Credit Approval Prediction

An Application of Logistic Regression Analysis

ABSTRACT

In the backdrop of the fact that Loan disbursement is one of the fundamental aspects of all banks and financial companies as this profit is a major source of bank's assets, which calls for adopting a rigorous process for validation and corroboration of a costumer but still does not guarantee that the customer chosen will get the loan approved. To reduce the risk, there is a need to predict whether the customer is a loan defaulter to reduce banks non-profitable assets. Hence, this paper approaches identifying the right customers to be prioritized for granting loans that can be easily detected by evaluating their chances for loan eligibility.

Since choosing the appropriate customer out of many applicants that apply for a loan in any bank or financial company is a herculean task, The primary objective of this validation is to choose appropriate customers so that their assets can be in good hands and help people who are in need. In this work, we propose a loan recommendation system that will provide an immediate and simple way to choose the right applicant based on the validation of features.

This research aims to predict a model for loan disbursement by using a regression model. The data is collected from the Analytics Vidhya for studying and prediction. Logistic Regression models have been performed, and the different measures of performances are computed. The model is marginally better because it includes variables other than checking account information (which shows the wealth of a customer) that should be taken into account to calculate the probability of default on a loan.

This paper aims to predict credit risk by conducting a bivariate data analysis to observe the impact of each variable on the target variable using the stacked bar plots. Various models are constructed based on the significance of the variables on the target variable, and the applicants' credit risk is calculated.

INTRODUCTION

Distribution of the loans is a crucial business part of almost every bank. A significant portion of the bank's assets is directly from the profit earned from the loans distributed by the banks. The prime goal in the banking domain is to invest their assets in safe hands. Lending money to unsuitable loan applicants results in credit risk.

In the present scenario, the cost of assets is increasing day by day, and the capital required to purchase an entire asset is very high. So, purchasing it out of your savings is not possible. The easiest way to get the necessary funds is to apply for a loan. Also, the loan needs to be approved manually by a bank representative to check whether the person is eligible for the loan or not by calculating the risk associated with it. With a motive to decrease the approval time and the risk associated with the loan, this loan prediction model is introduced.

Many banks approve loans after a lengthy procedure of verification, yet there is no guarantee whether the picked candidate is the right one or not. As it is done manually, it is a time-consuming process and is susceptible to errors. Estimating the risk, which is involved in a loan application, is one of the most significant concerns of the banks in order to survive in the highly competitive market. If the loan is not repaid, then it accounts for a loss to the bank, as banks earn most of their profits by the interest paid to them. If the banks lose too much money, then it would result in a banking crisis, which consequently may affect the economy of the country. Therefore, it is very important that the loan should be approved with the least amount of error in risk calculation while taking up the least time possible. Moreover, a loan prediction model is required that can predict quickly whether the loan can be passed or not with the least amount of risk possible.

"The logistic regression algorithm is most accurate for predicting loan eligibility. It can be seen that the logistic regression algorithm gives high accuracy than the other algorithms in the prediction of loan default and has strong ability of generalization."

Logistic regression is a mathematical approach used to describe the relationship between an independent variable to a numerous dependent variable or a dichotomous dependent. The regression function is employed because the proposed covariates are a combination of continuous and categorical random variables, whereas the dependent variable is dichotomous by default.[1]

"With the wide availability of sophisticated statistical software for high-speed computers, the use of logistic regression is increasing. Generally, logistic regression is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables.

During the last decade, logistic regression has been gaining popularity. Such popularity can be attributed to researchers' easy access to sophisticated statistical software that performs comprehensive analyses of this technique. It is anticipated that the application of the logistic regression technique is likely to increase".[2] In this paper, we demonstrate that logistic regression can be a powerful analytical technique when the outcome variable is dichotomous.

A prediction model uses data mining, statistics, and probability to forecast an outcome. Every model has some variables known as predictors that are likely to influence future results. The data is collected from various resources and then a statistical model is made. As more data becomes available, the model becomes more refined, and the error decreases that implies the data will be able to predict with the least risk and consume as little time as it can. The prediction model helps the banks by minimizing the risk associated with the loan approval system and helps the applicant by decreasing the time taken in the process.

Through our proposed model, we can predict whether that specific customer is safe or not. Data mining algorithms are used to study the loan-approved data and exact patterns, which would help in predicting the reasonable defaulters, thereby helping the banks for making better choices in the future.[3]

The objective of this paper is to predict the loan approval using logistic regression that can reduce the loan approval time and decrease the risk associated with it. It is done by predicting if the loan can be given to that person based on various parameters like credit history, income, age, marital status, gender, etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.

REVIEW OF LITERATURE

The author, Vaidya, Ashlesha[4] uses logistic regression as a machine learning tool in paper and shows how predictive approaches can be used in real-world loan approval problems. His paper uses a statistical model (Logistic Regression) to predict whether the loan should be approved or not for a set of records of an applicant. Logistic regression can even work with power terms and nonlinear effects. Some limitations of this model are that it requires independent variables for estimation, and a large sample is required for parameter estimation.

