SMART LENDER - APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL USING IBM WATSON

Developed By:

D. Anusha,
J. Teja,
Ch. Vikram,
MD. Feroz Uddin

Smart Bridge – Mini Project Report

INDEX

1.INTRODUCTION

2.LITERATURE SURVEY

3.THEORITICAL ANALYSIS

4.EXPERIMENTAL ANALYSIS

5.FLOWCHART

6.RESULT

7.ADVANTAGES & DISADVANTAGES

8.APPLICATIONS

9.CONCLUSION

10.FUTURE SCOPE

11.BIBILOGRAPHY

12.APPENDIX

1. INTRODUCTION

1.1 Overview

Banks are making major part of profits through loans. Though lot of people are applying for loans. It's hard to select the genuine applicant, who will repay the loan. While doing the process manually, lot of misconception may happen to select the genuine applicant. Therefore we are developing loan prediction system using machine learning, so the system automatically selects the eligible candidates. This is helpful to both bank staff and applicant. The time period for the sanction of loan will be drastically reduced.

We will be using classification algorithms such as Decision tree, Random forest, KNN, and xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

1.2 Purpose

A loan is the core business part of banks. The main portion the bank's profit is directly come from the profit earned from the loans. Though bank approves loan after a regress process of verification and testimonial but still there's no surety whether the chosen hopeful is the right hopeful or not. This process takes fresh time while doing it manually. We can prophesy whether that particular hopeful is safe or not and the whole process of testimonial is automated by machine literacy style. Loan Prognostic is really helpful for retainer of banks as well as for the hopeful also.

2.LITERATRUE SURVEY

2.1 Existing Problem

We start our literature review with more general systematic literature reviews that focus on the application of machine learning in the general field of Banking Risk Management. Since the global financial crisis, risk management in banks has to take a major role in shaping decision-making for banks. A major portion of risk management is the approval of loans to promising candidates. But the black-box nature of Machine learning algorithms makes many loan providers vary the result. Martin Leo, Suneel Sharma and k. Maddulety's [1] extensive report has explored where Machine Learning is being used in the fields of credit risk, market risk, operational risk, and liquidity risk only to conclude that the research falls short of extensive research is required in the field.

Classification and Regression Trees are referred to as CART (in short) introduced by Leo Breiman. It best suits both predictive and decision modeling problems. This Binary Tree methodology is the greedy method is used for the selection of the best splitting. Although Decision trees gave us a similar accuracy.

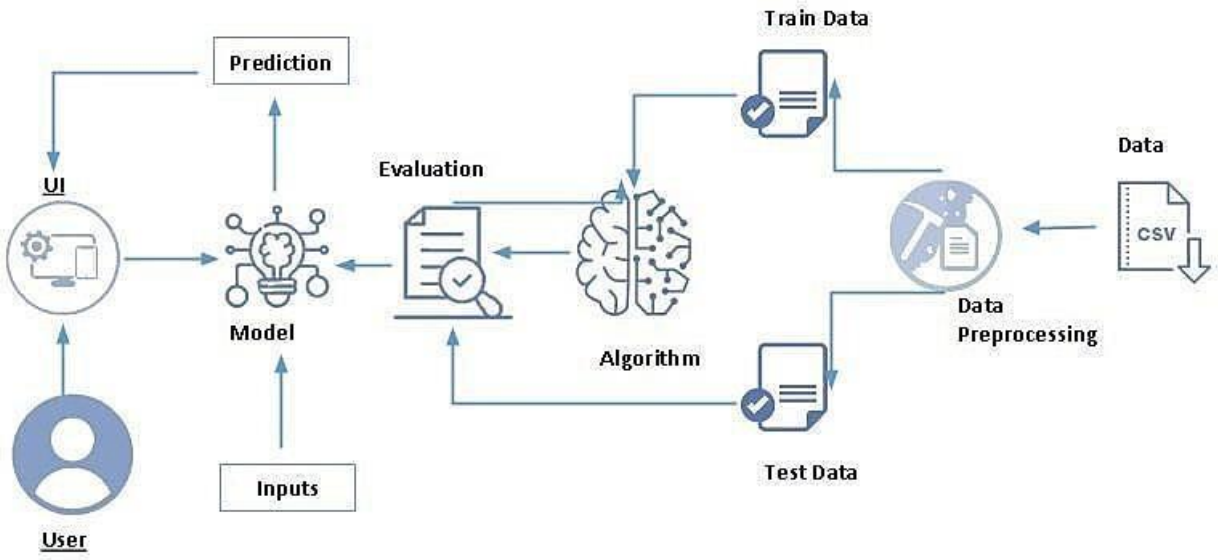
The benefits of Decision Trees, in this case, were due to the latter giving equal importance to both accuracy and prediction. This model became successful in making a lower number of False Predictions to reduce the risk factor.

