorithm with best performance, we use the following performance metrics:

A. Confusion Matrix

Confusion matrix is one of the performance metrics used to find the correctness and accuracy of the model. It has the following four parameters:

B. False Positives (FP)

A person who will pay is predicted as defaulter where actual class is no, and predicted class is yes. E.g., if actual class says a person is not married but predicted class tells that the person is married.

C. False Negatives (FN)

A person who will pay is predicted as defaulter where actual class is yes but predicted class is no. E.g., if actual class value indicates that the person is married, and predicted class tells that person is not married.

D. True Positives (TP)

A person who will not pay is predicted as defaulter where the value of actual class is yes, and the value of predicted class is also yes. E.g., if actual class value indicates that the person is married, and predicted class also tells the same thing.

E. True Negatives (TN)

A person who pays is predicted as defaulter where the value of actual class is no, and value of predicted class is also no. E.g., if actual class says the person is not married and predicted class also tells the same thing.

True Positive Rate (TPR) = True Positives / (True Positives + False Negatives)

False Positive Rate (FPR) = False Positives / (False Positives + True Negatives)

F. Accuracy

Accuracy is the most important performance metrics which is the ratio of observations that are correctly predicted to the total observations. Higher accuracy means that the model produces accurate results but only when we have symmetric datasets where values of false positive and false negatives are almost the same.

Accuracy = (TP+TN) / (FP+FN+ TN+TP)

G. Precision

Precision is the ratio of positive observations correctly predicted to the positive observations totally predicted. High precision rates relates to the low false positive rates of the dataset. We have got 0.876 precision which is really good.

Precision = TP / (FP+TP)

H. Recall

Recall is the ratio of correctly predicted positive observations to the all observations in actual class – yes i.e., the proportion of positively observed values correctly predicted which is nothing but the proportion of actual defaulters that the model will correctly predict.

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

I. F1 Score

F1 Score is basically the average weight of Precision and Recall. Therefore, the F1 score takes the values of both false positives and false negatives into consideration. F1 score can be found out if there is an uneven class distribution. If the values of false positives and false negatives are too different, it's better to have a look at both Precision and Recall.

$$\text{F1 Score} = \{(\text{Precision} * \text{Recall}) * 2\} / (\text{Precision} + \text{Recall})$$

(4) LOGISTIC REGRESSION:

Logistic regression is a machine learning classification supervised algorithm that is used for predicting the probability of a categorical dependent variable. It is a statistical method that is used for analysing a dataset where there are one or more independent variables which determines an outcome. The outcome is measured with a dichotomous variable (which means there are only two possible outcomes). The primary objective of logistic regression is to find the best fitting model for describing the relationship between dependent variables and a set of independent variables. In logistic regression, the

dependent variable is binary that contains data which is coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).



(5) DECISION TREE CLASSIFIER:

Decision tree is used for building classification models in the form of a tree structure. It basically breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. A decision node has two or more branches which ends in another decision node or a leaf node, and a leaf node represents the final decision. The topmost decision node in a tree called root node corresponds to the best predictor. Decision trees are used for handling both categorical and numerical data. It uses an if-then rule set which is mutually exclusive and exhaustive for classification of datasets. The rules are being learnt sequentially using the training dataset one at a time. Every time a rule is learned, the tuples which are covered by the rules are being removed. This process is continued on the training set until it reaches a terminating condition. It is constructed in a top-down recursive divide-and-conquer manner and all attributes in the dataset should be categorical. Or else, they should be discretized in advance. Attributes in the top of the tree have more impact towards classification which are identified using the information gain concept.



(6) RANDOM FOREST ALGORITHM:

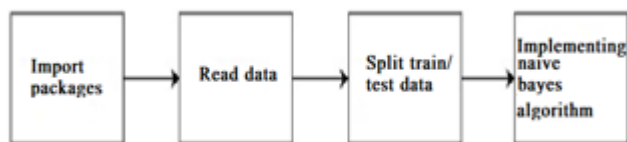
Random forest is one of the supervised machine learning algorithm which is based on ensemble learning method. Ensemble learning is a type of learning method where we can join different types of algorithms or join same algorithm multiple times to form a very powerful prediction model. It operates by constructing a multitude of decision trees at training time and outputting the class i.e., the mode of the classes of the individual trees. Random decision forests correct the decision trees' habit of over fitting their training set. In classification problems, each tree in the forest is used for predicting the category to which the new record belongs. Finally, the new record is being assigned to the category that wins the majority of the votes.