It is hard for bank employees to check and predict whether the customer is trustworthy to approve the loan with interest rate. The Data analysis helps in reducing the complexity of data and gives appropriate results helpful for the user. The data analysis technique used to analyze the behaviour of data is quite exploratory.[5], [6]

The main motive behind the paper is to identify the nature of loan applicants; The Loan prediction system is a very useful tool for the bank employee and also for the customer. It reduces the risk factor. The loan prediction system provides results with the help of trained models for loan approval. The datasets used to train the models are the past records collected.[4],[6] As the banking sector improves, the count of loan applicants is increasing. Still, due to limited funds available in banks, it is necessary to identify who is safer to approve the loan.

“Vaidya had suggested a method for approving loan forecasts using logistic regression”. “Logistic Regression is one of the most popular and very useful classification-based algorithms”. “The purpose or the importance of using Logistic Regression was that it uses the concept of predictive analysis which was suitable enough for describing the data”.[7]

“As we know, credit risk evaluation is crucial, and there are a variety of techniques used for risk level calculation. In addition, credit risk is one of the main functions of the banking community”.[8]

“In this paper, we have taken the information of past clients of different banks to whom on a bunch of boundaries advance were endorsed. So, the LR model is prepared on that record to get precise outcomes as it is well-known that the circulation of credits is the fundamental business of each bank. Our fundamental goal of this examination is to anticipate the wellbeing of credit”.[8]
The primary segment of a bank's beneficial resource straightforwardly comes from the benefit acquired from the advances being circulated by the bank.[8]

“One of the disadvantages of this model is that it emphasizes different weights to each factor, but in real life sometimes loan can be approved based on single strong factor only, which is not possible through this system”.[8]

There may be various benefits that the bank can obtain, such as setting a time limit for applicants to check and ensure whether their loan will be sanctioned or not. This prediction system may be helpful in the sense that it gives the right to bankers to focus more on valuable assets for the bank, not focus on the poor applicants. It will reduce the time for the loan application process of the applicant.[8]

METHODOLOGY

Objective:

Predictive Analysis: The paper’s main objective is to create a prediction model for loan approval and classify whether an applicant applying for a loan is eligible or not or a person is a loan defaulter using data collected by a bank employee.

Risk Minimization: In the present, the counts of people applying for loans are incrementing due to financial conditions. It is hard for employees to select the rightful applicant for a loan. Hence risk behind the loan approval can be minimized with the usage of these prediction models.

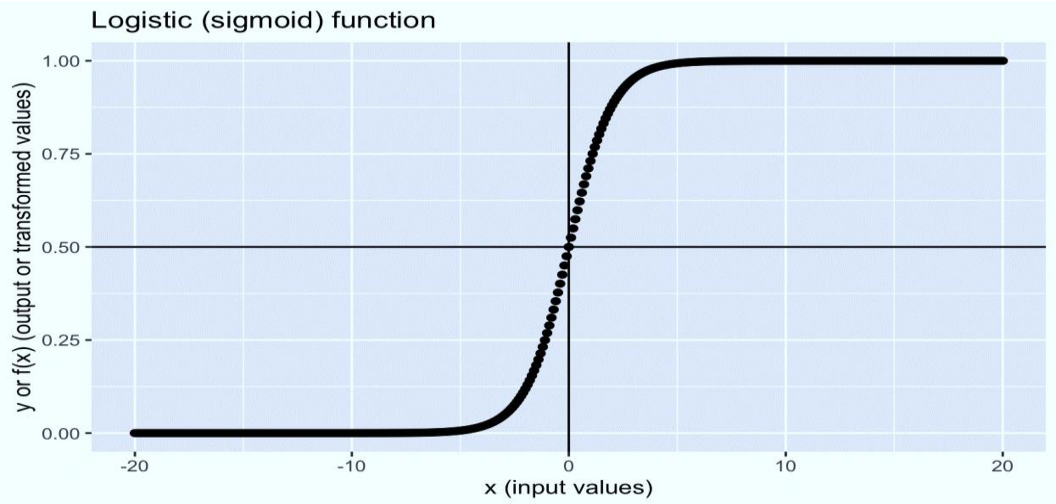
Logistic Regression

Logistic regression is one of the popular Machine Learning algorithms based on the Supervised Learning technique. This technique is applied to a categorical dependent variable using a given set of independent variables. Thus, the outcome must be a categorical or discrete value. The output can be either Yes or No, 0 or 1, true or false, and the like, but instead of giving the exact value as 0 or 1, it showcases probabilistic values which lie between 0 and 1. Logistic regression is much similar to linear regression except for how it is used. Whereas Linear Regression is used to solve regression problems, Logistic regression is used to solve classification problems.
However, in Logistic regression, rather than fitting a regression line, we fit an "S" shaped logistic function, which predicts the two most significant values (0 or 1). The curve from the logistic function demonstrates the probability. It is a significant algorithm since it

not only can provide probabilities and classify the use of different types of data but also easily determines the most influential variables that are used for classification.

