**Loan Approval Prediction Using Machine Learning**

**K. SRINIVASAN1, B. MANOJKUMAR2**

1Assistant Professor, Department of CSA, SCSVMV [Deemed to be University], Enathur, Kanchipuram, Tamilnadu, India.

2PG Student, Department of CSA, SCSVMV [Deemed to be University], Enathur, Kanchipuram, Tamilnadu,

India.

**ABSTRACT**

As we all know that now-a-days there's a rising in banking sector, ensuing several folks applying for bank loans. Looking for the mortal to whom the loan should be approved could be a troublesome. Thus, in this paper, we've proposed a model that predicts the authorization or rejection of associate degree mortal. This will be done by taking into account some machine learning techniques by predicting the model with the information of the previous records of the folks applied for loan.

 **Key Words:** Banking Sector, loan, predict, machine learning, logistic Regression

1. **INTRODUCTION**

Loans area unit the core business of banks. The most profit comes directly from the loan’s interest. The loan corporations grant a loan with an associate degree of intensive method of verification and validation. However, they still don’t have assurance if the mortal is in a position to repay the loan with no difficulties. Loan Prediction is extremely useful for worker of banks similarly as for the mortal additionally. The aim of this Paper is to supply fast, immediate method for the meriting candidates. It will offer special advantage to the bank.The Loan Prediction System will calculate the load of every option participating in loan process and on new check information same options area unit processed with relevance of their associated weight .A threshold time will be set for the mortal to see whether or not his/her loan is sanctioned or not. Loan Prediction System permits to jump to specific application so it is checked on priority basis.

1. **METHODOLOGY**

 **1.Collection of Data**

The input dataset is the whole bank dataset of customers who applied for the loan approval. The dataset is a CSV file. The dataset can be read into the python environment by using the read\_csv() method in pandas. So, we should import pandas into the present python environment. Some features of the Customers dataset are Loan ID, Married, Gender, Dependents, Education, Self-employed, Applicant Income, Applicant Income, Loan\_Amount\_Term, Loan Amount, Credit History, Loan Status and Property Area.

**Fig -2:** Attributes Of Dataset

,

* 1. **Data Visualizations**

The value count was done and so information was envisioned. Even when the data analysis, there's still no distinctive issue to work out loan standing. Categorical information was regenerated into numerical information.



* 1. **Data Modelling**

After the data is visualized, the data is modeled/trained. For this, the packages of 3 algorithms (Logistic regression, Decision tree and Random forest) were then imported. The model was then outlined and also the accuracy score was evaluated. Logistic Regression was the simplest work with the very best accuracy score of eighty-three.

It is applicable for categorical dependent variables employing a given set of freelance variables. Thus, the end result should be a categorical or distinct price. The output is often either Y or N, 0 or 1, true or false, etc. however rather than giving the precise price as zero or one, it offers some probabilistic values that lie between zero and one. In logistic regression, instead of fitting a curve, we have a tendency to work an"S" formed logistic operator, that predicts 2 greatest values (0 or 1). The curve from the logistic operation demonstrates the likelihood of one thing, as an example, despite whether or not the cells square measure harmful or not, a mouse is rotund or not supported on its weight, and so on. It's a major rule as a result of it will offer possibilities and classify the employment of various kinds of information and simply determines the foremost effective variables that

square measure used for classification. The S-structure curve is additionally referred to as the sigmoid operation or the logistic operation.

log(1/1-y) = b0+b1x1+b2x2+b3x3+……+bnx

**1.3 EXPERIMENTAL ANALYSIS**

Based on the information given by the loan applicant, we will predict whether or not the loan of the applicant is approved or not. The applicant must provide these values, and supported by these, the model can predict whether or not the loan is going to be approved or not.

After we have a tendency to apply Label encryption on Dependents and Property Area and Dummy encryption on the remaining options

Train-Test split applies to any supervised learning rule. Here the total dataset is get divided into 2 datasets as Train and check to create a model and to ascertain the performance of a model and every dataset gets divided into 2 once more as independent options and dependent options (only in supervised).

* Train Dataset: to train the machine learning model for learning functions.
* Test Dataset: to understand the performance of a trained machine learning model. Logistic function: G(z)=1/(1+e-z )

Accuracy = (True Positives + True Negatives)/Total Sample) Our accuracy was 0.8317667058123509.

Precision: exactness = (Number of True Positive)/(True Positive + False Positive) Our exactness score was 0.8117647058823529.

Recall: Recall = (True Positives)/(True Positive + False Negative) Our Recall score was 0.9726573293456218.

F1 Score: F1 Score = 2/((1/Precision) + (1/Recall)) Our F1 Score was 0.8317654110854729

**2**.**CONCLUSION**

In our model by employing a logistic regression model we finally predicted whether or not the loan is approved or not. So, to implement this, numerous input variables were required to get the output.When a program takes the computer file input it offers the output within the type of binary i.e., either 0 or 1. If the output is one then '1' is going to be displayed and it indicates that the loan is approved. If the output is zero then '0' is going to be displayed and it indicates that the loan isn't approved. Here, we have a tendency to had enforced a loan credibility prediction system that helps the organizations in creating the proper call to approve or reject the loan request of the purchasers. In this model, a Logistic Regression rule or algorithm is employed for the prediction. Incorporation of alternative techniques that vanquish the performance of common data processing models needs to be enforced and tested for the domain.

1. **REFERENCES**
2. [1]. Dileep B. Desai, Dr. R.V.Kulkarni “A Review: Application of Data Mining Tools in CRM for Selected Banks”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (2), 2013, 199 –201.M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
3. [2]. J.H. Aboobyda, and M.A. Tarig, “Developing Prediction Model of Loan Risk in
4. BanksUsing Data Mining”, MachineLearning and Applications: An International Journal (MLAIJ), vol. 3, no.1, pp. 1–9, 2016. K. Elissa, "Title of paper if known," unpublished.
5. [3]. A.B. Hussain, and F.K.E. Shorouq, “Credit risk assessment model for Jordanian commercial banks: Neuralscoring approach”, Review of Development Finance, Elsevier, vol. 4, pp. 20–28, 2014. JAC: A JOURNAL OF COMPOSITION THEORYVolume XIII, Issue V, MAY 2020ISSN: 0731-6755Page No: 324
6. [4]. T. Harris, “Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions”, Expert Systems with Applications, vol. 40,
7. pp. 4404–4413, 2013.