(7) NAIVE BAYES CLASSIFIER:

Naive Bayes is a supervised machine learning statistical classification technique based on Bayes Theorem. It is one of the most simplest supervised machine learning algorithms. Naive Bayes classifier is the fast, accurate and reliable algorithm among other algorithms. Naive Bayes classifiers have very high accuracy and speed on very large datasets. Naive Bayes classifier assumes that the effect

of a specific feature in a class of a dataset is always independent of other features. For example, a loan applicant gets loan or not depending on his/her income, previous loan and transaction history, age, and location. Even if the above features are interdependent, these features are still considered independent. This assumption simplifies computation which is the reason why it is considered as naive. The above assumption is called class conditional independence. The probability of a class value when the value of an attribute is given is called the conditional probability. By multiplying the conditional probabilities together for each attribute for a given class value of a dataset, we have a probability of a data instance belonging to that specific class. To make a prediction, we need to calculate probabilities of the instance belonging to each class and select the class value that has the highest probability.



(8) DEPLOYMENT USING FLASK:

After finding the algorithm that has the maximum performance among the four algorithms using performance metrics, we convert that model into a PKL file and then we deploy that model using FLASK framework in Python to build a user interface. So, when a customer enters his/her details in the user interface, and enters the submit button, he/she will get the result if a loan can be approved or not for that person.



(9) FUTURE WORKS:

We can make the Bank Loan Approval prediction to connect with Cloud for future use to optimize the work to implement in Artificial Intelligence environment.

CONCLUSION:

This application is working properly and meeting to all Banker requirements. This component can be easily plugged in many other systems. It works correctly and fulfills all requirements of bankers and can be connected to many other systems. There were multiple malfunctions in the computers, content errors and fixing of weight in computerized prediction systems. In the near term, the banking software could be more reliable, accurate, and dynamic in nature and can be fit in with an automated processing unit. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system more secure, reliable and dynamic weight adjustment. The system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time. Machine learning helps to understand the factors which affect the specific outcomes most. Other models like neutral network and

discriminate analysis can be used individually or combined for enhancing reliability and accuracy prediction.

REFERENCES:

- [1] Arun Kumar, Ishan Garg, and Sanmeer Kaur, \"Loan Approval Prediction Using Machine Learning Approach,\" 2018.
- [2] K. Hanumantha Rao, G. Srinivas, A. Damodhar, and M. Vikas Krishna at International Journal of Computer Science and Telecommunications published an article titled \"Implementation of Anomaly Detection Technique Using Machine Learning Algorithms\" (Volume2, Issue3, June 2011).
- [3] G. Arutjothi and C. Senthamarai, \"Prediction of loan status in commercial banks using machine learning classifier,\" International Conference on Intelligent Sustainable Systems (ICISS), 2017.
- [4] \"AzureML based analysis and prediction of loan applicants creditworthy,\" by Alshouiliy K, Alghamdi A, and Agrawal D P I n 2020, Third International conference on information and computer technologies.
- [5] \"Developing prediction model of loan risk in banks using data mining Machine Learning and Applications,\" Hamid A J and Ahmed T M, 2016.

[6] M. Li, A. Mickel, and S. Taylor \"Should this loan be approved or denied?\" published a paper in the Journal of Statistics Education in 2018.

[7] A. Vinayagamoorthy, M. Somasundaram, and C. Sankar, \"Impact of Personal Loans Offered by Banks and Non-Banking Financial Companies in Coimbatore City,\" 2012.

[8] M. Cary Collins, Ph.D., and Frank M. Guess, Ph.D., MIT's Information Quality Conference, 2000, \"Improving information quality in loan approval processes for fair lending and fair pricing.\"

[9] Arun Kumar, Ishan Garg, and Sanmeet Kaur, \"Loan approval prediction based on machine learning approach,\" National Conference on Recent Trends in Computer Science and Information Technology, 2016.

[10] Sivasree M S and Rekha Sunny T, \"Loan Credibility Prediction System Using Decision Tree Algorithm,\" International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 09, September-2015.

[11] <https://machinelearningmastery.com/random-forest-ensembles-with-xgboost/>

[12] <https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057>

[13] <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>