Analysis of Loan Eligibility for a Prospective Retail Borrower

Sanjay Basu[1], Divyesh Jaiswal[1], Sai Danoosh[1], Dr. Prianka RR[2]

[1]Department of Artificial Intelligence and Machine Learning, New Horizon College of Engineering, Bengaluru, India

[2]Sr Asst.Professor, New Horizon College of Engineering Bengaluru, India

ABSTRACT

This project aims to offer an individual a general estimate or idea as to the viability of home loan repayment for the loan amount incurred. This is achieved by taking into consideration the individuals qualitative and quantitative parameters. As a thumb rule, one's insurance pay out shouldn't exceed more than thirty-five percent of their income. An individual's probable EMI amount can amount to fifty percent of one's salary as per the general bank standard.

Keywords : Loan-prediction, Prospective borrower, Financial Literacy, Logistic regression, Python Flask

I. INTRODUCTION

Deep learning grants us the power to train algorithms (models) that can tackle problems based on data classification and prediction by deriving (learning) knowledge from raw data.To classify data, we will be using a Logistic regression which will be implemented using multiple in-built libraries in python like pandas, NumPy, matplotlib, sklearn and seaborn. Logistic regression is easier to implement, interpret, and very efficient to train.This project is aimed to prevent prospective borrowers from borrowing money beyond their means by incorporating machine learning.

Existing Systems: In today's day and age, people are very short-sighted when it comes to availing loans.

They make rushed and ill-informed decisions with little research and, without any regard for long term repercussions.

Moreover, what most people fail to realise is, that banks as per their respective loan eligibility models, seek to extract as much EMI pay out as possible for a lengthy duration of time.

This existing approach of impulsive borrowing on part of the borrower is very problematic, leading to a decrease in their standard of living, compromising on important purchases, decreased wealth creation, possibility of loan default, increase in retirement age, etc.

Proposed System: To make an application which helps borrowers make an informed decision on the loan amount taken, based on certain qualitative and quantitative parameters.
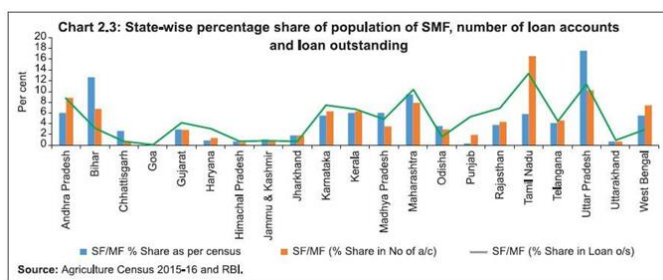
Loan Amount Determination: As a thumb rule, one's insurance payout shouldn't exceed more than thirty-five percent of their income.

An individual's probable EMI amount can amount to fifty percent of one's salary as per the general bank standard.

An estimate to an individuals loan eligibility potential is ascertained by the formula,

Loan amt = (Monthly Fixed Income * 0.5) – (Other EMI)

Per Lack EMI

Chart 2.3: State-wise percentage share of population of SMF, number of loan accounts and loan outstanding

Source: Agriculture Census 2015-16 and RBI.

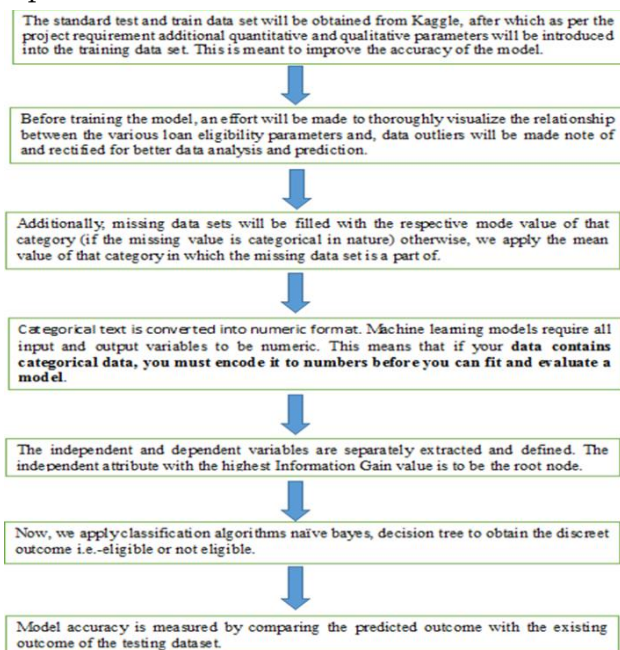## II. Flowchart

Logistic Regression:

As here we wasn't to classify between the people who have taken loan or not we have used Logistic Regression. The purpose of this algorithm is to find a plane that separates two types. Y variable belongs to 1 or 0. In 2–d we have to figure out a line that exactly separates two classes.

