GOVERNMENT OF TAMILNADU

Naan Muthalvan - Project-Based Experiential Learning

PREDICTING PERSONAL LOAN APPROVAL USING MACHINE LEARNING

Submitted by

M.GAYATHRI – (20626ER043)

V.ABARNA – (20626ER037)

M.GEETHA– (20626ER044)

D.SORNAMBIKA – (20626ER064)

C.SWATHI-(20626ER065)

Under the guidance of

Mrs.K.P.KAVITHA, MCA.,M.Phil.,(Ph.D)

Department of Computer Science

GOVERNMENT ARTS COLLEGE FOR WOMEN , NILAKOTTAI

B(Affiliated To Mother Teresa Women's University, Kodaikanal)

Reaccredited with "C" Grade by NAAC

NILAKOTTAI - 624208

APRIL - 2023

GOVERNMENT ARTS COLLEGE FOR WOMEN, NILAKOTTAI
(Affiliated to Mother Teresa Womens University, Kodaikanal)
Reaccredited with 'C' Grade by NAAC
Nilakottai - 624208

DEPARTMENT OF COMPUTER SCIENCE

BONAFIDE CERTIFICATE

This is to certify that this is a bonafide record of the project entitled **PREDICTING PERSONAL LOAN APPROVAL USING MACHINE LEARNING** done **By MS.M.GAYATHRI - (20626ER043), MS.V.ABARNA – (20626ER037), MS.M.GEETHA - (20626ER044), MS.D.SORNAMBIKA - (20626ER064), MS.C.SWATHI-(20626ER065)** This is submitted in partial fulfillment for the award of the degree of **Bachelor of Science in Computer Science in GOVERNMENT ARTS COLLEGE FOR WOMEN, NILAKOTTAI** during the period of December 2022 to April 2023.

Project Mentor(s) **Head of the Department**

Submitted for viva-voce Examination held on _______________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

CONTENT

PREDICTING PERSONAL LOAN APPROVAL USING MECHINE LEARNING

1. Introduction

A loan is the major source of income for the banking sector of financial risk for banks. Large portions of a bank's assets directly come from the interest earned on loans given. The activity of lending loans carry great risks including the inability of borrower to pay back the loan by the stipulated time. It is referred as "credit risk". A candidate's worthiness for loan approval or rejection was based on a numerical score called "credit score". Therefore, the goal of this paper is to discuss the application of different Machine Learning approach which accurately identifies whom to lend loan to and help banks identify the loan defaulters for much-reduced credit risk.

1.1 Overview

Have you ever thought the apps which can predict whether you will get your loan approved or not work? We are going to develop one such model which can predict whether a person will get his/her loan approved or not by using some of the background information of the applicant like the applicant's gender, marital status, income, etc.

1.2 Purpose

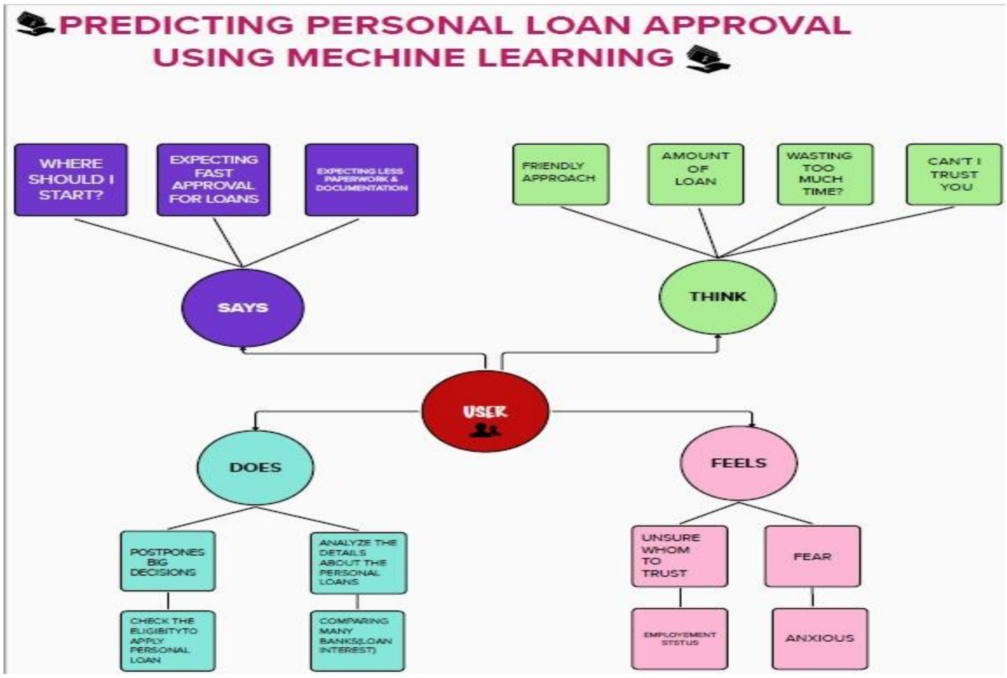
LOANS are the major requirement of the modern world. By this only, Banks get a major part of the total profit. It is beneficial for students to manage their education and living expenses, and for people to buy any kind of luxury like houses, cars, etc.

But when it comes to deciding whether the applicant's profile is relevant to be granted with loan or not. Banks have to look after many aspects.