2.2 Proposed Solution

Data Mining is also becoming popular in the field banking sector as it extracts information from a tremendous amount of accumulated data sets. & nbsp ; Aboobyda Jafar Hamid and Tarig Mohammed Ahmed [10] focused on implementing data mining techniques using three models j48, bayesNet, and naiveBayesdel for classifying loan risk in the banking sector. The author implemented and tested models using the Weka application. In their work, they made a comparison between these algorithms in terms of accuracy in classifying the data correctly. The operation of sprinting happened in a manner that 80% represented the training dataset and 20% represented the testing dataset. After analyzing the results the author came up with the results that the best algorithm among the three is the J48w algorithm in terms of high accuracy and low mean absolute error.

3. THEORITICAL ANALYSIS

3.1 Block Diagram



Project workflow:

1. Install required packages and libraries
2. Understanding the data.
3. Model Building
4. Application Building
5. Final UI

3.2 Hardware/Software Designing

For running a machine learning model on the system you need a system with minimum of 16 GB RAM in it and you require a good processor for high performance of the model. In the list of **software requirements** you must have:

1. Jupyter Notebook for programming, which can be installed by Anaconda IDE.
2. Python packages
3. A better software for running the html and css files for application building phase e.g. spyder.

4. EXPERIMENTAL INVESTIGATIONS

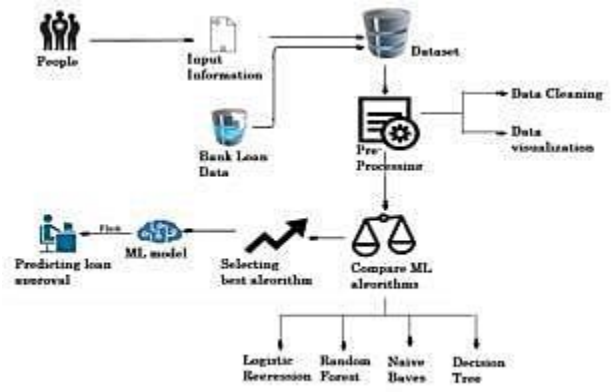
Data preprocessing

Data Analysis of Visualization

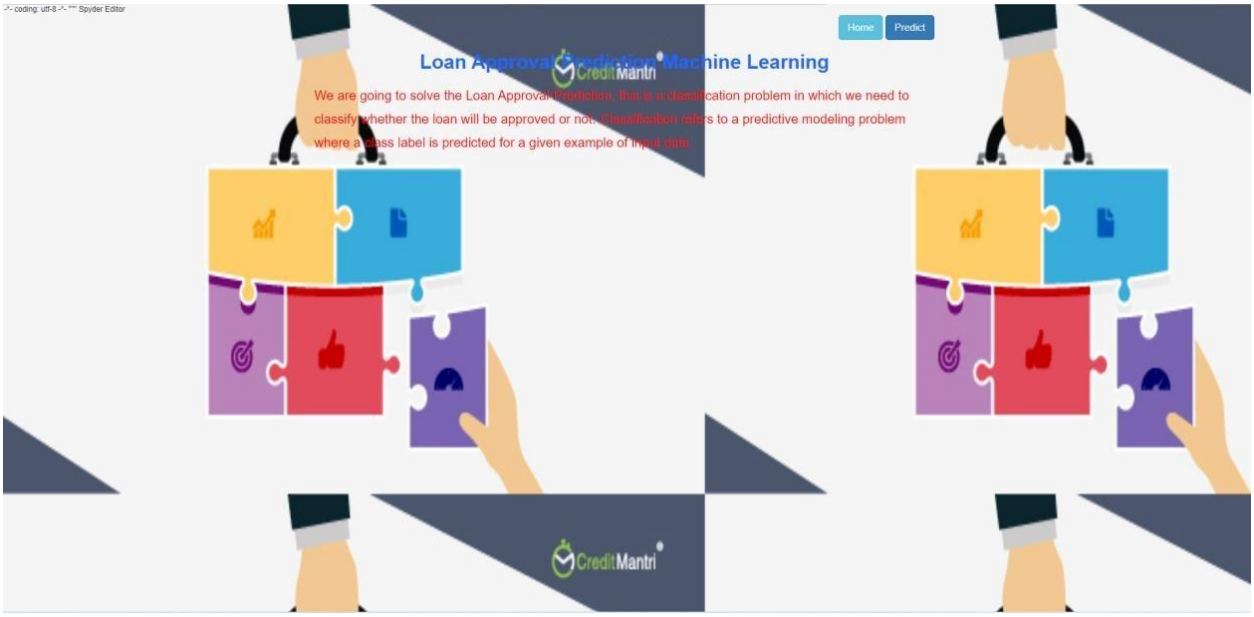
Comparing Algorithms

Deployment Using Flask

5. FLOWCHART



6.RESULT



-*- coding: utf-8 -*- """ Created on Thu Nov 10 20:08:16 2022 @author: DIVYA """

7. ADVANTAGES

1.Performance and accuracy of the algorithms can be calculated and compared.

2. Class imbalance can be dealt with machine learning approaches.

DISADVANTAGES

1. They had proposed a mathematical model and machine learning algorithms were not used.
2. Class Imbalance problem was not addressed and the proper measure were not taken.

8. APPLICATIONS

