

Smart Lender - Applicant Credibility Prediction for Loan Approval

1. Anant Shinde (2019)

Abstract:

As the needs of people are increasing, the demand for loans in banks is also frequently getting higher every day. Banks typically process an applicant's loan after screening and verifying the applicant's eligibility, which is a difficult and time-consuming process. In some cases, some applicants default and banks lose capital. The machine learning approach is ideal for reducing human effort and effective decision making in the loan approval process by implementing machine learning tools that use classification algorithms to predict eligible loan applicants

Literature Survey

According to the authors, the forecasting process begins with data clean-up and processing, missing value substitution, data set experimental analysis, and modelling, and continues to model evaluation and test data testing. A logistic regression model has been executed. The highest accuracy obtained with the original dataset is 0.811. Models are compared based on performance measurements such as sensitivity and specificity. As a result of analysing, the following conclusions were drawn. However, other characteristics of customers that play a very important role in lending decisions and forecasting defaulters should also be

2.Shubham Nalawade(2018)

Abstract:

Today a lot of people/companies are applying for bank loans. The core business part of every bank is the distribution of loans. The main objective of the banking sector is to give their assets in safe hands. But the banks or the financial companies take a very long time for the verification and validation process and even after going through such a regress process there is no surety that whether the applicant chosen is deserving or not. To solve this problem, we have developed a system in which we can predict whether the applicant chosen will be a deserving applicant for approving the loan or not. The system predicts on the basis of the model that has been trained using machine learning algorithms. We have even compared the accuracy of different machine learning algorithms. We got a percentage of accuracy ranging from 75-85% but the best accuracy we got was from Logistic Regression i.e., 88.70% The system includes a user interface web application where the user can enter the details required for the model to predict. The drawback of this model is that it takes into consideration many attributes but in real life sometimes the loan application can also be approved on a single strong attribute, which will not be possible using this system.

3.Suraj Andhe(2017)

Abstract:

Finally, we apply the Naive Bayesian formula to find the frequency table based on the responses which we have to predict further.

4.Prof. Amruta Sankhe(2020)

Abstract:

The less the variance, the less is the fluctuation of scores and the model will be more precise and stable. Only the K Nearest Neighbor Classifier is used in [5]. The process of Min-Max Normalization is used. It is a process of decomposing the attributes values. The highest accuracy they got was 75.08% when the percentage of dataset split was 50-50% with k to be set as 30. In [6] Logistic Regression is the only algorithm used. They didn't calculate the accuracy of the algorithm.

5.Siddhesh Parab(2013)

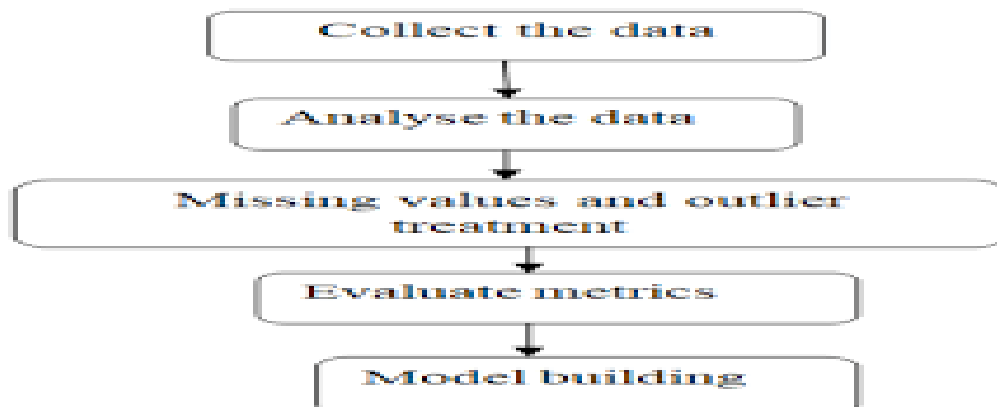
Abstract:

This algorithm is also used for classification as well as regression problems. This algorithm consists of a tree diagram wherein the leaf nodes resemble the class label and internal nodes specify the attributes. The motive of using the decision tree is to train a model that will help to predict the class or value of the target variable by remembering simple decision tree rules which we get from the training data. Decision tree uses many specific algorithms to split a node into various sub nodes.