Logistic Function (Sigmoid Function)

The Sigmoid Function is a numerical function that outlines the predicted values to probabilities. It maps any real value to another value between 0 and 1. The value must be between 0 and 1; that is, it cannot exceed the limit, then it forms a curve that takes the "S" form. This S-structure curve is known as the Sigmoid function or the Logistic function. In Logistic regression, we employ the concept of the threshold value, which characterizes probability. The values below the threshold value tend to be 0.



A logistic function or logistic curve is a typical S-shaped curve with the following equation,

$f(x) = \dfrac{L}{1+e^{-k(x-x_0)}}$, where,

x_0, the x value of the Sigmoid's midpoint;

L, the curve's maximum value;

k The logistic growth rate or steepness of the curve.

The standard logistic function is the logistic function with parameters, $k=1$, $x_0=0$ and $L=1$, which yields,

$$f(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} = \frac{1}{2}+\frac{1}{2}\tanh\left(\frac{x}{2}\right)$$

Logistic Regression Equation

The equation can be obtained from the Linear Regression equation. The mathematical steps to obtain the equations are given below:

The equation of the straight line can be written as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots\ldots + b_n x_n$$

In Logistic Regression, y can be between 0 and 1 only, so let's divide the equation by (1-y).

$$\frac{y}{1-y} = \begin{cases} 0, & for\ y = 0 \\ \infty, & for\ y = 1 \end{cases}$$

But we need a range between $-\infty$ to ∞, then take the logarithm of the equation it will become:

$$\log\left(\frac{y}{1-y}\right) = b_0 + b_1 x_1 + b_2 x_2 + \ldots\ldots + b_n x_n$$

Above one is the final equation of logistic regression.

The Odds Ratio

For a continuous independent variable, the odds ratio can be defined as:

$$OR = \frac{odds(x+1)}{odds(x)} = \frac{\dfrac{P(x+1)}{1-P(x+1)}}{\dfrac{P(x)}{1-P(x)}} = \frac{e^{\beta_0+\beta_1(x+1)}}{e^{\beta_0+\beta_1 x}} = e^{\beta_1}$$

This exponential relationship provides an interpretation for. The odds multiply by e^{β_1} for every 1-unit increase in x. The odds ratio expresses how many times the probability that an event will occur increases or decreases if a unit change of the independent variable occurs (with determined values of other independent variables). This ratio is calculated as follows:

$$S(A) = p(A)\ /1 - p(A)$$

The odds ratio refers to the situation when the occurrence of a given phenomenon is studied in two independent groups. It is expressed by the ratio of the chance of this phenomenon occurring in group A or $O(A)$, to the chance of this phenomenon occurring in group B or $O(B)$. The formula for the odds ratio is:

$$OR = \frac{O(A)}{O(B)} = \frac{\frac{p(A)}{1-p(A)}}{\frac{p(B)}{1-p(B)}}$$

where:

$p(A)$ - the probability of occurrence of the event (value of 1 dependent variable) in category A (value of 1 independent variable);

$p(B)$ - the probability of occurrence of the event (value of 1 dependent variable) in category B of observation (value of 0 of the independent variable).

It is assumed that if:

- $OR > 1$, then X stimulates the event to occur. OR shows how much the probability of 1 in the dependent variable increases when the predictor value increases by one unit;
- $OR < 1$, then X, de-stimulates the occurrence of the event. OR shows how much the probability of 1 in a dependent variable decreases when the predictor value increases by one unit;
- $OR=1$, then X does not affect the occurrence of the event.

To assess the degree of adjustment of the logistic regression model to empirical data, the R^2 counting measure is used, which takes values from the range (0,1) defined as follows.

$$R^2_{count} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

The closer the value of this measure to 1, the better the logistic model will fit the empirical data of the studied phenomenon. In such a method indicates the percentage of rightly classified cases. The model works well in forecasting the studied phenomenon, which implies that the classification based on the model is better than random. The quality of the built logistic regression model can also be assessed by various other measures.

PROBLEM STATEMENT:

Account firms and banks need to automatize the credit qualification activity that is essentially dependent on data given by customers when rounding out an online structure. Sex, Marital Status, Education, Number of Dependents, Salary, Loan Amount, Credit History, and other such different subtleties are incorporated. “Approval of Loan is a very common real-life problem that every company faces in their lending operations. If the loan approval process is automated, it can save a lot of man-hours and improve the speed of service to the customers. The increase in customer satisfaction and savings in operational costs are significant” [9]. “However, the rewards can only be realized if the bank has a sturdy model in place to accurately forecast which client's loans it should accept and which it should reject, in order to reduce potential risk. Hence the primary aim of our paper is to minimize the risk present in approving the loan to the applicant. Therefore, by using the logistic regression analysis approach, the right customers to be targeted for granting loans can be easily detected by evaluating their likelihood of default on loan.