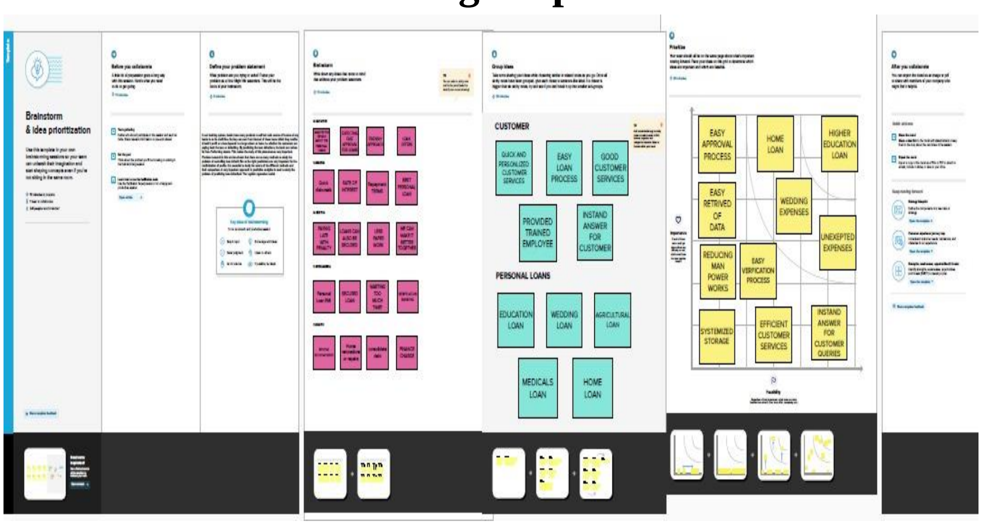
So, here we will be using Machine Learning with Python to ease their work and predict whether the candidate's profile is relevant or not using key features like Marital Status, Education, Applicant Income, Credit History, etc.