3.1 General Flow chart

Fig 3.2 depicts the general flow of execution of the application

3.2 Description of project flow:

Note: Elaborates on Fig. 3.2

In the innit .py file, we have created a flask application and, initialised its secret key i.e. that which encrypts or secures the session data.

The main.py file is where the website package is imported and the created file function defined earlier is invoked. This is what is used to create and run the application.

The views.py file stores the routes for the website i.e., where the users can access the various pages of the webapp. Similarly, auth.py file is where routes of pages relation the user authentication is stored.

Thus, all the webpages created are stored in the templates folder and, their respective routes are defined in the views.py or auth.py files depending on the purpose they serve.

Web page structuring and content addition are done primarily using html, Bootstrap (an open-source CSS frame work).

There is also a special templating language used in Flask that allows us to write a little bit of Python inside the HTML documents (Eliminates the need for java script for this project).

A base template is also defined as seen in the base.html file. This basically helps in setting up the theme of the website (like navigation bars, header, footer, etc...). Thus, when creating more templates for more webpages we have overridden parts of the base template with more specific templates. Additionally, stylesheets imported into the base templates can be used by other templates through this overriding.

The above points provide an overview as to how we went about including webpages into the web application.

The ML model and its function have been typed into the views.py file. Upon entering the details in the train page as seen in the webapp and, clicking the 'predict!' button, all the data is sent into the views.pred function.

Thereupon, the model is trained and, the data form the input field are accordingly stored into successive variables. Then, the function call is made to the predict function which is defined in views.py file.

After which, using the already trained model, an individual's loan status is accordingly displayed.

4.1 Pseudocode for the ML model:

Step 1: Start, Initialize "model" with the type of classifier, i.e., logistic regression. Initialize "train file" and "test file" with the path of the training dataset and the testing dataset respectively. Initialize the columns in multiple lists based on whether the null values will be replaced using the mode or the median and the columns which have to be encoded

Step 2: Define module "load train dataset". Read the training file and store it in "df train". Replace the null values in the dataset with the mode or median as necessary and then print it.

Return "df train" and end the module "load train dataset". Define module "load test dataset"

Define module "train with cv". Initialize "X" as a dataframe that contains all the relevant features and "Y" holding the target variable i.e., Loan Status.

Step 3: Convert the categorical values to numerical values. Split the training dataset such that 75% will be used to train the model and 25% is used to test the model and "x cv" will hold the x values for cross validation and "y cv" holds the loan status for the corresponding "x cv" values. Train the model using the fit function. Predict the Loan status using the predict function and passing "x cv". Check the accuracy using "y cv". End "train with cv" module . Define module "predict test". Initialize "x" with only the relevant features. Convert the categorical values to numerical in "x". Predict the result by passing X to the predict function and print it. End "predict test" module

Step 4: Define module "train". Clean the dataset using the "load train dataset" module and store in "df train". Initialize "x" and "y" with x being only the relevant features and y being the target variable i.e., Loan Status. Convert categorical variables to numeric. Train the model using the fit function, End "train" module. Define "Predict" module. Accept the information from the user and convert it into a data frame (named "x").

Step 5: Convert categorical values to numerical. Call the predict function and then print it. Return the result. End "predict" module.End

III. EXAMPLES

Ramesh is a fifty year old male looking to purchase some commercial property to rent out to companies and earn a steady stream of passive income from his hard assets

Lets say he has to rent out 500sqft in Noida. So the only way the WTC will approve this purchase is if he pays upfront 17 lakhs and pays the remaining 10 lakhs over the course of two years utmost.

The issues is he doesn't have substantial savings to pay the money upfront as most of his wealth is tied to non-liquid assets like jewelry, other real estate (on which he is paying EMI still) and debentures with a fixed holding period of 10 years.