6.Soni PM, Varghese Paul(2018)

Abstract:

Now a day's people approach or select bank loans to fulfill their needs, which are very common. This practice has been increasing day by day especially for business, education, marriage, agriculture as well. But several people take advantage and misuse the facilities given by the bank. With technology developing at such a peak stage in these days, data mining plays a key role in computer science to solve such issues. Classification is the most suitable predictive modeling technique in data mining to predict the loan repayment capability of a customer in a banking industry. There are various methods to improve the accuracy of a classification algorithm. The accuracy of random forest classification algorithm can be improved using Ensemble methods, Optimization techniques and Feature selection. Various feature selection methods are available.



7.Nidhi Singh(2016)

Abstract:

The cost of assets is increasing day by day and the capital required to purchase an entire asset is very high. So purchasing it out of your savings is not possible. The easiest way to get the required funds is to apply for a loan. But taking a loan is a very time consuming process. The application has to go through a lot of stages and it's still not necessary that it will be approved. To decrease the approval time and to decrease the risk associated with the loan many loan prediction models were introduced. The aim of this project was to compare the various Loan Prediction Models and show which is the best one with the least amount of error and could be used by banks in real world to predict if the loan should be approved or not taking the risk factor in mind. After comparing and analysing the models, it was found that the prediction model based on Random Forest proved to be the most accurate and fitting of them all.



8.Udaya Bhanu(2019)

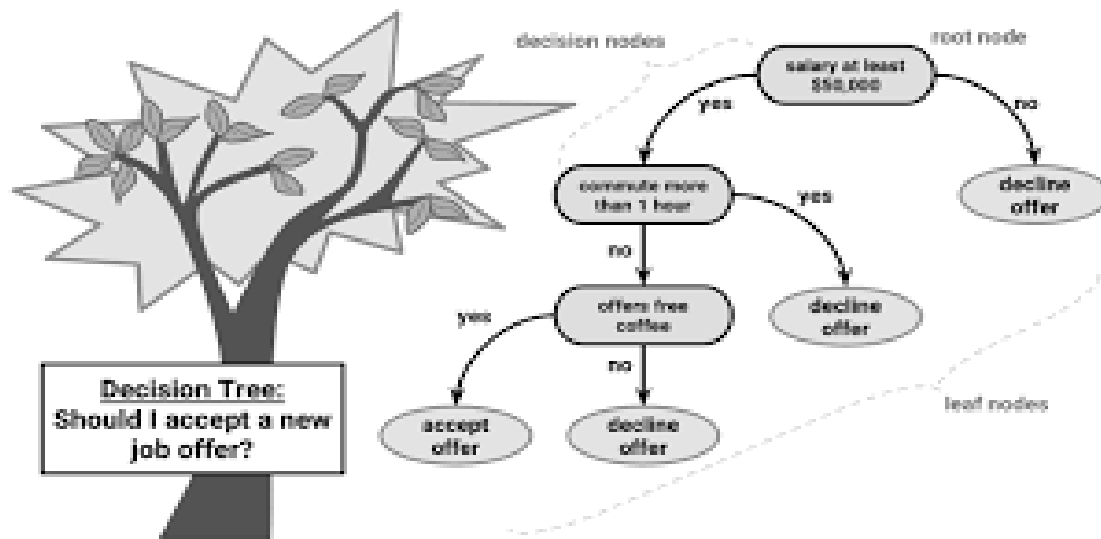
Abstract:

