IBM SMARTINTERNZ

Project Title: Smart Lender - Applicant Credibility Prediction for Loan Approval

Literature Survey:

Survey 1

Title: Loan Credibility Prediction System using Data Mining Techniques

Author: **Anuja Kadam, Pragati Namde, Sonal Shirke, Siddhesh Nandgaonkar, Dr.D.R. Ingle**

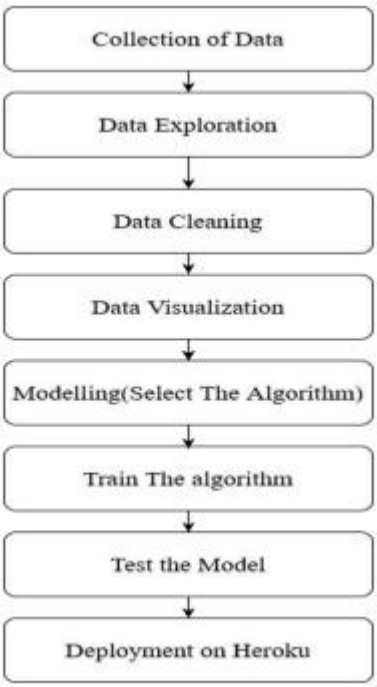
Published: **International Research Journal of Engineering and Technology (IRJET)**

Description:

The prime goal in banking domain is to invest their assets in safe hands. Lending money to unsuitable loan applicants results in the credit risk. Today many banks approve loans after a long procedure of verification, yet there is no guarantee whether the picked candidate is the right candidate or not. Estimating the risk, which is involved in a loan application, is one of the most significant concerns of the banks in order to survive in the highly competitive market. Data mining algorithms are used to study the loan-approved data and exact patterns, which would help in predicting the reasonable defaulters, thereby helping the banks for making better choices in the future. Loan Prediction is extremely useful for employee of banks and for the applicant also. The main aim of this model is to provide a speedy, immediate and simple approach to pick the deserving applicants.

There are numerous areas in which data mining can be used in the banking industry, which include customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, cash management and forecasting operations, optimizing stock portfolios, and ranking investments. In addition, banks may use data mining to identify their most profitable credit card customers or high-risk loan applicants. To help bank to retain credit card customers, data mining is used. By analysing the past data, data mining can help banks to predict customers that are likely to change their credit card affiliation so they can plan and launch different special offers to retain those customers. Credit card spending by customer groups can be identified by using data mining.

Flow Model:



Features of dataset:

Features	Descriptions
Loan_ID	Unique Loan ID
Gender	Male/Female
Married	Married(Yes)/not married(No)
Dependents	Number of dependents
Education	Education(Graduate/Not graduate)
Self_Employed	Self employed(Yes/No)
ApplicantIncome	Applicant income
CoapplicantIncome	Co-applicant income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	Credit history meets guidelines
Property_Area	Urban/semi urban/rural
Loan_Status	Loan approval(Y/N)

Here, we had implemented loan credibility prediction system that helps the organizations in making the right decision to approve or reject the loan request of the customers. In this model, Logistic Regression algorithm is used for the prediction. Incorporation of other techniques that outperform the performance of popular data mining models has to be implemented and tested for the domain.Survey.

Title: Analysis of Loan Availability using Machine Learning Techniques

Author: Sharayu Dosalwar, Ketki Kinkar, Rahul Sannat, Dr Nitin Pise

Published: International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Description:

In the banking system, banks have a variety of products to provide, but credit lines are their primary source of revenue. As a result, they will profit from the interest earned on the loans they make. Loans, or whether customers repay or default on their loans, affect a bank's profit or loss. The bank's Non-Performing Assets will be reduced by forecasting loan defaulters. As a result, further investigation into this occurrence is essential. Because precise forecasts are essential for benefit maximisation, it's crucial to analyse and compare the various methodologies. Banks have many products to sell in our banking system, but their main source of income is their credit lines. As a result, they are likely to profit from the interest on the loans they make. Loans, or whether customers repay or default on their loans, affect a bank's profit or loss. The bank can minimize its Non-Performing Assets by forecasting loan defaulters. Because precise predictions are crucial for maximising earnings, it's essential to look at the different methodologies and compare them. A logistic regression model is a critical approach in predictive analytics for analysing the problem of predicting loan defaulters. Kaggle data is taken in order to investigate and predict. Logistic Regression models were used to calculate the various performance measures. Model is significantly better because it includes variables (personal attributes of customers include graduation, dependents, credit score, credit amount, credit period, and so on.) other than checking account information (which indicates a customer's wealth) that should be considered when correctly calculating the probability of loan default. As a result, by evaluating the likelihood of default on a loan, the right customers to target for loan granting can be easily identified using a logistic regression approach. The model predicts that a bank should not solely approve loans to wealthy consumers rather should also consider a customer's other characteristics, which play an important role in credit decisions and predicting loan defaulters. As the demand for products and services rises, so does the amount of capital credit given, and people are more eager to take credit than ever before. As a result, computer software has replaced the human interface as more people from all over the world (Urban, Rural, and semi-urban) push for a high demand for credit.

A Machine Learning software algorithm has been developed in order to construct a robust and efficient software algorithm that classifies individuals based on 13 characteristics (Gender, Education, Number of Dependents, Marital Status, Employment, Credit Score, Loan Amount, and others) whether they would be eligible for a loan or not. Although this is the first line of command, it will undoubtedly lower the workload of all other bank employees because the process will be automated to identify client segments and those who are qualified for a loan amount, allowing them to target those clients individually. And this will indicate whether or not the loan applicant meets the eligibility criteria for loan approval based on those 13 elements. To provide a convenient, prompt, and accurate method

of selecting deserving applicants for loan eligibility. To determine the model's accuracy, calculate the accuracy score, which is the level of precision displayed by the model when forecasting the applicant's loan eligibility.

The gradient descent method for calculating the regression coefficient, and the Sigmoid function improvement. Therefore, the number of repetitions has been reduced, the classification impact has been improved, and the accuracy has remained nearly unchanged. Kumar Arun, Garg Ishan, and Kaur Sanmeet have demonstrated how to reduce the risk factor when picking a safe individual in order to save time and money for the bank. This is performed by mining Big Data of previous records of persons to whom the loan was previously provided, and the machine was taught to get the best accurate result using a machine learning model based on these records/ experiences.

Model used	Accuracy
Logistic Regression	0.785
Decision Tree Classifier	0.662
K Neighbors Classifier	0.619
Naive Bayes	0.779
Random Forest Classifier	0.773
Support Vector Machine	0.650
XGBoost Classifier	0.773

In our model prediction of whether the loan would be accepted or not, we achieved the highest accuracy from the logistic regression model. On the dataset, the best case accuracy attained is 0.785. Our model was able to forecast whether the applicants in the dataset would be eligible for the loan when the project was completed. It was also able to anticipate the loan eligibility of a specific applicant by pointing out his row number. Applicants with a high income and smaller loan requests are more likely to be approved, which makes sense because they are more likely to payback their debts. The Loan creditability prediction system can assist companies in making the best judgement on whether to approve or deny a customer's loan request.