He has a family of four to take care of on top of which being 50 he is at the tail end of his career as he plans on retiring by the age of 55. Now the crux of the matter is, is it a viable option for him to avail a loan of such a substantial figure from a bank and what adds to his conundrum is the fact that being 50 years of age his loan repayment rate will be quite high (as he is approaching retirement age and has multiple dependents)

Thus an individual like Ramesh might make use of an application like ours to properly weigh all of his options and make the decision on whether or not he can finance this loan until its full repayment without there being a decline in his or his families standard of living.

Or application is tailor made to cater to the needs of the prospective borrower to help them to make an informed decision on the availing of a loan.

IV. CHALLENGES

In today's economic climate access to readily available and cheap credit has been the primary driver of economic growth in most developed and developing countries.

The issue lies in the fact that credit or debt has become so easy to get that this has created the systemic issue of non- performing assets in a lot of the major financial institutions all across the developing and developed world.

It is in the best interest of banks to provide loans especially those loans which can be easily accounted for by pledging the borrower's collateral.

The cheap credit system has encouraged the common man to recklessly avail one loan after another without taking into account the potential ramifications on his financial health.

Should the borrower default on his loan the banks are protected as they can simply seize the loan defaulters' assets. Thus, leaving the borrower in a precarious situation.

Nowadays among millennials there is a noticeable trend of them misusing their credit cards. This issue is indicative of the spendthrift mindset that has plagued todays youth.

An example of the above would-be millennials pilling upon credit card debt to pay for their fast fashion accessories (these include the weekly line of clothes that are advertised by major clothing franchises, designer purses to name a few.

This fast fashion brands look to wholly exploit their primary customer base into buying their weekly catalogues. Thus, these fast fashion brands have tied up with credit card companies to ensure that customers using the above-mentioned credit card companies credit card to shop at these stores have limited restrictions imposed on the upper limit of the daily credit allowance

V. RESULTS

When all necessary details are entered into the input fields of the train page and the "Predict!" button is clicked, immediately the model is trained and the data is sent to the variables defined in the views.pred function

Further, the data is sent to the predict function in views.py and a clear-cut response is generated i.e., if that person is eligible or non-eligible
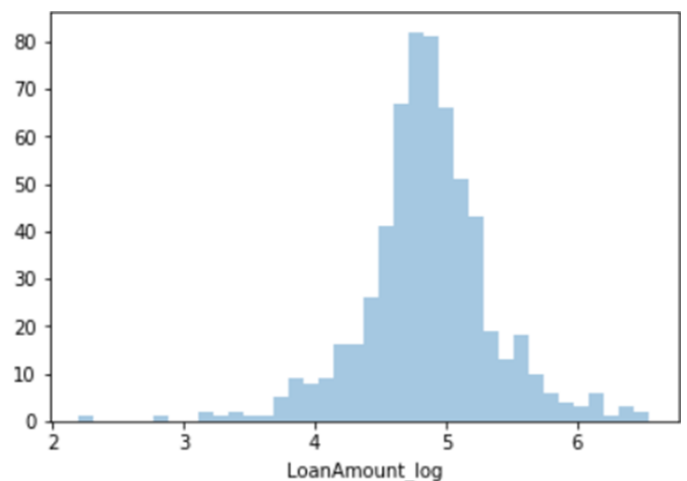
VI. CONCLUSION

Currently we have not thought of feature additions to our program, but we do have plans to increase the accuracy of our model and streamlining our program to the max so we can get as perfect of a result as possible.

In the future we might include more outliers in the model, and a few more variables if possible (if it makes a noticeable difference) to make the program more applicable to a wide range of variables/situations.

On top of these under the hood enhancements, we would also make some cosmetic updates which would make the UI easier to use for the average person, while also being very informative.

VII. REFERENCES

[1]. J. R. Quinlan. Induction of Decision Tree. Machine Learning, Vol. 1, No. 1. pp. 81-106., 1086.

[2]. A. Goyal and R. Kaur, "A survey on Ensemble Model for Loan Prediction", International Journal of Engineering.

[3]. G. Shaath, "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R".

[4]. A. Goyal and R. Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models".

[5]. https://www.experian.com/blogs/ask-experian/credit-   education/score-basics/what-is-a-good-credit-score/

[6]. https://machinelearningmastery.com/types-of-classification-in-machine-learning/   computer applications

[7]. T. Harris, "Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions", Expert Systems with Applications,vol.40,pp. 4404– 4413, 2013.

[8]. 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