EXPLORATORY DATA ANALYSIS

METHODOLOGY

For implementing the Logistic Regression using R, we had done the following steps:

1. Data Collection
2. Data Pre-processing
3. Fitting Logistic Regression to the Training set
4. Predicting the test result
5. Test accuracy of the result
6. Visualizing the test set result

1. Data Collection

The accuracy of our model is dependent on data collection, generally referred to as the representation of data. The data is collected from the website Kaggle, which will be used as a training and testing dataset. The data set used in this study is obtained from the data set available for the loan prediction problem in *Analytics Vidhya*. The data set was imported in ‘.csv’ format. It consists of various variables taken into consideration before approving the loan to the applicant. The attributes affecting the loan approval are mentioned in the table below.

Attributes	Description
Gender	Male or Female
Married	Yes or No
Dependents	Number of dependents
Education	Graduate or Not
Self_Employed	Yes or No
Credit_history	Meets guidelines or not
Property_Area	Urban or rural or semiurban

Table1: Categorical Variables

Attributes	Description
ApplicantIncome	Applicant's Income
Co-applicantIncome	Co-applicant's Income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months

Table 2: Numerical Variables

2. Data Pre-processing

Pre-processing includes removing null values, empty cells, missing values, conversion of data type, and randomization of data to remove, performed in order to stop misleading. The data collected in the previous step has been analyzed for a better comprehension of the structure of the data. Next, the character variables are converted into factors according to the suitable levels. The result is as follows.

```
> summary(data)
   Loan_ID              Gender      Married    Dependents        Education      Self_Employed ApplicantIncome CoapplicantIncome
 Length:614             : 13         :  3       : 15      Graduate     :480       : 32       Min.   :  150   Min.   :    0
 Class :character   Female:112   No :213    0 :345      Not Graduate:134    No :500      1st Qu.: 2878   1st Qu.:    0
 Mode  :character   Male  :489   Yes:398    1 :102                           Yes: 82      Median : 3812   Median : 1188
                                            2 :101                                        Mean   : 5403   Mean   : 1621
                                            3+: 51                                        3rd Qu.: 5795   3rd Qu.: 2297
                                                                                          Max.   :81000   Max.   :41667

   LoanAmount     Loan_Amount_Term Credit_History    Property_Area Loan_Status
 Min.   :  9.0   Min.   : 12      Min.   :0.0000   Rural    :179   N:192
 1st Qu.:100.0   1st Qu.:360      1st Qu.:1.0000   Semiurban:233   Y:422
 Median :128.0   Median :360      Median :1.0000   Urban    :202
 Mean   :146.4   Mean   :342      Mean   :0.8422
 3rd Qu.:168.0   3rd Qu.:360      3rd Qu.:1.0000
 Max.   :700.0   Max.   :480      Max.   :1.0000
 NA's   :22      NA's   :14       NA's   :50
```

The above summary of data shows that there are missing values (NA's) in the variables of the given data. These missing values (NA's) of respective variables are replaced by their median in the case of numerical variables and by mode in the case of categorical variables.

```
> summary(data)
   Loan_ID              Gender      Married    Dependents        Education      Self_Employed ApplicantIncome CoapplicantIncome
 Length:614         Female:112   No :213    0:360      Graduate     :480    No :532      Min.   :  150   Min.   :    0
 Class :character   Male  :502   Yes:401    1:102      Not Graduate:134    Yes: 82      1st Qu.: 2878   1st Qu.:    0
 Mode  :character                           2:101                                       Median : 3812   Median : 1188
                                            3: 51                                       Mean   : 5403   Mean   : 1621
                                                                                        3rd Qu.: 5795   3rd Qu.: 2297
                                                                                        Max.   :81000   Max.   :41667
   LoanAmount     Loan_Amount_Term Credit_History   Property_Area Loan_Status
 Min.   :  9.0   Min.   : 12.0    Min.   :0.000   Rural    :179   N:192
 1st Qu.:100.2   1st Qu.:360.0    1st Qu.:1.000   Semiurban:233   Y:422
 Median :128.0   Median :360.0    Median :1.000   Urban    :202
 Mean   :145.8   Mean   :342.4    Mean   :0.855
 3rd Qu.:164.8   3rd Qu.:360.0    3rd Qu.:1.000
 Max.   :700.0   Max.   :480.0    Max.   :1.000
```