Customer loan prediction is usually life time issue so; each and every retail bank faces the issue at the minimum lifetime. If done exactly, it can spare a lot's of man hours at the conclusion of a retail bank. If Company wants to semi automate the loan acceptability process (real time) based on customer detail provided while filling online application form. These subtle elements are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this method, they have given an issue to recognize the customers segments; those are allowed for loan amount total so they can clearly target these customers. We need to predict whether or not a loan would be approved. In a classification problem, we need to predict separate values based on a given set of self-sufficient variable(s). What's our objective is to implement machine learning model so as to classify, to the best doable degree of accuracy, and dataset gathered from Kaggle. Random forest classification method shows best

9.Dr. S. Narayana(2021)

Abstract:

supervised learning techniques for loan candidate as a valid or fail to pay customer. In this paper, various algorithms were implemented to predict customer loan. Optimum results were obtained using Logistic Regression, Random Forest, KNN, and SVM, decision Tree Classifier. Compare these five algorithms random forest is the high accuracy. From a correct analysis of positive points and constraints on the part, it can be safely ended that the merchandise could be an extremely efficient part. This application is functioning properly and meeting to all or any Banker necessities. This part is often simply obstructed in several different systems. There are numbers cases of computer glitches, errors in content and most significant weight of option is mounted in machine-driven prediction system, therefore within the close of future.

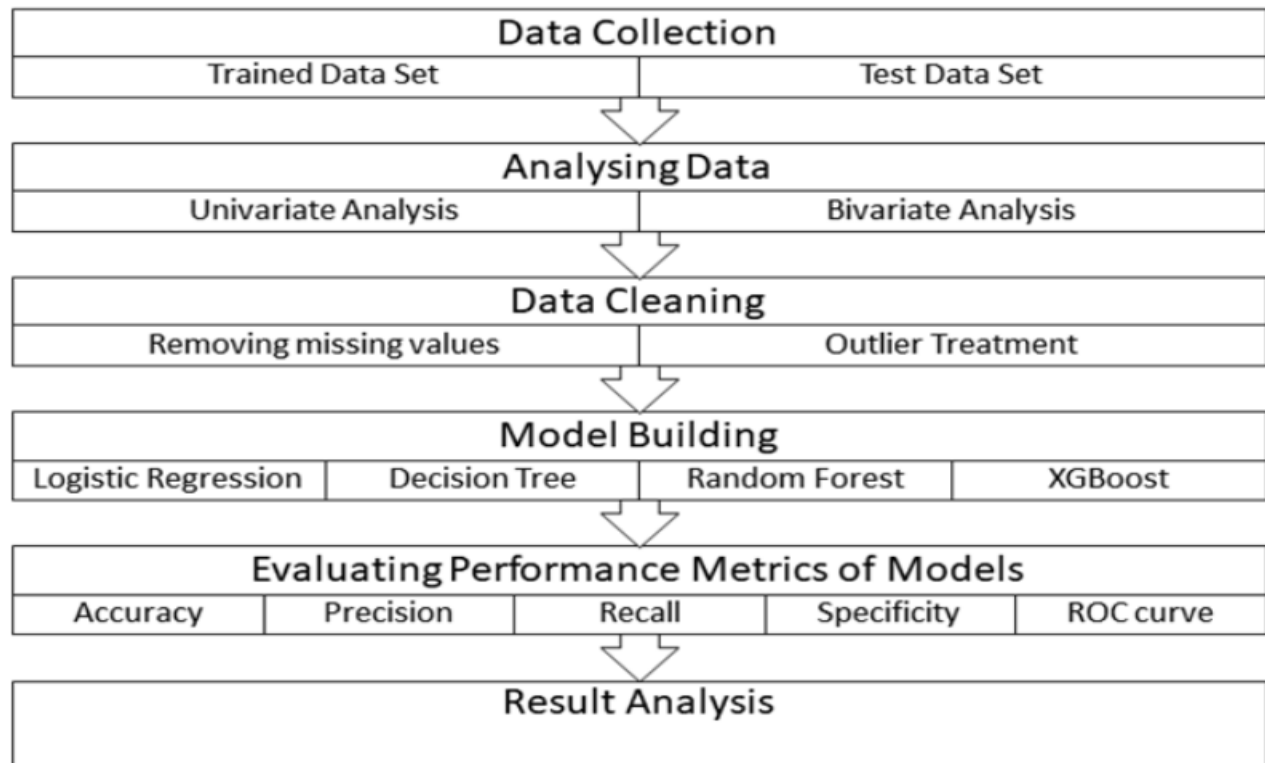


10.Rhishab Mukherjee(2021)