We are developing loan prediction system using machine learning, so the system automatically selects the eligible candidates. This is helpful to both bank staff and applicant. The time period for the sanction of loan will be drastically reduced. The dataset is obtained by gathering lot of required datasets and combining them to produce a generalised dataset. The dataset thus produced is pre-processed i.e., the dataset is cleaned before doing data visualization. Then the four algorithms are applied on the same pre-processed dataset and calculated for the best performed algorithms among them. Then the best algorithm is used to train the model and test it to check how accurate the algorithm can predict the output. Then we deploy that model to predict if bank loan can be approved or not for a specific candidate.

9. CONCLUSION

For the purpose of predicting the loan approval status of the applied customer, we have chosen the machine learning approach to study the bank dataset. We have applied various machine learning algorithms to decide which one will be the best for applying on the dataset to get the result with the highest accuracy. We also determined the most important features that influence the loan approval status. These most important features are then used on some selected algorithms and their performance accuracy is compared with the instance of using all the features. This model can help the banks in figuring out which factors are important for the loan approval procedure. The comparative study makes us clear about which algorithm will be the best and ignores the rest, based on their accuracy.

10. FUTURE SCOPE

We can make the Bank Loan Approval prediction to connect with Cloud for future use to optimize the work to implement in Artificial Intelligence environment. In future, this model can be used to compare various machine learning algorithm generated prediction models and the model which will give higher accuracy will be chosen as the prediction model. This paper work can be extended to higher level in future. Predictive model for loans that uses machine learning algorithms, where the results from each graph of the paper can be taken as individual criteria for the machine learning algorithm.

11. BIBILOGRAPHY

1. Arun Kumar, Ishan Garg, and Sanmeer Kaur, "Loan Approval Prediction Using Machine Learning Approach," 2018.
2. K. Hanumantha Rao, G. Srinivas, A. Damodhar, and M. Vikas Krishna at International Journal of Computer Science and Telecommunications published an article titled "Implementation of Anomaly Detection Technique Using Machine Learning Algorithms" (Volume2, Issue3, June 2011).
3. G. Arutjothi and C. Senthamarai, "Prediction of loan status in commercial banks using machine learning classifier," International Conference on Intelligent Sustainable Systems (ICISS), 2017.