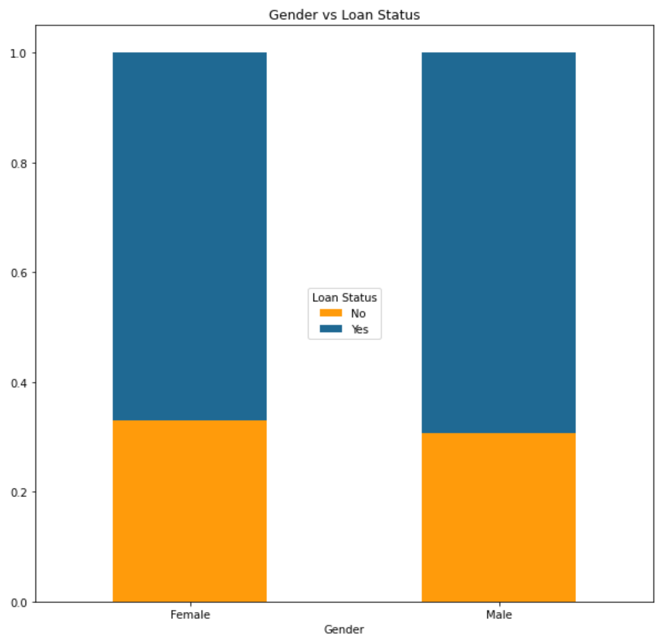
The next step is to divide the data set into two groups, viz.,

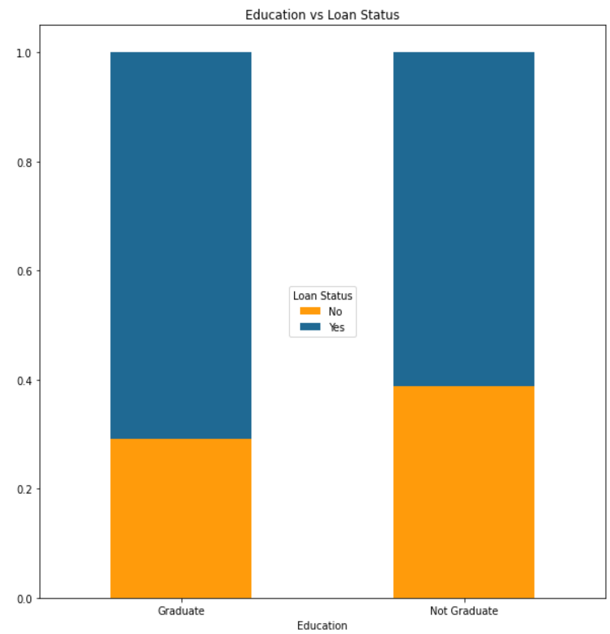
1. Training dataset for determining a logistic model and
2. Testing dataset for checking the accuracy of the predicted outcome of the model determined.

