Bank loan prediction using Logistic Regression

Ms. S. Sivaselvi AP/CSE, V. Anupriya, P. Babhu Ganesh, S. Rajapandi,
Dept of CSE, Erode Sengunthar Engineering College,
Erode, Tamil Nadu

ABSTRACT

For financial companies, the loan approval process is crucial. For numerous difficulties, the banking industry is always in need of a more precise predictive modelling system. For the financial sector, predicting credit defaulters is a challenging issue. The loan applications were approved or rejected by the system. A key determining factor in a bank's financial results is loan recovery. Predicting whether a consumer will pay back a debt is exceedingly tough. For vast amounts of data, machine learning (ML) techniques are highly helpful in predicting outcomes. Three machine learning methods, Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF).In this study, are utilised to predict whether consumers will be authorised for loans. The experimental results show that the Logistic Regression will be better whencompared to other two algorithms.

Keywords: - Loan, Machine Learning, Training,Testing, Prediction.

I. INTRODUCTION

Banks all throughout the world frequently examine bank credit risk. Risk level computation uses a range of methodologies since credit risk rating is so important. Additionally, Credit risk management is among the key functions of the banking business [1][2][3].

The main business of practically all banks is the distribution of loans. The vast bulk of a bank's assets are owed to earnings on loans that the business has disbursed [4][5]. The primary purpose in a banking setting is to place one's assets in trustworthy hands. Today, many banks and financial institutions grant loans after a protracted process of validation and verification, but it is still uncertain if the chosen candidate is the most worthy one among all applicants [6][7][8]. With the help of this system, we can determine whether a specific application is secure or not, and the entire process of feature validation is automated using machine learning [6][7]. This model's drawback is that it emphasizes distinct weights for each aspect, however in reality loans are occasionally accepted based entirely on a single, major criterion, which is not possible with this method [1].

Both applicants and bank employees benefit greatly from loan prediction. The goal of this study is to provide a quick, simple, and direct approach for choosing competent applicants [8]. It may provide the bank with distinct advantages. [2][3][4]. It may automatically weight each loan processing characteristic, and on fresh test data, the same characteristics are processed in line with their respective weights. [5][6][7]. It is feasible to give the applicant a deadline to ascertain if it is or not their loan will be accepted. When you navigate to a single application in it, you can evaluate it in priority order [8]. The controlling authority of the bank or finance firm is the only audience for this paper, and the entire prediction procedure is carried out in confidence without the involvement of any other parties [2]. Other bank departments can receive results for a specific Loan Id, enabling them to respond to applications in the most efficient way possible. This makes it easier for the other departments to complete other formalities[3].

II. DATA SET

The data set has now been sent to the machine learning model, and the model has been trained using this data set. The information given on the application form by each new applicant acts as a test set of data. Based on the inferences drawn from the
training data sets, the model predicts whether the new applicant is a suitable candidate for loan approval or not after testing.

Variable Name	Description of Variable
Loan ID	A Uniques Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Graduate/ Under Graduate
Self_Employed	Self Implloyed (Y/N)
Applicant Income	Applicant income
Coapplicant Income	Coapplicant income
Loan_Amount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi Urban/ Rural
Loan_Status	Loan Approved(Y/N)

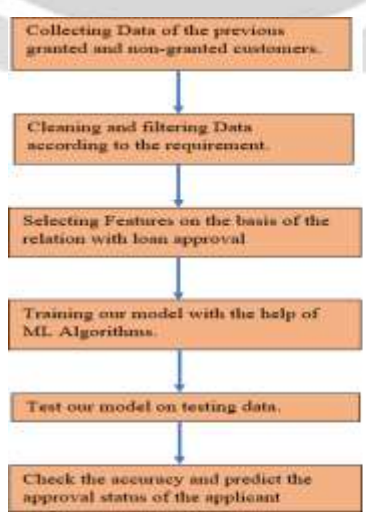
2.1. MACHINE LEARNING METHODS:

For the prediction of Android applications, six machine learning classification models have been utilised. The models can be found in the free R software. [6] A GNU GPL licence applies to R. Belowis a brief description of each model.