2. Problem definition & Design Thinking

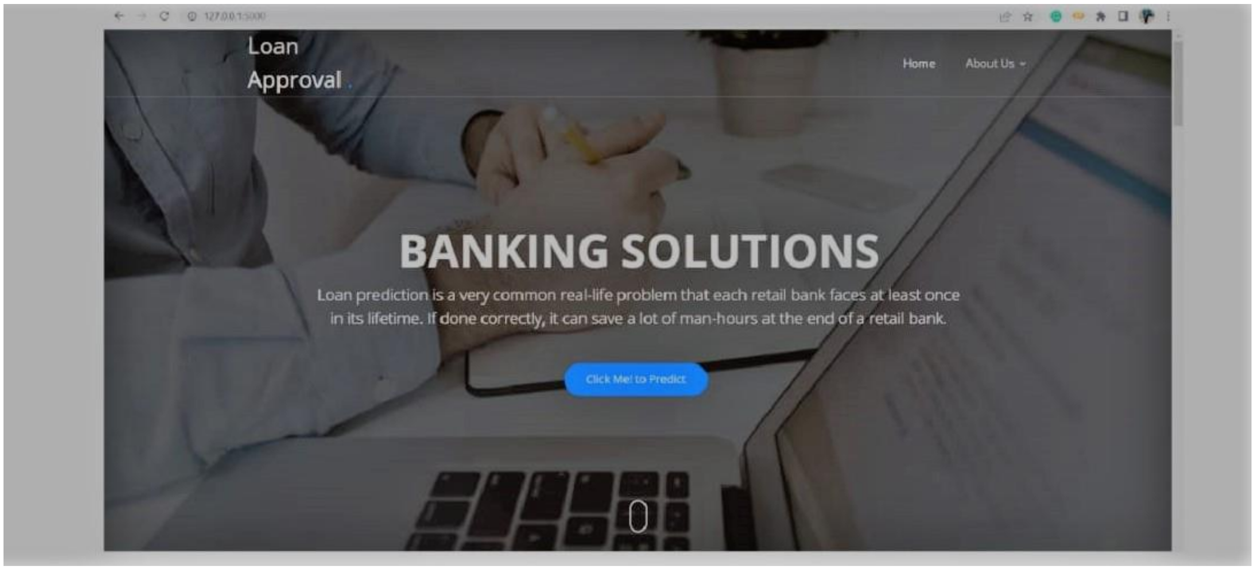
2.1Empathy map

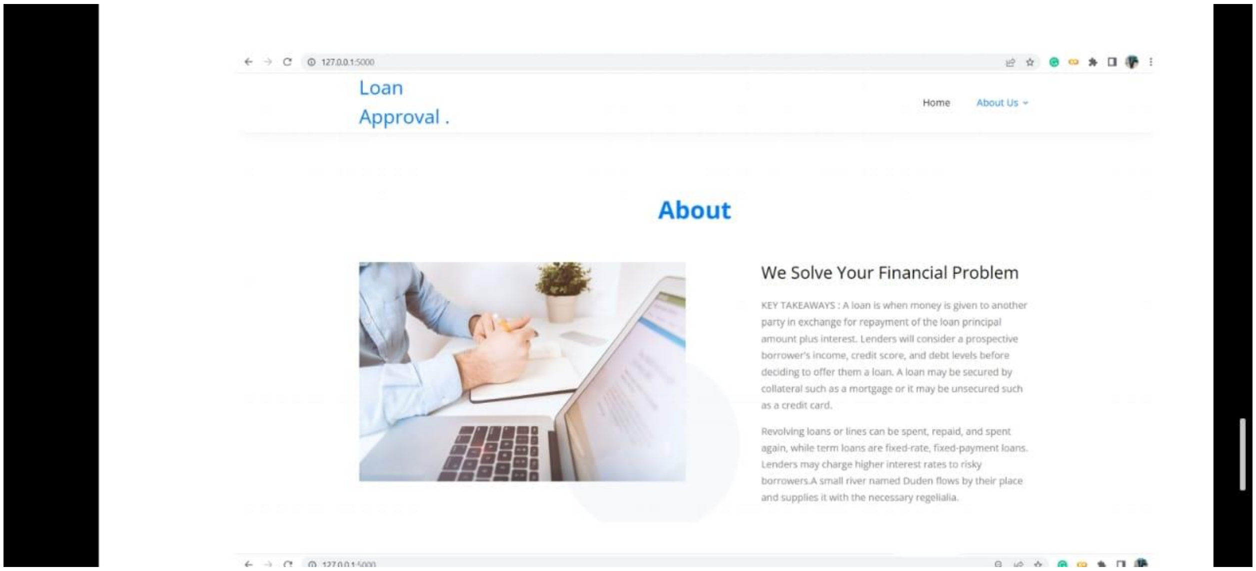


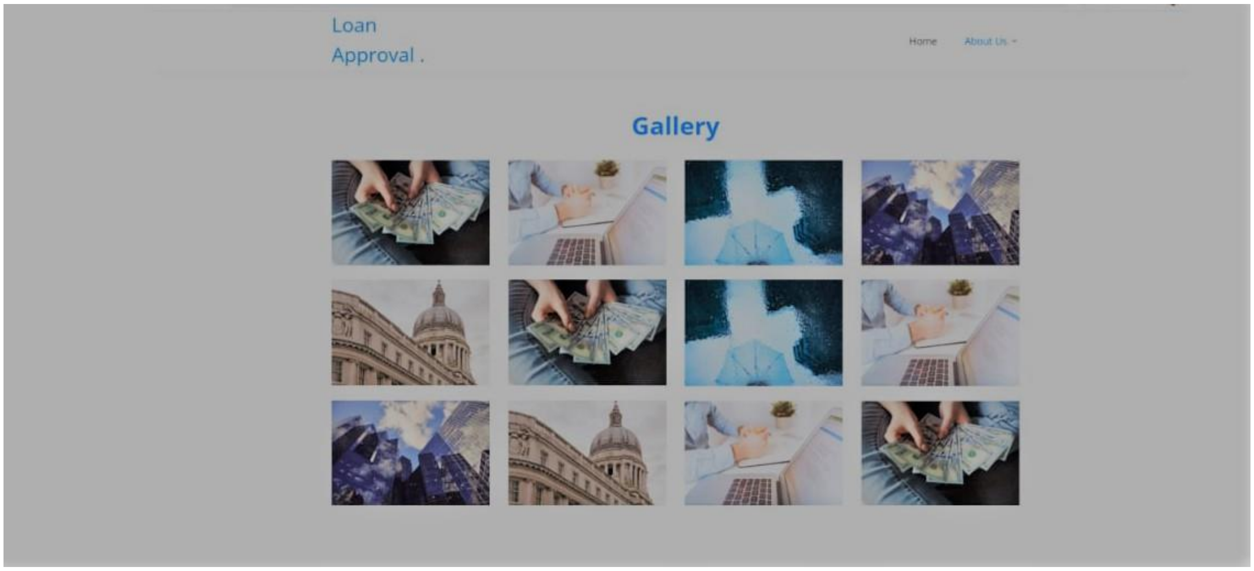
2.2 Ideation & Brainstorming Map



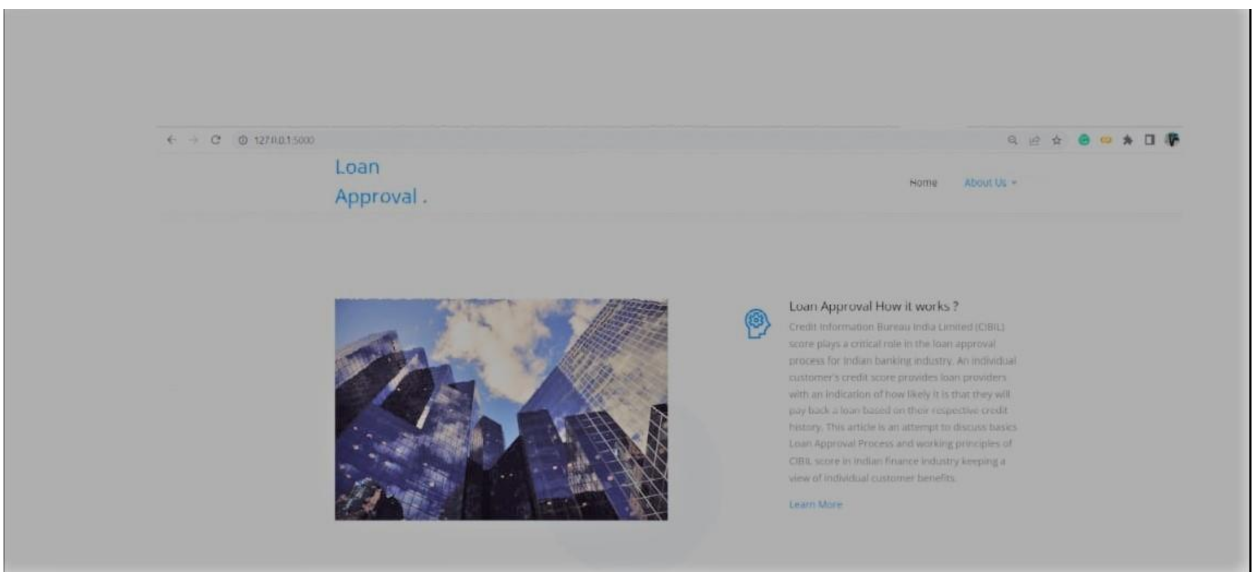
3.RESULT


Loan
Approval .
Home
About Us
BANKING SOLUTIONS
Loan prediction is a very common real-life problem that each retail bank faces at least once in its lifetime. If done correctly, it can save a lot of man-hours at the end of a retail bank.
Click Me! to Predict