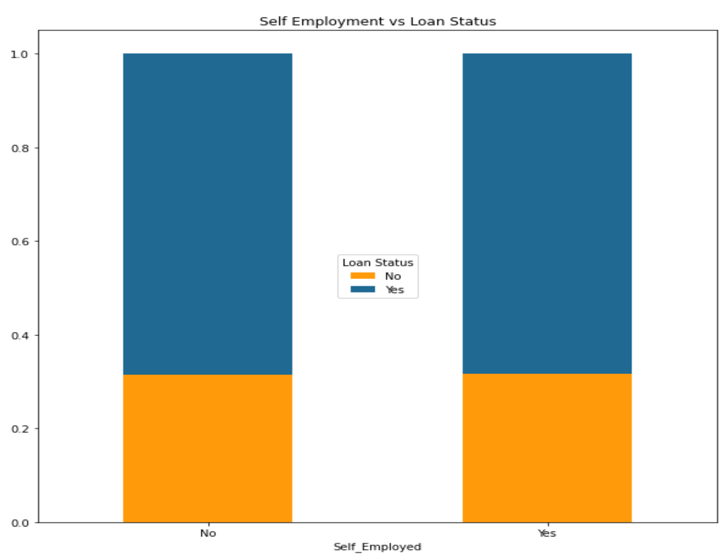
Explanatory Analysis

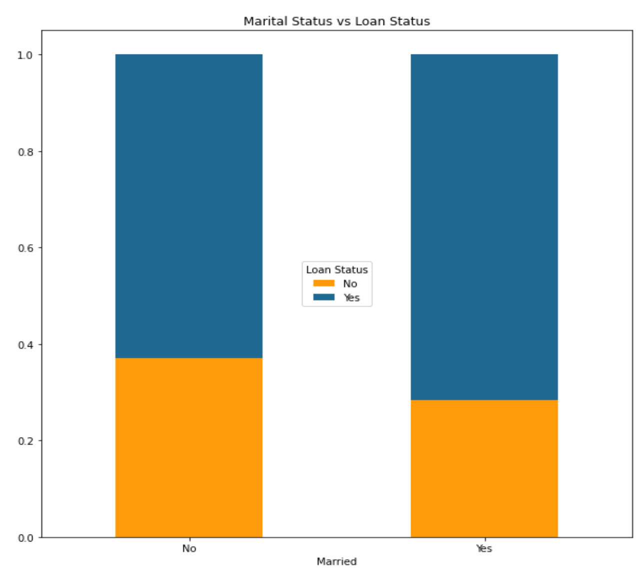
Bivariate Analysis

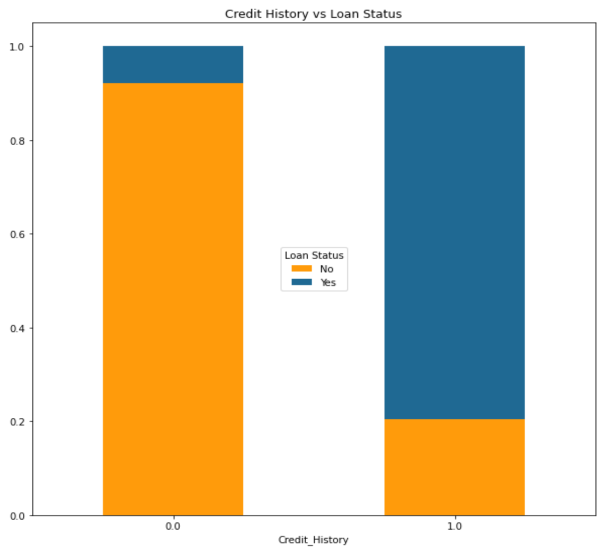
From the bivariate analysis of each variable against the target variable Loan_Status, it is evident that the variables Gender, Self_Employed are less significant and have less effect on the approval of the loan. The stacked bar charts for the significance of each attribute are portrayed below.