APPENDIX

```
#!/usr/bin/env python
# coding: utf-8
  import numpy as np import pandas as pd
import matplotlib.pyplot as plt
get_ipython().run_line_magic('matplotlib', 'inline') import seaborn as sns
from sklearn.model_selection import train_test_split
```

```
from sklearn.metrics import classification_report, confusion_matrix
  import os, types import
 pandas as pd

def __iter__(self): return 0

data = pd.read_csv (r'C:\Users\DIVYA\Desktop\mini
 project\dataset\train_dataset.csv.csv') data.head()
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'])

data.describe()


data.info()

data.isnull().sum()

data['Gender']=data['Gender'].map({'Female':1,'Male':0})
data['Property_Area']=data['Property_Area'].map({'Urban':2,'SemiUrban':1,'R ural':0})
data['Married']=data['Married'].map({'Yes':1, 'No':0})
data['Education']=data['Education'].map({'Graduate':1,'Not Graduate':0})
 data['Loan_Status']=data['Loan_Status'].map({'Y':1,'N':0})

data['Self_Employed']=data['Self_Employed'].map({'Yes':1, 'No':0})

data.head(5)
  data['Gender']=data['Gender'].fillna(0).astype('int64')
 data['Married']=data['Married'].fillna(0).astype('int64')
 data['Dependents']=data['Dependents'].astype('int64')
data['Self_Employed']=data['Self_Employed'].fillna(0).astype('int64')
 data['CoapplicantIncome']=data['CoapplicantIncome'].fillna(0).astype('int64
')
```

```
data['LoanAmount']=data['LoanAmount'].fillna(0).astype('int64')
data['Loan_Amount_Term']=data['Loan_Amount_Term'].fillna(0).astype('int64')
data['Credit_History']=data['Credit_History'].fillna(0).astype('int64')

from imblearn.combine import SMOTETomek

smote = SMOTETomek(0.90)

#  dividing the dataset into dependent and independent x and y respectively y = data['Loan_Status']
x = data.drop(columns=['Loan_Status'],axis=1)

x.shape

y.shape
pip install SMOTE
 from imblearn.over_sampling import SMOTE

sm = SMOTE(random_state = 42)
 from sklearn.ensemble import HistGradientBoostingClassifier
from sklearn.datasets import load_iris X, y =
load_iris(return_X_y=True)
clf = HistGradientBoostingClassifier().fit(X, y) clf.score(X, y)
1.0
from imblearn.pipeline import Pipeline as imbpipeline from sklearn.pipeline import
Pipeline

import numpy as np
from sklearn.impute import SimpleImputer
imp = SimpleImputer(missing_values=np.nan, strategy='mean')

oversample = SMOTE()



# #creating a new x and y variables for the balanced set
# x_bal,y_bal = oversample.fit_resample(x,y)

# #printing the values of y before balancing the data and after
# print(y.value_counts())
# print(y_bal.value_counts())
```

home.html

```
<!doctype html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initialscale=1">
```

```
  <meta http-equiv="X-UA-Compatible" content="ie=edge">
  <title>Home</title>    <link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min .css">    <style>       body
{       background-image: url("https://s3.ap-south-
1.amazonaws.com/img1.creditmantri.com/community/article/5-importanttips-for-a-successful-
business-loan-application.jpg");
              background-size: 1300px 900px;
             }       h3.big       {       line-
height: 1.8;
      }
  </style>
</head>
<body>
  <br>
  <div class="container">

      <div class="row">
        <div class="col-md-12 bg-light text-right">
          <a href="/home" class="btn btn-info btn-lg">Home</a>           <a href="/predict" class="btn
btn-primary btnlg">Predict</a>

        </div>
      </div>

      <center>

        <font color="#1C67FF"><h1><strong>Loan Approval Prediction
Machine Learning</strong></h1></font>

      </center>
      <h3 class="big">

       <p><font color = "red" font-family: "Arial" style="text-align: center;"> We are going to solve the
Loan Approval Prediction, this is a classification problem in which we need to classify whether the loan
will be approved or not. Classification refers to a predictive modeling problem where a class label is
predicted for a given example of input data.</font></p>

      </h3><br>

  </div>

  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js">
</script>    <script
src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.j s"></script>
```

```
</body>
</html>
<p style="color:#FFFFFF">
```