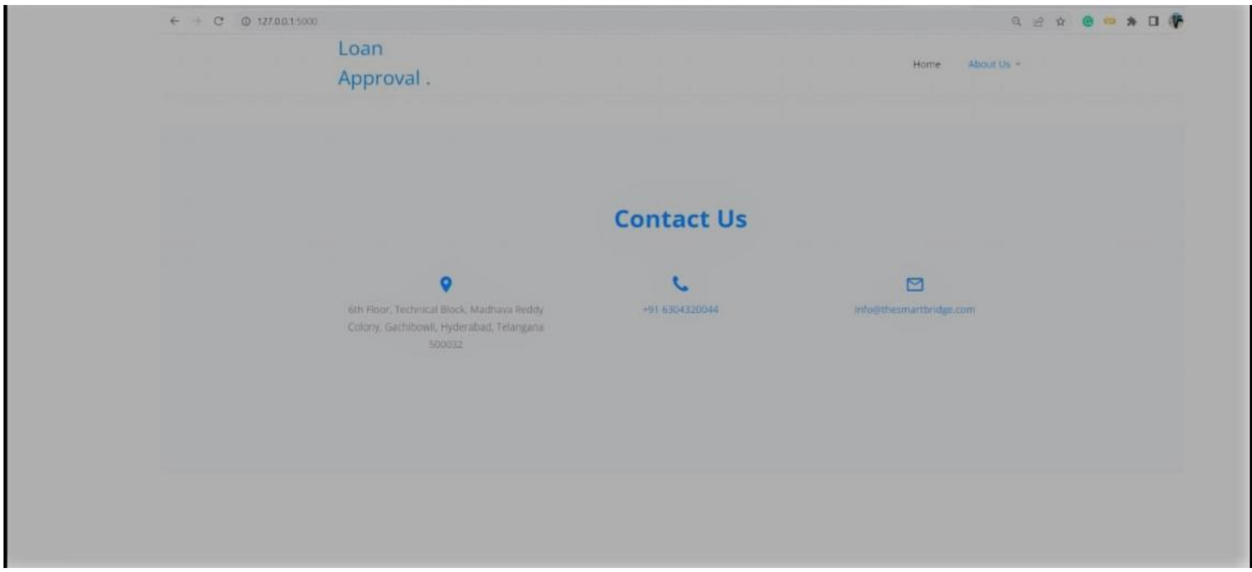

Loan
Approval .
Home
About Us
About
We Solve Your Financial Problem
KEY TAKEAWAYS : A loan is when money is given to another party in exchange for repayment of the loan principal amount plus interest. Lenders will consider a prospective borrower's income, credit score, and debt levels before deciding to offer them a loan. A loan may be secured by collateral such as a mortgage or it may be unsecured such as a credit card.
Revolving loans or lines can be spent, repaid, and spent again, while term loans are fixed-rate, fixed-payment loans. Lenders may charge higher interest rates to risky borrowers.A small river named Duden flows by their place and supplies it with the necessary regelialia.