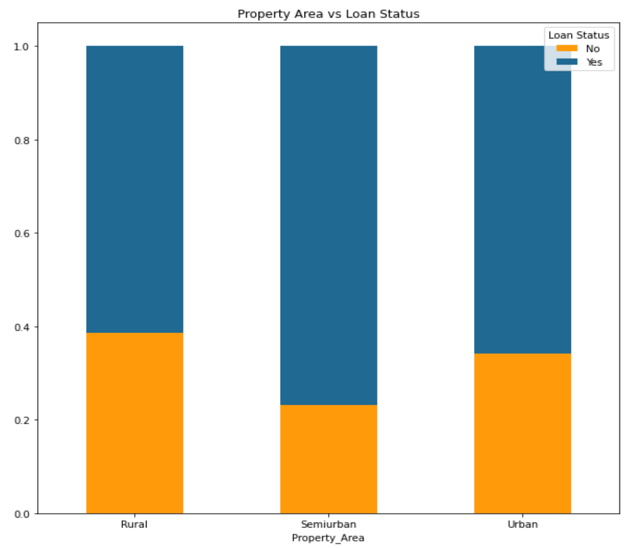




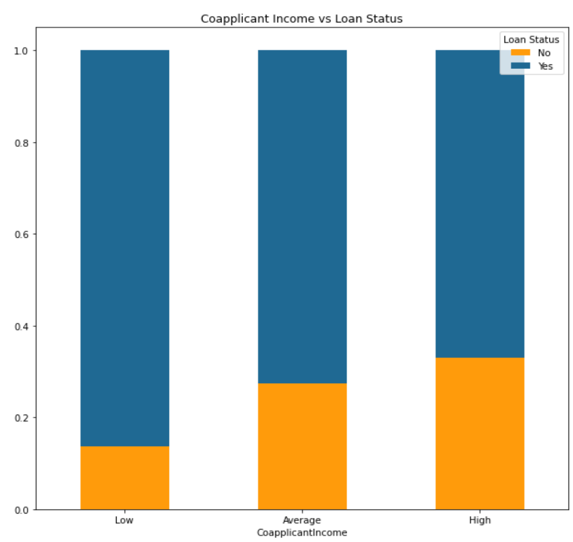


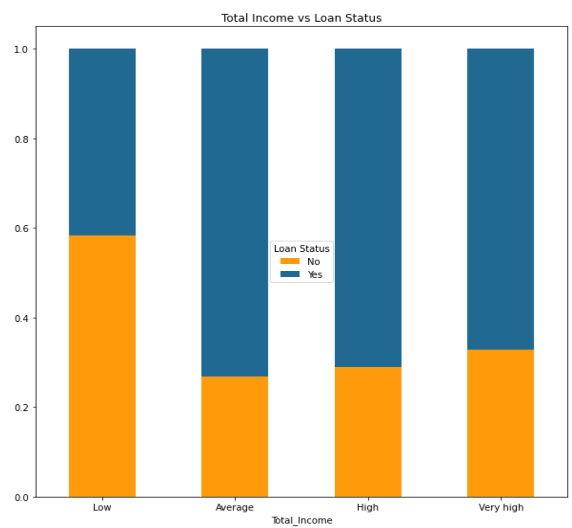






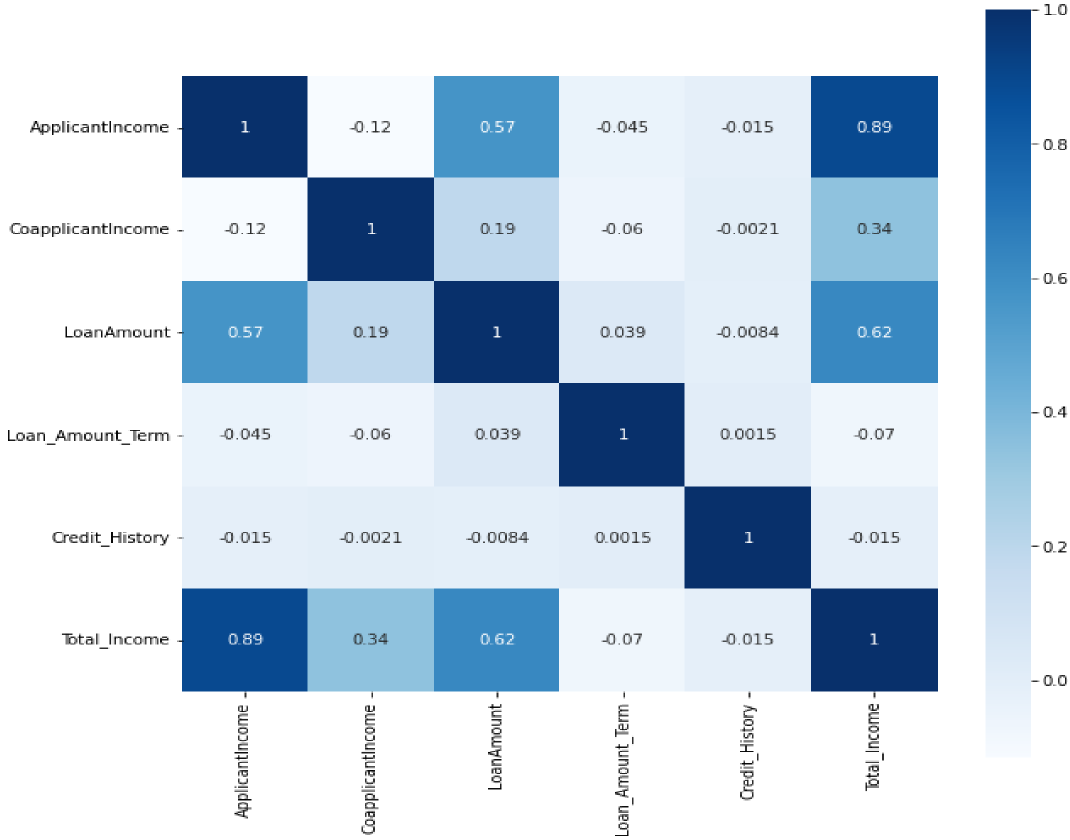
The Bivariate analysis of Co-Applicant Income against Loan_Status shows that the applicants with less co-applicant income have more chance for loan approval compared to the applicant with more co-applicant income. This might be because most people don't have a co-applicant. So, we combine the Applicant Income and Co-Applicant Income into a new variable, Total_Income.





The analysis of Total Income and Loan Status proves that the chance of approval of the loan for the applicant with less total income is less than that of the applicant with more total income. However, the chance of loan approval for the applicant with an average income is the highest. It must be emphasized that the exact significance of the attributes cannot be derived from the bar plots based on bivariate analysis, as the significance levels tend to fluctuate.

As a successive step to analyze numerical values, A Heat Map is made to calculate the correlation between the variables and decide which variables are more correlated and those that should be neglected.



As the variable Total Income is constructed from Applicant and Co-applicant Income, they are neglected from the model.

3. Fitting the Model:

Multiple machine learning approaches are used to get a proper result. In this paper, we have used Logistic regression with the highest accuracy to obtain a prediction.

MODEL CONSTRUCTION

Training the Model: Training of the models is done to learn the data patterns and relationships between variables. Larger the dataset for training, more evaluation needed to be done, and more accurate results.
A Logistic model is constructed on the training data set, and the summary of the model obtained is shown below:

```
> model=glm(Loan_Status~.,data=training, family="binomial")
> summary(model)

Call:
glm(formula = Loan_Status ~ ., family = "binomial", data = training)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1289  -0.4111   0.5533   0.7017   2.2754

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)            -4.96275    2.49927  -1.986  0.04707 *
Gender                  0.10567    0.32411   0.326  0.74440
Married                 0.59254    0.27767   2.134  0.03285 *
Dependents              0.01993    0.13181   0.151  0.87982
Education               0.43226    0.28540   1.515  0.12988
Self_Employed          -0.14430    0.33847  -0.426  0.66988
Credit_History          3.74723    0.45526   8.231  < 2e-16 ***
Property_AreaSemiurban  0.76033    0.28823   2.638  0.00834 **
Property_AreaUrban      0.10730    0.28622   0.375  0.70774
EMI                    -0.30334    0.24891  -1.219  0.22297
TotalIncome             0.13065    0.27345   0.478  0.63281
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 611.54  on 491  degrees of freedom
Residual deviance: 461.95  on 481  degrees of freedom
AIC: 483.95

Number of Fisher Scoring iterations: 5
```

The above summary exhibits the significance of the variables. Based on the significance level, it is evident that the variables Credit History, Property Area and Marital status are the only significant variables. After successive testing of attributes for a suitable model by constructing new models from the aforementioned variables using backward elimination procedure, we obtain the significant variables.

