

**KCG COLLEGE OF TECHNOLOGY**

Karapakkam, Rajiv Gandhi Salai, Chennai- 600097

**Smart Lender - Applicant Credibility Prediction for Loan Approval – Literature Survey**

IBM Team ID - PNT2022TMID27218

Team Members :

* Aakash. B (311019104001)
* Aakash. S (311019104002)
* Ajith. G (311019104007)
* Dharshan. S (311019104019)

**1.Loan Credibility Prediction System using Data Mining Techniques**

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

Published Month & Year : May 2021

Project Description : This model is implemented using the Logistic

Regression algorithm. Whenever program takes the input data it

gives the output in the form of binary i.e., either 0 or 1. If the output

is 1, it indicates that loan is approved. If the output is 0, then it

indicates that loan is not approved. Logistic Regression was the best fit

with highest accuracy score 81.12%.

Constraints :

* The major limitation of Logistic Regression is the assumption of linearity between the dependent variable and the independent variables.
* It is tough to obtain complex relationships using logistic regression. More powerful and compact algorithms such as Neural Networks can easily outperform this algorithm.

Possible solution :

Decision tree algorithm could be a possible solution to overcome the limitations for the following reasons -

* While utilizing a decision tree algorithm, it is not essential to standardize or normalize the data that has been collected. It can handle both continuous and categorical variables.
* The execution of a Decision tree algorithm must be possible without having to scale the data as well.
* The idea/ concept that drives the decision tree making model is more familiar and easier for developers/ programmers in comparison to other algorithms.

**2.An Approach For Prediction Of Loan Approval Using Machine Learning Algorithm**

Authors : Ms. Kathe Rutika Pramod, Ms. Panhale Sakshi Dattatray, Ms. Avhad Pooja Prakash, Ms. Dapse Punam Laxman, Mr. Ghorpade Dinesh B.

Published Month & Year : June, 2021.

Project Description : This model is implemented using Decision Tree algorithm. Decision trees are widely used in the banking industry due to their high accuracy and ability to formulate a statistical model in plain language. In Decision tree each node represents a feature (attribute), each link (branch) represents a decision (rule) and each leaf represents an outcome (categorical or continues value). Using different data analytics tools loan prediction and there severity can be forecasted.

Constraints :

* A small change in the data can cause a large change in the structure of the decision tree causing instability.
* For a Decision tree sometimes calculation can go far more complex compared to other algorithms.
* Decision tree often involves higher time to train the model.

Possible Solution :

Random forests are a strong modeling technique and much more

robust than a single decision tree. They aggregate many decision

trees to limit overfitting as well as error due to bias and therefore

yield useful results.

**3.Credit Risk Model Based on Central Bank Credit Registry Data.**

Authors : Fisnik Doko, Slobodan Kalajdziski, Igor Mishkovski.

Published Month & Year : March, 2021.

Project Description : In this Project, different algorithms and models like

Logistic Regression, Decision Tree, Random Forests, Support Vector

Machines (SVM), and Neural Networks, are evaluated and compared under

different cases of Datasets (Imbalanced Data without scaling, Imbalanced

Data with scaling, Balanced Data set without scaling).

Constraints :

* This paper uses only one dataset, and all countries have a similar dataset, which can vary by its requirements, laws and roles.
* There is not any research that uses data from credit risk, and we were unable to carry out such a comparison.
* It comparatively takes a lower execution time. But still it fails to provide results with better accuracy.

Possible Solution :

* Multiple classes of Datasets can be explored to extract more potential

Variables in order to lessen the impact of Dataset Bias problem.

* To gain Business insights, various analytical tools can be incorporated

and integrated into the model.

**4.An Approach for Prediction of Loan Approval using Machine Learning Algorithm**

Authors : Mohammad Ahmad Sheikh, Amit Kumar Goel

Published Year & Month : May, 2020.

Project Description : This prediction model takes variables like age,

Purpose, Credit history, Credit amount, Credit Purpose etc., apart from

traditionally considering only account information. The evaluation is

finally done by implementing Logistic Regression algorithm. The

output of the predicted model will be either 1 or 0. Predicted value 1

shows that the model is classified the application as accepted and

predicted value 0 implies that model classified the application has

not been accepted.

Constraints :

This model fails to take certain other potential variables into

consideration like -

* Gender
* Marriage History etc.

Possible Solution :

Exhaustive search for all possible variables can be done in order to

bring a more accurate and efficient solution.

**5.Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval**

Authors : Amruta S. Aphale, Dr. Sandeep R. Shinde.

Published Month & Year : August, 2020.

Project Description : In this project, different classification algorithms

are applied over testing datasets such as Neural Networks, Discriminant

Analysis, Naïve Bayes, K-Nearest Neighbor, Linear Regression, Ensemble

Learning/Method, Decision Trees. The experiment revealed that, apart from

the Nearest Centroid and Gaussian Naive Bayes, the rest of the algorithms

perform credibly well in term of their accuracy and other performance

evaluation metrics. Each of these algorithms achieved an accuracy rate

between 76% to over 80%.

Constraints :

* Even if using Neural Networks is deemed to be more accurate through

this experiment, they are more computationally expensive than traditional

Algorithms.

* Ensemble learning is less interpretable, the output of the ensembled model is hard to predict and explain. Hence the idea with ensemble is hard to sell and get useful business insights.

Possible solution :

Various Datasets can be used to train the model in the used algorithms to gain

much more refined and definite output. This can be further used for gathering

Accurate insights on the pre-established factors.

1. **C**redit Risk Model Based on Central Bank Credit Registry Dat