2.1.2. Loan Prediction Methodology

2.1.1. Decision Trees (DT):

Decision trees' fundamental algorithm [7] calls for the discretization of all attributes and features. The selection of features is dependent on which features provide the most information. IF-THEN regulations can be made used to describe the knowledge [8]. Thismodel is a development of Quinlan's C4.5 classification methods.



2.1.3. Random Forest (RF):

To characterise (and relapse), Random Forests
[3] generate a lot of Decision Trees. Both simultaneously period and provide a class that is the average of the classes produced by individual trees [3].

2.1.4. Logistic Regression (LR):

A statistical analysis method called logistic regression uses previous observations from a data set to predict a binary outcome, such as yes or no. By examining the correlation among a few, already present independent variables, a pattern forecasts a the dimensional data parameter.

s

2.1.5. Pandas:

The Python module Pandas [4] makes it quick, simple, and straightforward to work with structured as well as regression analysis. It's quite meant to act as the fundamental, advanced building block for Python-based useful, extensive data mining. The main goal of the project is to become the The much more versatile and effective open-source data analysis and modification tool in any domain. It is currently rapidly heading in this way.

III. CODING

```
"cells": [
  {
   "cell_type": "code", "execution_count": ,"metadata":
   {},
   "out puts"
   : [],
   "source":
   [
    "import pandasas pd\n", "importnumpy as np"
   ]
  },
  {
 "cell_type": "code", "execution_count": , "metadata": {}, "outputs": [], "source":
 [ "importos\n", "os.chdir(\"E:\\\\MY
DATA\\\\P\\\\batch\\\\2. LOAN APPROVAL\")"
 ]
 },
 {
 "cell_type": "code","execution_count": ,"metadata": {},
 "outputs": [],"source": [
  "loan_train=pd.read_csv(\"train.csv\")"
 ]
 },
 {
 "cell_type": "code","execution_count": ,"metadata": {}, "outputs": [
  {
   "data": { "text/plain": ["(614, 13)"
    ]
   },
   "execution_count": ,"metadata": {},
   "output_type": "execute_result"
  },
```

```
"source": [ "loan_train.shape"
]
},
{
"cell_type": "code", "execution_count": 5,"metadata":
{}, "outputs": [
 {
  "data": { "text/plain": ["0"
   ]
  },
  "execution_count": 5,"metadata": {},
  "output_type": "execute_result"
 }
],
   "source": [ "loan_train.duplicated().sum()"
   ]
  },
  {
   "cell_type": "code", "execution_count": 6,"metadata": {}, "outputs": [
    {
     "name": "stdout", "output_type": "stream","text": [
      "<class 'pandas.core.frame.DataFrame'>\n",

      "RangeIndex: 614 entries, 0 to 613\n", "Data columns (total 13 columns): \n", "Loan_ID 614 non-null object\n",
      "Gender            601 non-null object\n",
        "Married             611 non-nullobject\n", "Dependents
                            599 non-null object\n", "Education 614 non-nullobject\n", "Self_Employed
                               582 non-nullobject\n", "ApplicantIncome
                               614 non-null int64\n", "CoapplicantIncome 614 non-null float64\n", "LoanAmount
                               592 non-nullfloat64\n", "Loan_Amount_Term
                            600 non-nullfloat64\n","Credit_History
                             564 non-nullfloat64\n", "Property_Area
                             614 non-nullobject\n", "Loan_Status
                            614 non-null object\n", "dtypes: float64(4), int64(1),object(8)\n", "memory usage: 62.4+ KB\n"
       ]
       }"source": [

  "loan_train.isna().sum().sort_values(ascending=False
  )"
    ]
   }
        "dtype: int64"
       ]
      },
      "execution_count": 8,
      "metadata": {},
      "output_type": "execute_result"
     }
     ],
```