Nested models are created from the full model with all the variables and then their Akaike’s Information Criterion (AIC) values are compared. The model with the lower AIC value among the nested model is considered a better model. Later likelihood ratio tests are conducted to verify the nested models’ significance and determine the model with the best fit.

Finally, after the comparison of AIC values of the models and likelihood ratio tests, it is obtained that the final variables of significance are Credit History, Property Area and Marital status since this model exhibits the least AIC value and has the goodness of fit among the nested models created.

5. Evaluate the Model: In the evaluation step, the trained models are tested with the already separated testing dataset to evaluate the accuracy.

Then the decided model is now used to predict the chances of approval of the loan for the testing dataset. The threshold value is set at 0.5, and the prediction values of the target variable are obtained and compared to the actual values of the dataset using a confusion matrix to obtain the accuracy of the constructed model.

```
Confusion Matrix and Statistics

          Reference
Prediction  N  Y
         N 19  1
         Y 19 83

               Accuracy : 0.8361
                 95% CI : (0.7582, 0.8969)
    No Information Rate : 0.6885
    P-Value [Acc > NIR] : 0.0001559

                  Kappa : 0.5608

 Mcnemar's Test P-Value : 0.0001439

            Sensitivity : 0.9881
            Specificity : 0.5000
         Pos Pred Value : 0.8137
         Neg Pred Value : 0.9500
             Prevalence : 0.6885
         Detection Rate : 0.6803
   Detection Prevalence : 0.8361
      Balanced Accuracy : 0.7440

       'Positive' Class : Y
```

CONCLUSION

In our logistic regression model, we have finally predicted whether the loan is approved or not. Through this research model, we had implemented a loan credibility prediction system that helps the organizations make the right decision to approve or reject the loan request of the customers. This will help the banking industry to open up efficient delivery channels.

The observation inferred through this analysis is that the variable Credit History affects the chance of approval of a loan by a huge difference. The applicants with the Property Area as Sub-urban and those married applicants are more likely to get their loan approved. The accuracy of the predicted model is found to be 83.6%.

Those candidates whose credit score rating was most noticeably awful will neglect to get the advance endorsement because of a higher likelihood of not repaying the credit sum. More often than not, those candidates who have top-level salary and requests for a lower measure of advance are bound to get affirmed, which bodes well, bound to take care of their credits.

However, one disadvantage of the system being in a real-life scenario is that the recommendation system may be biased toward one dominant feature or attribute. The model concludes that a bank should not only target the rich customers for granting loans, but it should also assess the other characteristics of a customer as well, which play a very important part in credit granting decisions and predicting the loan defaulters. Therefore, using a logistic regression approach, the right customers to be targeted for granting loans can be easily detected by evaluating their likelihood of defaulting on a loan.

REFERENCES

[1] Loan prediction model using logistic regression, decision tree, random forest [1]Prof. Satish jaywant manje, [2]Mr. Dhiraj Raghunath Manje, [3]mr. Rahul Pandurang Bhere, [4]Mr. Rahul Kailas Pawade, international journal for research in engineering application & management (ijream) ISSN: 2454-9150 vol-07, special issue, May 2021

[2] An introduction to logistic regression analysis and reporting chao-ying joanne peng kuk lida lee gary m. Ingersoll Indiana university-Bloomington, the journal of educational research,2002

[3] Logistic regression-based loan approval prediction sai aparna vangaveeti1, naga likitha venna2, prasanna naga sri ramyayajamanam kidambi3, harika marneni4, naga satish kumar maganti5, journal of composition theory

[4] Ashlesha Vaidya, “Predictive and probabilistic approach by logistic regression: prediction of loan approval,” Computer Science Engineering SRM University, Chennai July 3-5, 2017, IIT Delhi

[5] “An Approach for Prediction of Loan Approval about Machine Learning Algorithm,” Mohammad Ahmad Sheikh, Proceedings of the International Conference on Electronics and Sustainable Communication Systems (2020). International Journal for Research in Engineering Application & Management (IJREAM) ISSN: 2454-9150 Vol-07, Special Issue, MAY 2021 49 | IJREAMV07I02SJ009 © 2021, IJREAM All Rights Reserved.

[6] PidikitiSupriya, “Loan Prediction using Machine Learning Models,” International Journal of Engg. and Techniques, Mar-Apr 2019.

[7] Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017