Abstract:

We start our literature review with more general systematic literature reviews that focus on the application of machine learning in the general field of Banking Risk Management. Since the global financial crisis, risk management in banks has to take a major role in shaping decision-making for banks. A major portion of risk management is the approval of loans to promising candidates. But the black-box nature of Machine learning algorithms makes many loan providers vary the result.

Methodology:



Machine Learning and Concepts:

Four machine learning models have been used for the prediction of loan approvals. Below are the description of the models used. This algorithm only works with the quantitative variable. It is a gradient boosting algorithm which forms strong rules for the model by boosting weak learners to a strong learner. It is a fast and efficient algorithm which recently dominated machine learning because of its high performance and speed.

Conclusion:

We did Exploratory data Analysis on the features of this dataset and saw how each feature is distributed.

We did bivariate and multivariate analysis to see impact of one another on their features using charts.

We analysed each variable to check if data is cleaned and normally distributed. We cleaned the data and removed NA values. We also generated hypothesis to prove an association among the Independent variables and the Target variable. And based on the results, we assumed whether or not there is an association. We calculated correlation between independent variables and found that applicant income and loan amount have significant relation. We created dummy variables for constructing the model. We constructed models taking different variables into account and found through odds ratio that credit history is creating the most impact on loan giving decision. Finally, we got a model with coapplicant income and credit history as independent variable with highest accuracy. We tested the data and got the accuracy of 83 .

References :

- [1] M. Sheikh, A. Goel, T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," International Conference on Electronics and Sustainable Communication Systems (ICESC), (2020).
- [2] S. M S, R. Sunny T, "Loan Credibility Prediction System Based on Decision Tree Algorithm," International Journal of Engineering Research & Technology (IJERT) Vol. 4 Issue 09, (2015).
- [3] A. Kumar, I. Garg and S. Kaur, "Loan Approval Prediction based on Machine Learning Approach," IOSR Journal of Computer Engineering, (2016).
- [4] Dr K. Kavitha, "Clustering Loan Applicants based on Risk Percentage using K-Means 6Clustering Techniques," IJARCSSE - Volume 6, Issue 2, (2016).
- [5] P. Dutta, "A STUDY ON MACHINE LEARNING ALGORITHM FOR ENHANCEMENT OF LOAN PREDICTION", International Research Journal of Modernization in Engineering Technology and Science, (2021).
- [6] G. Arutjothi, Dr C. Senthamarai, "Prediction of Loan Status in Commercial Bank using Machine Learning Classifier," Proceedings of the International Conference on Intelligent

Sustainable Systems, (2017).

[7] P. Supriya, M. Pavani, N. Saisushma, N. Kumari and K. Vikas, "Loan Prediction by using Machine Learning Models," International Journal of Engineering and Techniques, (2019).

[8] R. Salvi, R. Ghule, T. Sanadi, M. Bhajibhakare, "HOME LOAN DATA ANALYSIS AND VISUALIZATION," International Journal of Creative Research Thoughts (IJCRT), (2021).

[9] B. Srinivasan, N. Gnanasambandam, S. Zhao, R. Minhas, "Domain-specific adaptation of a partial least squares regression model for loan defaults prediction," 11th IEEE International Conference on Data Mining Workshops, (2011).

[10] M. V. Reddy, Dr B. Kavitha, "Neural Networks for Prediction of Loan Default Using Attribute Relevance Analysis," International Conference on Signal Acquisition and Processing, (2010).