

# **LOAN ELIGIBILITY CLASSIFICATION USING MACHINE LEARNING**

(Under the guidance of Mrs. Y. Bhagya Lakshmi, Assistant Professor, B.tech,M.tech., Department of Information Technology, Tirumala Engineering College)

**N.Bhavani, G.Srilekha, R.Harika , SK.Jasmine , V.Dinesh**

IT Department , Tirumala Engineering College

## **Abstract**

In banking sector lots of people are applying for bank loans but the bank has limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted and which will be safer option for the bank is a typical process. So, we try to reduce This risk factor by selecting the safe person to save bank efforts and assets. This is done by using the previous records of the people to whom the loan was granted before and based on these records the machine was trained using the machine learning model which give the most accurate result. The record contains customer details such as Gender, Marital Status, Education, Number of Dependents, Income, of Dependents, Income, Loan Amount, Credit History, Property Area and others.

## **I. INTRODUCTION**

Machine learning is one of the applications of artificial intelligence (AI) that provides computers, the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to learn from themselves. The main aim is to allow computers to learn automatically without human intervention and also adjust actions accordingly. Loan prediction is very helpful for employee of banks as well as for applicants also. It provides quick, immediate and easy way to choose the deserving applicants. Customer will apply for the loan first then the machine learning model automates the loan eligibility process based on the customer details provided while filling the application.

## **II LITERATURE SURVEY**

Machine learning is one of the applications of artificial intelligence (AI) that provides computers, the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to learn from themselves. The main aim is to allow computers to learn automatically without human intervention and also adjust actions accordingly.

**Loan Approval Prediction based on machine learning approach** the main objective is to [1] predict whether assigning the loan to particular person will be safe or not. This is divided into four sectors i.e., Data Collection, Comparison of machine learning models on collected data, Training of system on most promising model.

**Loan Prediction using machine learning model** [2] whether it will be safe to allocate the loan to a particular person or not. This is divided into three sections such as Collection of data, Data Cleaning and Performance evaluation. By using machine learning models we try to reduce the risk factor behind selecting the safer person so as to save lots of bank efforts and assets.

**Loan Prediction using machine learning algorithms** in India the number of people or organization [3] applying for a loan gets increased every year. The bank have to put lot of efforts to predict whether the customer can pay back the loan amount or not in the given time. The aim is to find the nature or background or credibility of customer applying for the loan.

#### **Importance of machine learning in banking sector:**

It improves decision making, as compared to traditional methods, machine learning helps banks to calculate credit score accurately. The main reason machine learning can do this is that it can provide an objective evaluation without any bias. The huge amount of data collected from the potential borrower assists banks in making better decisions.

Credit card fraud is a huge problem in the banking industry. Machine learning for banking can significantly lower the number of fraudulent activities. The majority of fraud occurs when customers pay for products, whether online or offline. Machine learning prevents this from happening in several ways.

It reduces risk for both customers and banks through accurate reporting. It can also make predictions based on transaction history after giving credit to customers. Employees have more insights into credit risk testing.

### **III EXISTING SYSTEM**

Banks maintain all the records of customers. Even though, those records are not used in an efficient manner for the loan eligibility process. To maintain the records in an efficient error free manner, the new proposed system is introduced.

#### **DISADVANTAGES**

- Computation time is very high
- Difficult to maintain customer details
- Risk factor is very high
- Finding out to whom the loan can be granted will be the safer option is a typical process

### **IV PROPOSED SYSTEM**

This system will predict whether the loan applied by the customer is approved or denied. So, this system provides the banking sector to predict the loans in an easier way with less risk. This proposed system not only predicts the loan accurately but also reduces time for prediction. We predict the loan data by using some machine learning algorithms like Logistic Regression, Decision Tree, Random Forest, Support Vector Machine to get accurate result.

#### Advantages

- Generates accurate and efficient results
- Computation time is greatly reduced
- Easy maintenance of customer records
- Reduces manual work

## V MODULES

### **Cleaning of the Dataset**

Data cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. But, as we mentioned above, it isn't as simple as organizing some rows or erasing information to make space for new data. Data cleaning is a lot of muscle work. There's a reason data cleaning is the most important step if you want to create a data culture, let alone make airtight predictions

### **Pre-processing of the Dataset**

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine-learning model. When creating a machine learning project, it is not always the case that we come across clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put it in a formatted way. So for this, we use data preprocessing task.

