

# **SmartLender – Applicant Credibility Prediction for Loan Approval**

## **Introduction**

A loan is the major source of income for the banking sector of financial risk for banks. Large portions of a bank's assets directly come from the interest earned on loans given. The activity of lending loans carry great risks including the inability of borrower to pay back the loan by the stipulated time. It is referred as "credit risk". A candidate's worthiness for loan approval or rejection was based on a numerical score called "credit score". Therefore, the goal of this paper is to discuss the application of different Machine Learning approach which accurately identifies whom to lend loan to and help banks identify the loan defaulters for much-reduced credit risk.

## **Literature Survey**

In [1] they have used only one algorithm; there is no comparison of different algorithms. The algorithm used was Logistic Regression and the best accuracy they got was 81.11%. The final conclusion reached was only those who have a good credit score, high income and low loan amount requirement will get their loan approved. Comparison of two machine learning algorithms was made in [2]. The two algorithms used were two class decision jungle and two class decision and their accuracy were 77.00% and 81.00% respectively. Along with these they also calculated parameters such as Precision, recall, F1 score and AUC. The [3] shows a comparison of four algorithms. The algorithms used were Gradient Boosting, Logistic Regression, Random Forest and CatBoost Classifier. Logistic Regression gave a very low accuracy of 14.96%. Random forest gave a good accuracy of 83.51%. The best accuracy we got was from CatBoost Classifier of 84.04%. There was not much difference between Gradient Boosting and CatBoost Classifier in terms of accuracy. Accuracy of Gradient Boosting was 84.03%. Logistic Regression, Support Vector Machine, Random Forest and Extreme Gradient Boosting algorithms are used in [4]. The accuracy percentage didn't vary a lot between all the algorithms. But the support vector Machine gave the lowest variance. 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. Logistic Regression is the only algorithm used.

**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.

### *Advantages*

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

### *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.

### **Conclusion**

The analysis starts from data cleaning and processing missing value, exploratory analysis and finally model building and evaluation of the model. The best accuracy on public test set is when we get higher accuracy score and other performance metrics which will be found out. This paper can help to predict the approval of bank loan or not for a candidate.