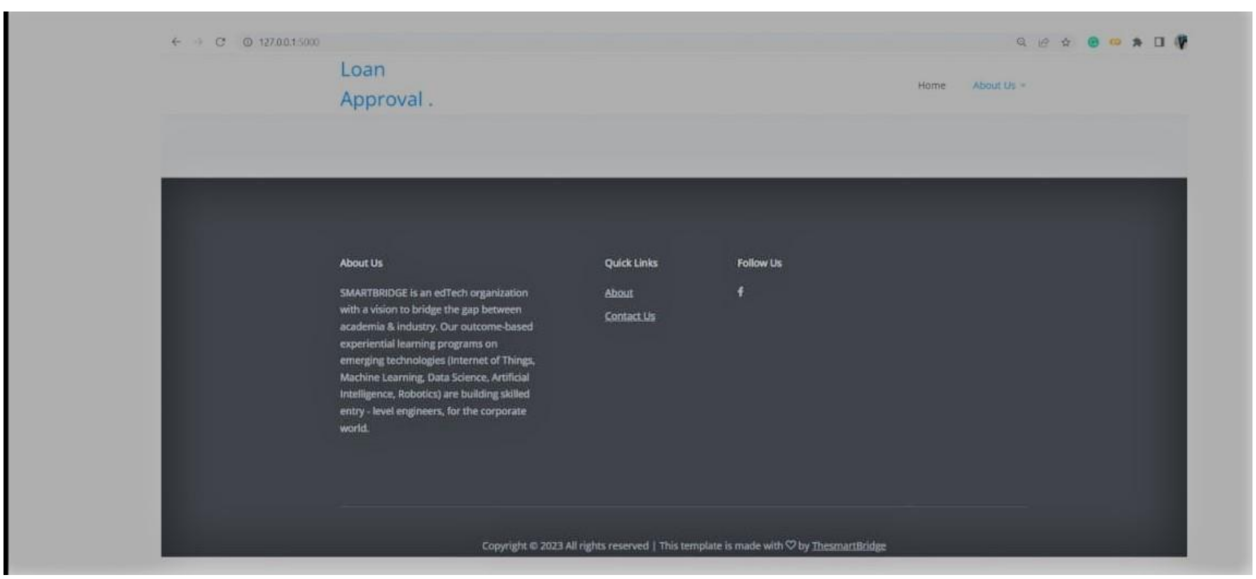


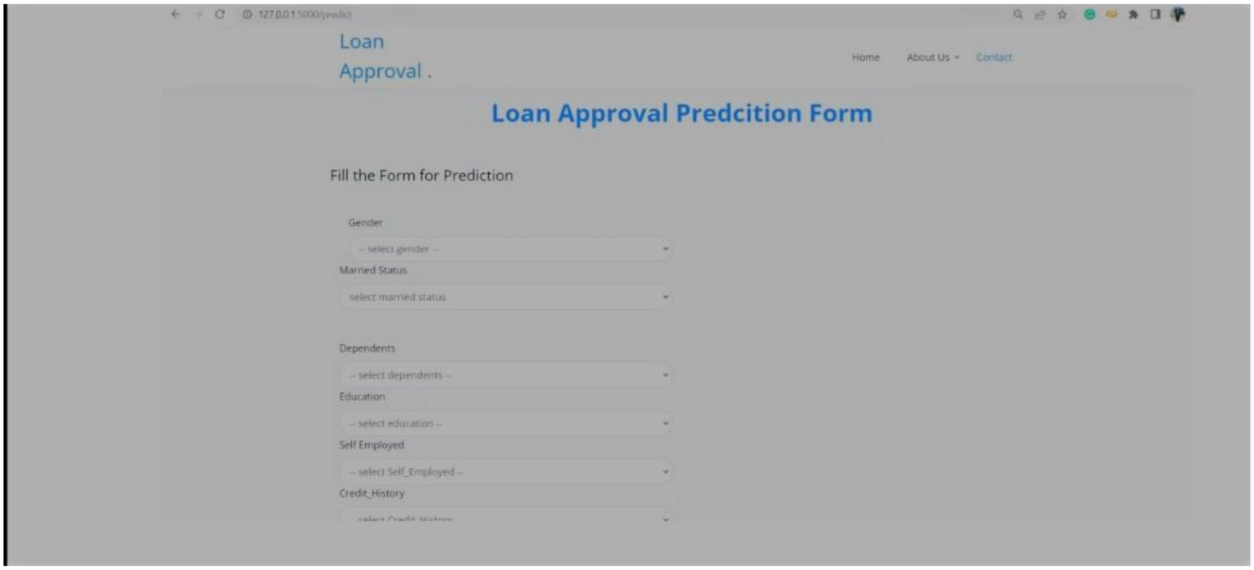
Loan
Approval .
Home
About Us
Gallery

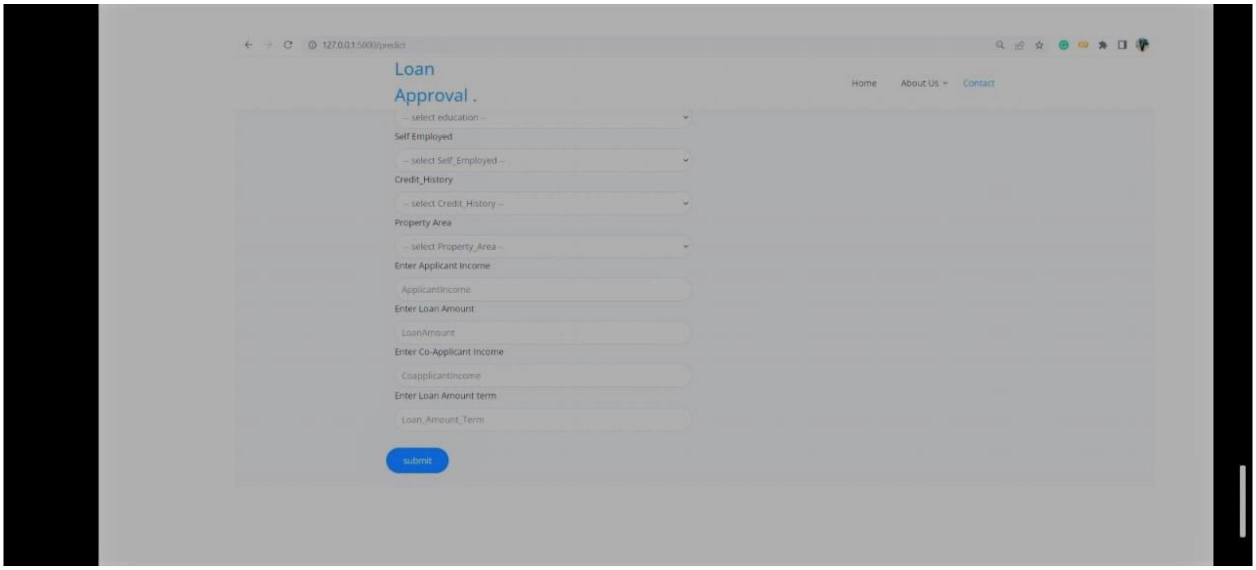


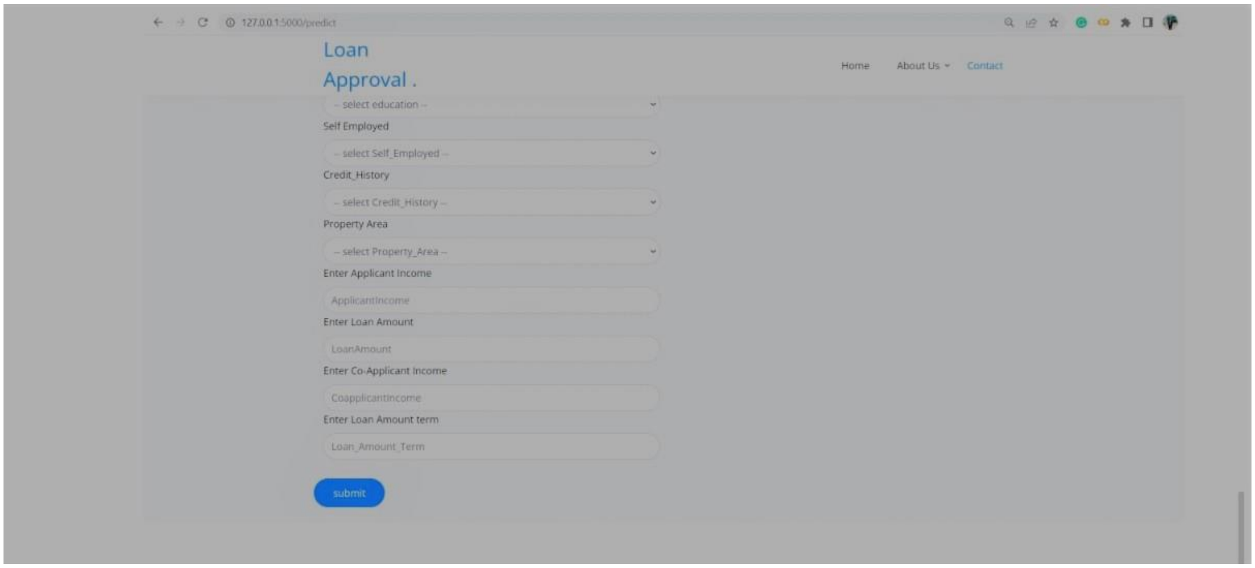
127.0.0.1:5000
Loan
Approval .
Home
About Us
Loan Approval How it works ?
Credit Information Bureau India Limited (CIBIL) score plays a critical role in the loan approval process for Indian banking industry. An individual customer's credit score provides loan providers with an indication of how likely it is that they will pay back a loan based on their respective credit history. This article is an attempt to discuss basics Loan Approval Process and working principles of CIBIL score in Indian finance industry keeping a view of individual customer benefits.
Learn More

