

SMART LENDER - APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL

MRIDUL MISHRA 1904093

MUTHU SHANMUGAM 1904094

PUSHKAR PANDEY 1904105

SUJITHA.R. 1904115



PROBLEM STATEMENT

- One of the most important factors which affect our country's economy and financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognized at banks across the globe. As we know credit risk evaluation is very crucial, we need to make a techniques which can be effectively used for risk level calculation.
- Credit risk is one of the main functions of the banking community. The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement.



EXISTING SOLUTIONS

- Machine learning algorithms such as Random forest, Logistic regression, Decision tree and Naive bayes algorithm are used on the same dataset. 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.
- Using data pre processing data mining and data filtering efficient Decision Tree is formulated with Decision Tree Induction Algorithm. It produces a model with the most relevant 6 attributes. A decision is made at each node and the leaf node gives us the final result. That is, if the customer possess the credibility for loan repayment on due time. The main benefit of applying Data Mining is that we can always rely on the result of the algorithm to know the credibility of loan applicant.
- Algorithms such as random forest, naïve bayes, k-nearest neighbour are used. random forest classifier is selected to build a model for the loan repayment credibility of a customer. Random forests form a family of methods that consist in building an ensemble (or forest) of decision trees grown from a randomized variant of the tree induction algorithm. In this algorithm actual learning is performed by the fit method.



INFERENCE FROM EXISTING SOLUTIONS

- Logistic regression and Decision tree shows their best accuracy output when compared to Naive bayes algorithm as it has performed poorly and is least recommended method among all the methods.
- For an individual, the historical behavior data are often too limited for profiling his/her behavior pattern and because of the heterogeneous nature of transaction data, there is no uniform treatment to various attribute values, which is a potential barrier for the model
- Anomaly detection often suffers from poor generalization ability and a very high false alarm rate. The transaction data are highly skewed, and it becomes a challenge for utilizing the label information effectively.



IDEATION

- There is an increasing rate of loan defaulters and banks are not able to correctly handle the loan request. To avoid this problem a machine learning algorithm is developed
- The system automatically selects the credible candidates to approve the loan and it will improve the speed, efficacy, and accuracy of loan approval processes.
- This help the user(Lender) to accurately identify whom to lend the loan and also help the banks to identify the loan defaulter for much-reduced credit risk.