### **Machine Learning Classifiers**

Classification is the process of predicting the class of given data points. Classes are sometimes called targets/ labels or categories. Classification predictive modelling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).

### **Picking Best Classifiers**

The pickle module implements binary protocols for serializing and de-serializing a Python object structure. "Pickling" is the process whereby a Python object hierarchy is converted into a byte stream, and "unpickling" is the inverse operation, whereby a byte stream (from a binary file or bytes-like object) is converted back into an object hierarchy.

### **Building Application**

Streamlit is an open-source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as sci-kit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc.

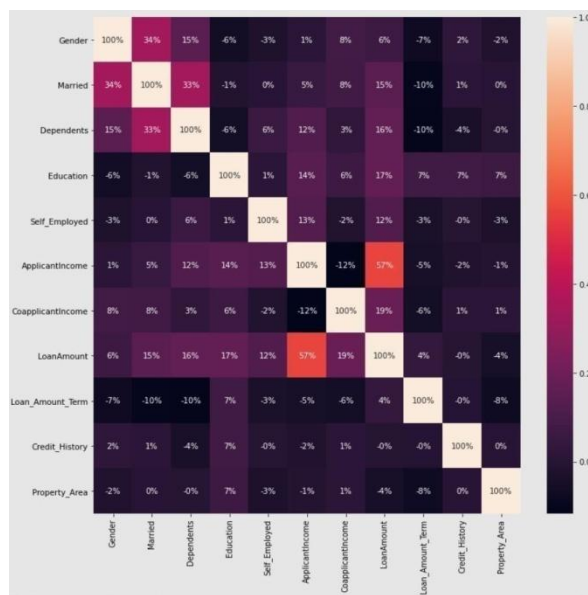
## VI TRAINING DATASET

Before feeding data to an algorithm we have to apply transformations to our data which is referred as pre-

processing. By performing pre-processing the raw data which is not feasible for analysis is converted into clean data. In-order to achieve better results using a model in Machine Learning, data format has to be in a proper manner. The data should be in a particular format for different algorithms. For example, if we consider Random Forest algorithm it does not support null values. So that those null values have to be managed using raw data.

**Missing values**

Filling missing values is one of the pre-processing techniques. The missing values in the dataset is represented as NaN. Pandas can detect the missing values. We have filled that missing values of Gender with Female, Married with Yes, Dependents with 0, Self\_Employed with No, Loan\_Amount\_Term with 360.0, Credit\_History with 1.0 .



**VII CONCLUSION**

We have used six algorithms like Logistic Regression, Decision Tree, Random Forest, Naive Bayes, KNN, SVM and in-order to predict approved or denied of loan. The accuracy varies for different algorithms. The accuracy for SVM algorithm is 82.70. The accuracy of Decision tree algorithm is 61.62. The accuracy for Random Forest algorithm is 81.62. The accuracy for Naive Bayes algorithm is 82.16. The accuracy for KNN algorithm is 79.45. The accuracy for Logistics Regression algorithm 82.24. The highest accuracy is given when we have used XG- boot algorithm with an accuracy of 83.24 percent.



### BIBLIOGRAPHY

1. Kumar Arun, Garg Ishan, Kaur Sanmeet, —Loan Approval Prediction based on Machine Learning Approachll , IOSR Journal of Computer Engineering(IOSR-JCE), Vol. 18, Issue 3, pp. 79-81, Ver. I (May-Jun. 2016).
2. S. Vimala, K.C. Sharmili, —Prediction of Loan Risk using NB and Support Vector Machinell , International Conference on Advancements in Computing Technologies(ICACT 2018), vol. 4, no. 2, pp. 110-113, 2018.
3. Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi VimalaKumari,kVikash,“LoanPredictionby using MachineLearning Models”, InternationalJournalofEngineering andTechniques.Volume 5 Issue 2, Mar-Apr 2019.
4. Aafer Y, Du W &Yin H 2013, DroidAPIMiner: ‘Mining API-Level Features forRobust Malware Detection in Android’, in: Security and privacy in CommunicationNetworks Springer, pp 86-103
5. J. R. Quinlan. Induction of Decision Tree. Machine Learning, Vol. 1, No. 1. pp.81-106., 1086.