***LOAN APPLICATION STATUS PREDICTION***

PROBLEM DEFINITION

Loan approval prediction using machine learning is a cutting edge technology in lending.It involves harnessing machine learning models to predict loan approvals by learning patterns from historical data. This approach can provide quicker and more data driven decisions by automating the loan approval process. In this data regarding gender having different relationship status having different qualification background having different loan amount credit history and loan status has been given. By analysing this data we have to make an informed decision regarding lending.

DATA ANALYSIS

In this phase of data analysis whole of the data has been analysed. Data cleaning has been done which all columns are required which columns can be dropped which are not helping us in our visualisation. Loan\_id has been dropped.

EXPLORATORY DATA ANALYSIS

This steps includes checking for data types looking for null values and finding the mean,variance,standard deviation for comparison of various attributes.In the given model there are certain null values present in 1) Gender, 2)Dependents, 3)Self employed, 4)Loan\_amount ,5)Loan\_amount\_term and 6)Credit\_history. Here much of the missong data belongs to categorical value so IDBI dropped all the missing values.

DATA VISUALISATION

Various graphs have been made to draw conclusions from the data. Various countplots drawing comparison of married,gender,occupation,education,self\_employed,credit\_history,property\_area have been done with loan status to fetch some useful information for our analysis.correlation is also calculated to make a comparison. Even outliers are present in the data we remove the outliers while using ZSCORE. Various graphs showing univariate and bivariate analysis has been drawn to check for outliers and skewness.

PRE PROCESSING PIPELINE

In this step first LABEL ENCODER has been used since categorical data is present in the data. 'Gender','Married','Dependents','Education','Self\_employed', 'Property\_area',’Loan\_status’ has been converted into integer data.STANDARD SCALING TECHNIQUE has been used to scale the feature so that mean and std dev can be zero.

BUILDING MACHINE LEARNING MODELS

Data is split into X and Y while taking loan status as a target variable. Various classification models have been used to predict the accuracy of the model.Data has been tested while using 1 LOGISTIC REGREESSION 2 DECISION TREE CLASSIFIER 3 KNEIGHBORS CLASSIFIER 4 RANDOM FOREST CLASSIFIER.Amongst all the models accuracy of LOGISTIC REGRESSION is the highest with the accuracy of .83. so we check the accuracy oby using using HYPERPARAMETER TUNING USING LOGISTIC REGRESSION. Then accuracy is checked by using auc roc curve and auc score where the score is.79 which depicts our modelis predicting well.

CONCLUDING REMARKS

* BANKS ARE GIVING MORE LOANS TO MALE. MARRIED ONES ARE GIVEN PREFERENCE IN APPROVING LOANS.
* GRADUATES ARE GIVEN MORE PREFERENCE FOR SANCTIONING LOANS AS COMPARED TO UNDERGRADUATES
* SALARIED PERSONS ARE GIVEN MORE PREFERENCE THAN SELF EMPLOYED DUE TO CERTAINITY OF RECOVERY IN CASE OF SALARY
* LOAN PREFERENCE IS OBVIOUSLY GIVEN TO THE PERSONS WHOSE CREDIT SCORE IS 1 WHERE THERE IS CERTAINITY FOR BANK TO RECOVER LOAN
* SEMI URBAN PROPERTY HOLDERS ARE GIVEN PREFERENCE