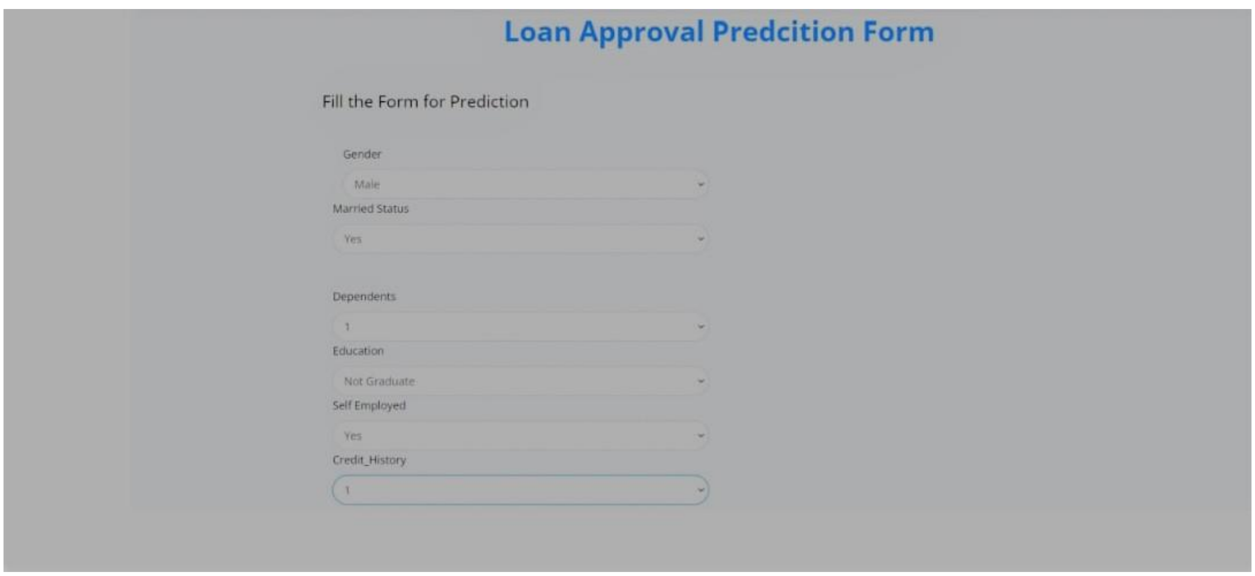

127.0.0.1:5000
Loan Approval .
Home
About Us
Contact Us
6th Floor, Technical Block, Madhava Reddy Colony, Gachibowli, Hyderabad, Telangana 500032
+91 6304320044
info@thesmartbridge.com

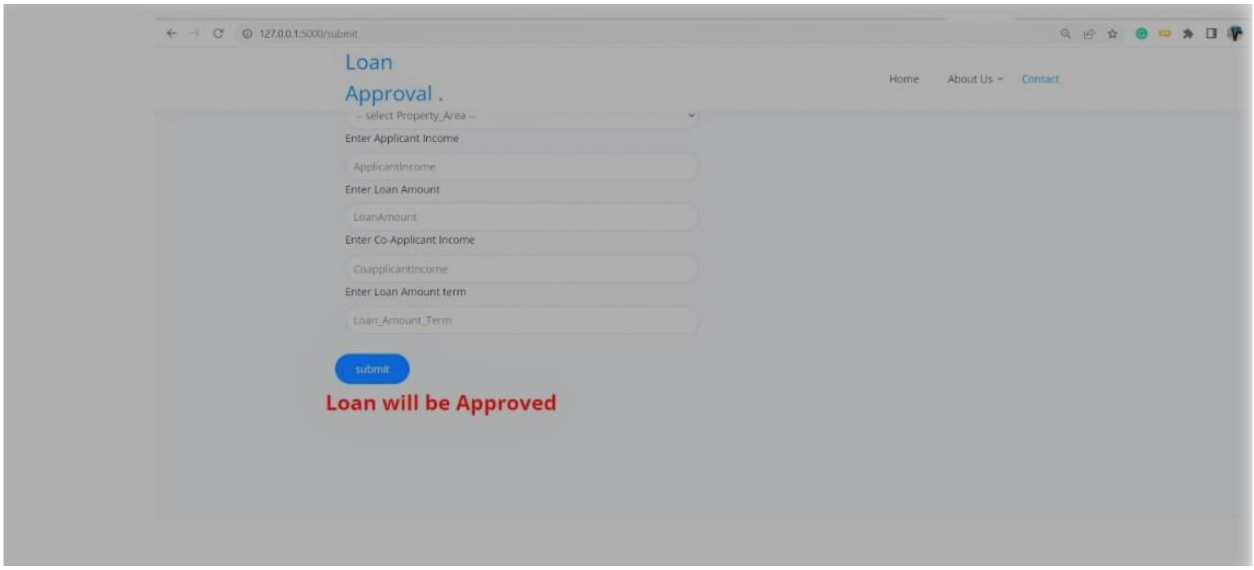

127.0.0.1:5000
Loan Approval .
Home
About Us
About Us
SMARTBRIDGE is an edTech organization with a vision to bridge the gap between academia & industry. Our outcome-based experiential learning programs on emerging technologies (Internet of Things, Machine Learning, Data Science, Artificial Intelligence, Robotics) are building skilled entry - level engineers, for the corporate world.
Quick Links
About
Contact Us
Follow Us
Copyright © 2023 All rights reserved | This template is made with ♡ by ThesmartBridge