IV. RESULT

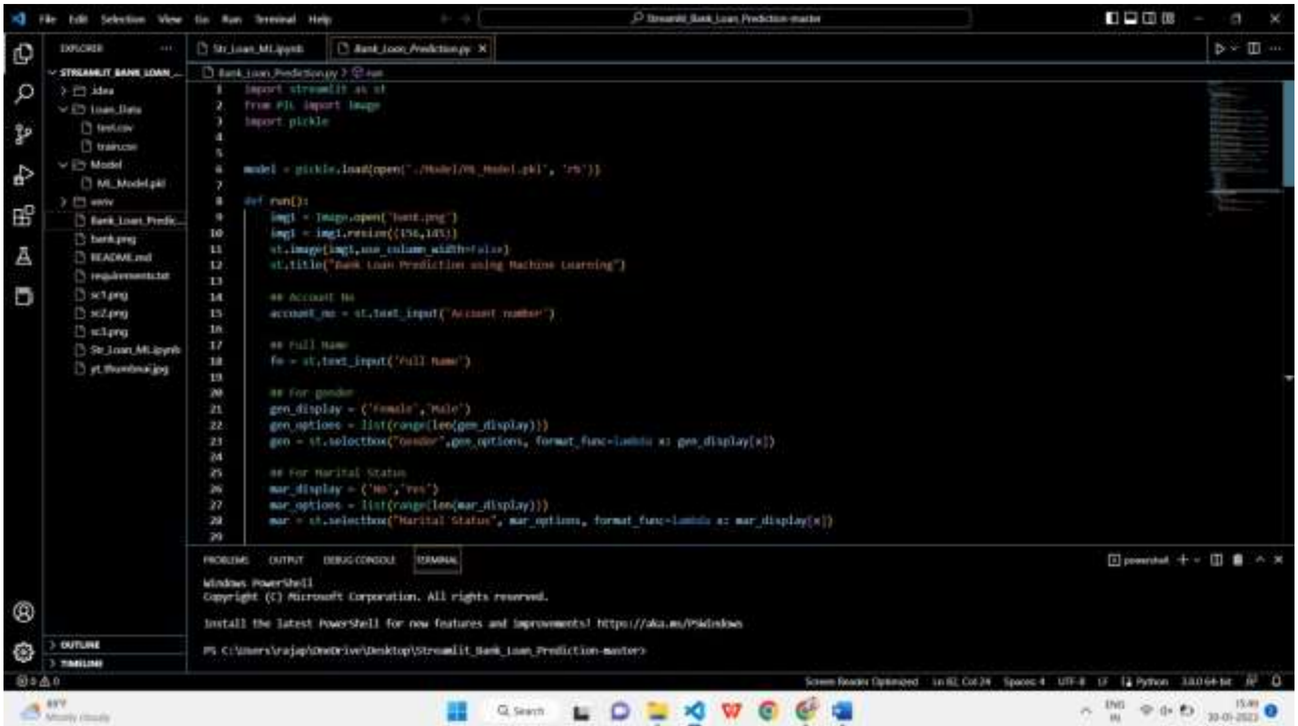


Figure 1: Loan Train Data set

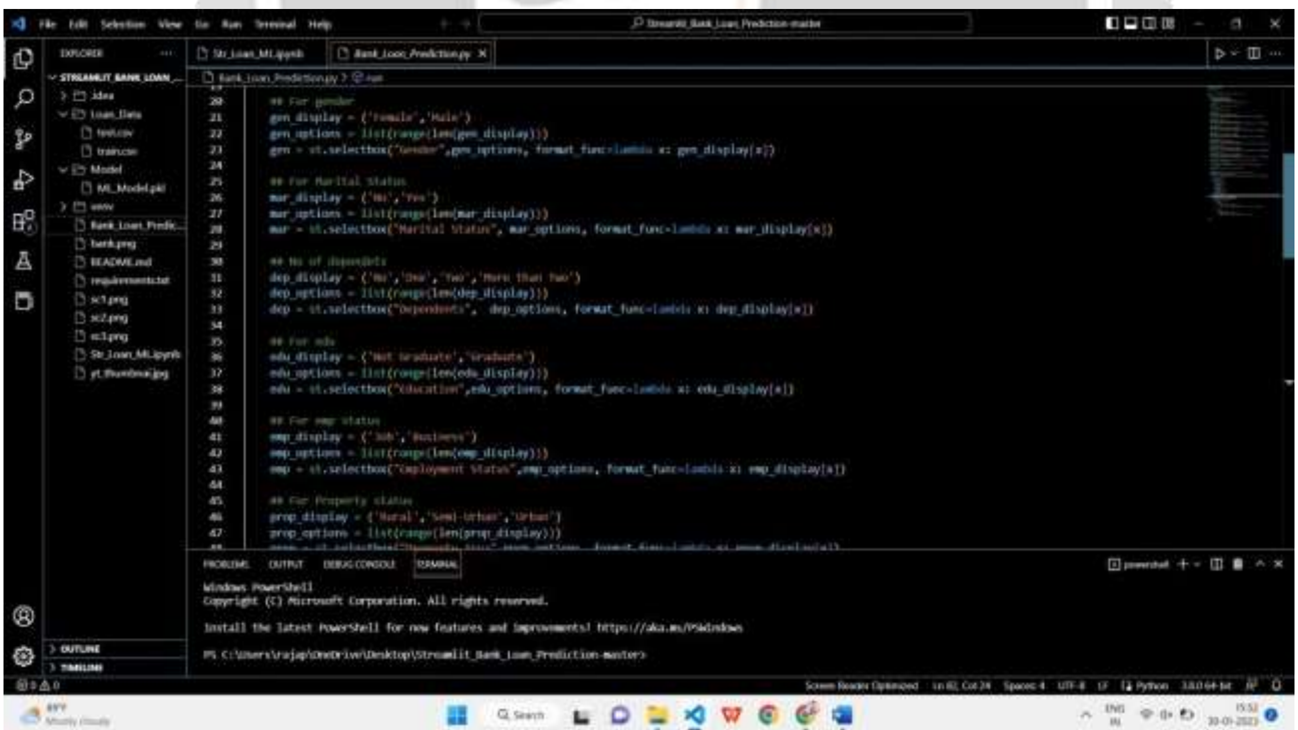


Figure 2: Loan Train Data set on gender and Dependents

Figure 3: Live at streamlit Filling Information

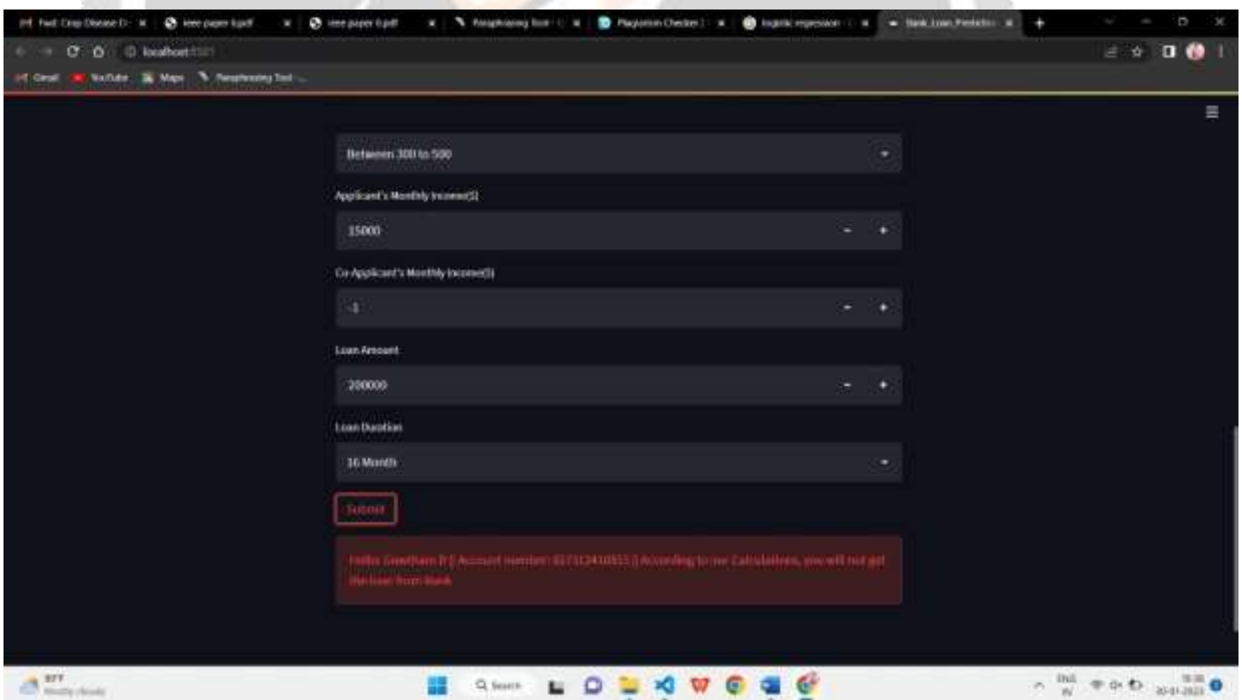


Figure 4: Live at streamlit Final output

V. CONCLUSION

Following a comprehensive assessment of the component's benefits and downsides, it is safe to conclude that the prediction is an efficiently high element. This module is functional and satisfies all lender criteria. This component is simple to include into a number of other systems. There have been several reports of computer errors, content issues, and, most importantly, the weight of features being changed in automated likely systems. Like a result, the software described above may be upgraded in the near future to be more safe, reliable, and