input.html

```
<html lang="en">
<head>
  <meta charset="UTF-8">
  <title>Predict</title>  <link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min .css">  <style>  body
{
    background-image: url("https://thumbs.dreamstime.com/b/loanapproval-businessman-financial-
standing-application-form-lenderborrower-help-investment-bank-estate-concept-157237960.jpg");
        background-size: 1300px 900px;
    }  h3.big  {  line-
height: 1.8;
    }
  </style>
</head>
<body>
<font color="#1C67FF"><h1><strong>Enter the details for Loan
Approval</strong></h1><br></font>
<div class="container">

    <form action="submit", method="post">


      <div>
      <label for="Gender" style="font-size:
20px">Gender</label><br>

      <select name="Gender" id="Gender">
        <option value="0">Male</option>
        <option value="1">Female</option>
       </select></div></br>

     <div>
     <label  for="Married" style="font-size:
20px">Married</label><br>
      <select  name="Married" id="Married">
        <option value="0">No</option>
        <option value="1">Yes</option>
       </select></div></br>
```

```
          <div><label  for="Dependents" style="font-size:
20px">Dependents</label><br>
          <input type="number" id="Dependents" min=0 max=10 name="Dependents" placeholder="No
of Dependents on you....."></div></br>


          <div>
          <label  for="Education" style="font-size:
20px">Education</label><br>
          <select  name="Education" id="Education">
            <option value="0">Not Graduate</option>
            <option value="1">Graduate</option>
           </select></div></br>

         <div>
         <label  for="Self_Employed" style="font-size:
20px">Self_Employed</label><br>
         <select  name="Self_Employed" id="Self_Employed">
            <option value="0">No</option>
            <option value="1">Yes</option>
          </select></div></br>

          <div><label  for="ApplicantIncome" style="font-size:
20px">ApplicantIncome</label><br>
          <input type="number" id="ApplicantIncome" min=1000  name="ApplicantIncome"
placeholder="Your Income....."></div></br>                    <div><label
for="CoapplicantIncome" style="font-size:
20px">CoapplicantIncome</label><br>
          <input type="number" id="CoapplicantIncome" min=100  name="CoapplicantIncome"
placeholder="Your Income....."></div></br>

          <div><label  for="LoanAmount" style="font-size:
20px">LoanAmount</label><br>
         <input type="number" id="LoanAmount" min=0  name="LoanAmount" placeholder="Enter the
                                  Loan Amount....."></div></br>

          <div><label  for="Loan_Amount_Term" style="font-size:
20px">Loan_Amount_Term</label><br>
          <input type="number" id="Loan_Amount_Term" min=30 max=1500
name="Loan_Amount_Term" placeholder="Enter the Loan Amount
Term in Days....."></div></br>

          <div><label  for="Credit_History" style="font-size:
```

```
20px">Credit_History</label><br>
          <input type="number" id="Credit_History" min=0 max=5  name="Credit_History"
placeholder="Enter your previous Credit
History.....></div></br>

        <div><label  for="Property_Area" style="font-size:
20px">Property_Area</label><br>
        <select  name="Property_Area" id="Property_Area">
          <option value="0">Rural</option>
          <option value="1">Semiurban</option>
          <option value="2">Urban</option>
         </select></div></br>




       <button type="submit" style="font-size:
15px">Submit</button>


    </form>

</div>



</body>
</html>
```

submit.html

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <title>Output</title>   <link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min .css">   <style>       body {
     background-image: url("https://encrypted-
tbn0.gstatic.com/images?q=tbn:ANd9GcS___7W3sK8ZEqTU3HIHvVGuCRne8lZAJEa5McqCU44EATLC1pP
ZZrVyvGUtu4F1RlNyI&usqp=CAU");
         background-size: 1300px 900px;
     }       h3.big       {
     line-height: 1.8;
     }
  </style>
```

```
</head>
<body>
  <br>
    <div class="container">

      <div class="row">
        <div class="col-md-12 bg-light text-right">          <a href="/home" class="btn
btn-info btnlg">Home</a>
          <a href="/predict" class="btn btn-primary btnlg">Predict</a>

        </div>
      </div>
      <br>
      <font color="#1C67FF"><h1><strong>Prediction for Loan
Approval</strong></h1></font><br>
      <p><font font-family: "Arial" color = "#C500E0"><h3>      Based on the given input, it
predicts LOAN APPROVAL is {{prediction_text}}.
      </h3>
      </font></p>

    </div>

</body>
</html>
```