127.0.0.1:5000/predict
Loan
Approval .
Home
About Us
Contact
Loan Approval Predcition Form
Fill the Form for Prediction
Gender
-- select gender --
Married Status
select married status
Dependents
-- select dependents --
Education
-- select education --
Self Employed
-- select Self_Employed --
Credit_History


127.0.0.1:5000/predict
Loan
Approval .
Home
About Us
Contact
-- select education --
Self Employed
-- select Self_Employed --
Credit_History
-- select Credit_History --
Property Area
-- select Property_Area --
Enter Applicant Income
ApplicantIncome
Enter Loan Amount
LoanAmount
Enter Co-Applicant Income
CoapplicantIncome
Enter Loan Amount term
Loan_Amount_Term
submit


127.0.0.1:5000/predict
Loan
Approval .
Home
About Us
Contact
-- select education --
Self Employed
-- select Self_Employed --
Credit_History
-- select Credit_History --
Property Area
-- select Property_Area --
Enter Applicant Income
ApplicantIncome
Enter Loan Amount
LoanAmount
Enter Co-Applicant Income
CoapplicantIncome
Enter Loan Amount term
Loan_Amount_Term
submit


Loan Approval Predcition Form
Fill the Form for Prediction
Gender
Male
Married Status
Yes
Dependents
1
Education
Not Graduate
Self Employed
Yes
Credit_History
1



4.ADVANTAGES & DISADVANTAGES

ADVANTAGES

- Spread the cost of a significant purchase safely
- Can help you manage your personal finances
- Ideal if you have struggled to save in the past
- Unsecured loans are not tied to assets

DISADVANTAGES

- Long-term commitment
- Good product requires a good credit score
- Certain loan types are riskier than others
- Will never get 0% interest - unlike a credit card or finance deal

5. APPLICATION

Dream Housing Finance company deals in all kinds of home loans. They have a presence across all urban, semi-urban and rural areas. The customer first applies for a home loan and after that, the company validates the customer eligibility for the loan.

The company wants to automate the loan eligibility process (real-time) based on customer detail provided while filling out online application forms. These details are Gender, Marital Status, Education, number of Dependents, Income, Loan Amount, Credit History, and others.

To automate this process, they have provided a dataset to identify the customer segments that are eligible for loan amounts so that they can specifically target these customers.

6.CONCLUSION

The analysis starts from data cleaning and processing missing value, exploratory analysis and finally model building and evaluation of the model. The best accuracy on public test set is when we get higher accuracy score and other performance metrics which will be found out. This paper can help to predict the approval of bank loan or not for a candidate.

7.FUTURE SCOPE

Analyzing personal and transaction data gives Banks the opportunity to understand customers' needs today and anticipate future ones. Personalization then adds the ability to deliver those insights to customers in a contextual manner. The most obvious application of these techniques is to increase sales targeting and effectiveness according to a defined business strategy.

A low balance with upcoming bills might call for a personal overdraft offer, a high balance on a current account might suggest appetite for a fixed deposit, recurring visits to the mortgage loan information page might indicate plans to purchase a home, a frequent traveller may be interested in a travel insurance, a fine dining lover might appreciate discounts at a popular restaurant, etc. The opportunities to leverage customer-centric data analytics and personalization for targeted cross-selling or merchants-based campaigns are numerous.