capable of dynamic weight modification. In the near future, this prediction module will be linked to the automated processing system module.

REFERENCES

[1]. Anshika Gupta, Vinay Pant, Sudhanshu Kumar and Pravesh Kumar Bansal: “Bank Loan Prediction System Using Machine Learning” -2020, IEEE System Modeling &Advancement in Research Trends.

[2]. Mohammad Ahmad Sheikh, Amit Kumar, Tapas Kumar: “An Approach for Prediction of Loan Approval Using Machine Learning Algorithm”-2020, IEEE Electronics and Sustainable Communication Systems.

[3] A Peda Gopi and Lakshman Narayana Vejendla, "Certified Node Frequency in Social Networks Using Parallel Diffusion Methods," Ingénierie des Systèmes 'Information, Vol. 24, No. 1, 2019, pp.113-117. The DOI is 10.18280/isi.240117.

[4] Santhi Sri Kurra and Ande Prasad N.Veeranjaneyulu and R S M Lakshmi Patibandla (2015) published “unstructured data (2015) published "Unstructured Data: QualitativeAnalysis" in J. of Computing In Biosciences AndEngineering, Vol. 2, No. 3, pp.1-4.

[5]. L.Jagajeevan Rao, M. Venkata Rao, T.Vijaya Saradhi (2016), “How the Smartcard Makes the Certification Verification Easy” Journal ofTheoretical and Applied Information Technology, Vol.83. No.2, pp. 180-186.

[6]. K.Santhi Sri and PRSM Lakshmi,(2017),“DDoS Attacks, Detection Parameters andMitigation in Cloud Environment”, IJMTST, Vol.3, No.1, pp.79-82.

[7]. VellalacheruvuPavani and I. Ramesh Babu (2019),” Three Level Cloud Storage Scheme for Providing Privacy Preserving using Edge Computing”,International Journal of Advanced Science and Technology Vol. 28, No. 16, pp. 1929 – 1940.

[8]. B.V.Suresh kumar,Sk.Reshmi Khadherbhi,BIG-IOT Framework Applications and Challenges: A Survey Volume 7, Issue VII,JULY/2018 pg.no 1257-1264.