8. APPENDIX

A. Source code

```
import pandas as pd
import numpy as np
import pickle
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import sklearn
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier,RandomForestClassi
fier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import RandomizedSearchCV
import imblearn
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score,classification_report,confusion
_matrix,f1_score


data=pd.read_csv('/content/train_u6lujuX_CVtuZ9i.csv')
data

data.info()

data.isnull().sum()

data['Gender']=data['Gender'].fillna(data['Gender'].mode()[0])

data['Married']=data['Married'].fillna(data['Married'].mode()[0])

data['Dependents']=data['Dependents'].str.replace('+','')

data['Dependents']=data['Dependents'].fillna(data['Dependents'].mode()[0])

data['Self_Employed']=data['Self_Employed'].fillna(data['Self_Employed'].m
ode()[0])

data['LoanAmount']=data['LoanAmount'].fillna(data['LoanAmount'].mode()[0])
```

```
data['Loan_Amount_Term']=data['Loan_Amount_Term'].fillna(data['Loan_Amount
_Term'].mode()[0])

data['Credit_History']=data['Credit_History'].fillna(data['Credit_History'
].mode()[0])

from imblearn.combine import SMOTETomek

smote=SMOTETomek(0,90)

y=data['Loan_Status']
x=data.drop(columns=['Loan_Status'],axis=1)

x_bal,y_bal=smote.fit_resample(x,y)

data.describe()

plt.figure(figsize=(12,5))
plt.subplot(121)
sns.distplot(data['ApplicantIncome'],color='r')
plt.subplot(122)
sns.distplot(data['Credit_History'])
plt.show

plt.figure(figsize=(18,4))
plt.subplot(1,4,1)
sns.countplot(data['Gender'])
plt.subplot(1,4,2)
sns.countplot(data['Education'])
plt.show

plt.figure(figsize=(20,5))
plt.subplot(131)
sns.countplot(data['Married'],hue=data['Gender'])
plt.subplot(132)
sns.countplot(data['Self_Employed'],hue=data['Education'])
plt.subplot(133)
sns.countplot(data['Property_Area'],hue=data['Loan_Amount_Term'])

sns.swarmplot(data['Gender'],data['ApplicantIncome'],hue=data['Loan_Status
'])

x_train,x_test,y_train,y_test=train_test_split(x_bal,y_bal,test_size=0.33,
random_stste=42)
```

```
def DecisionTreeClassifier(x_train,x_test,t_train,y_test):
  dt=DecisionTreeClassifier()
  dt.fit(x_train,y_train)
  yPred=dt.predict(x_test)
  print('***DecisionTreeClassifier***')
  print('confusion matrix')
  print(confusion_matrix(x_test,yPred))
  print('Classification report')
  print(classification_report(y_test,yPred))

def RendomForestClassifier(x_train,x_test,t_train,y_test):
  rf=RandomForestClassifier()
  rf.fit(x_train,y_train)
  yPred=dt.predict(x_test)
  print('***RandomforestClassifier***')
  print('confusion matrix')
  print(confusion_matrix(x_test,yPred))
  print('Classification report')
  print(classification_report(y_test,yPred))

def KNN(x_train,x_test,t_train,y_test):
  knn=KNeighborsClassifier()
  knn.fit(x_train,y_train)
  yPred=dt.predict(x_test)
  print('***KNeighborsClassifier***')
  print('confusion matrix')
  print(confusion_matrix(x_test,yPred))
  print('Classification report')
  print(classification_report(y_test,yPred))

def xgboost(x_train,x_test,t_train,y_test):
  xg=GradientBoostingClassifier()
  xg.fit(x_train,y_train)
  yPred=xg.predict(x_test)
  print('***GradientBoostingClassifier***')
  print('confusion matrix')
  print(confusion_matrix(x_test,yPred))
```

```
  print('Classification report')
  print(classification_report(y_test,yPred))


import tensorflow
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense


classifier=Sequential()
classifier.add(Dense(units=100,activation='relu',input_dim=11))
classifier.add(Dense(units=50,activation='relu'))
classifier.add(Dense(units=1,activation='sigmoid'))
classifier.compile(optimizer='adam',loss='binary_crossentropy',metrics=['a
ccuracy'])
model_history=classifier.fit(x_train,y_train,batch_size=100,validation_spl
it=0.2,epochs=100)
dtr.predict([[1,1,0,1,1,4276,1542,145,240,0,1]])
rfr.predict([[1,1,0,1,1,4276,1542,145,240,0,1]])
knn.predict([[1,1,0,1,1,4276,1542,145,240,0,1]])
xgb.predict([[1,1,0,1,1,4276,1542,145,240,0,1]])
classifier.save("Loan.h5")
y_pred=classifier.predict(x_test)
y_pred
y_pred=(y_pred > 0.5)
y_pred
sample_value=np.array(sample_value)
sample_value=sample_value.reshape(1,-1)
sample_value=sc.transform(sample_value)
return classifier.predict(sample_value)
sample_value=[[1,1,0,1,1,4276,1542,145,240,0,1]]

if predict_exit(sample_value)>0.5
  print('Prediction: High chance of Loan Approval')
else:
  print('Prediction: Low chance Loan Approval')

sample_value=[[1,0,1,1,1,45,14,45,240,1,1]]
if predict_exit(sample_value)>0.5:
   print('Prediction: High chance of Loan Approval')
else:
  print('Prediction: Low chance of Loan Approval')

def compareModel(x_train,x_test,y_train,y_test):
```

```
  decisionTree(x_train,x_test,y_train,y_test)
  print('-'*100)
  RandomForest(x_train,x_test,y_train,y_test)
  print('-'*100)
  XGB(x_train,x_test,y_train,y_test)
  print('-'*100)
  KNN(x_train,x_test,y_train,y_test)
  print('-'*100)

compareModel(x_train,x_test,y_train,y_test)
```

```
yPred=classifier.predict(x_test)
print(accuracy_score(y_pred,y_test))
print("ANN Model")
print("Confusion_Matrix")
print(confusion_matrix(y_test,y_pred))
print("Classification Report")
princt(classification_report(y_test,y_pred))

from sklearn.model_selection import cross_val_score

rf =RandomForestClassifier()
rf.fit(x_train,y_train)
ypred = rf.predict(x_test)

f1_score(ypred,y_test,average = 'weighted')

cv= cross_val_score(rf,x,y,cv=s)

np.mean(cv)

pickle.dump(model,open('rdf.pl1','wb'))

from flask import Flask, render_template,request
import numpy as np
import pickle

app = Flask(_name_)
model = pickle.load(open(r.'rdf.pkl','rb'))
scale = pickle.load(open(r,'scale1.pkl','rb'))
```

```
def home():
  return render_template('home.html')

def submit():
  input_feature=[int(x) for x in request.form.values()]
  input_feature=[np.array(input_feature)]
  print(input_feature)
  names=['Gender','Married','Dependents','Education','self_Employed','Appl
ication','CoapplicantIncome','LoanAmount','Loan_Amount_Term','Credit_Histo
ry','Property_Area']
  data= pandas.DataFrame(input_feature,coumns=names)
  print(data)
  prediction=model.predict(data)
  print(prediction)
  prediction=int(prediction)
  print(type(prediction))
  if(prediction==0):
    return render_template("output.html",result="Loan will Not be Approval
")
  else:
    return render_template("output.html",result="Loan will be Approval")
```

```
if__name__==”__moin__”:
        port=int(os.environ.get(‘PORT’,5000))
        app.run(debug=False)
```