

# **CREDIT CARD APPROVAL PREDICTION USING IBM WATSON MACHINE LEARNING**

## **1. INTRODUCTION:**

### **1.1 Overview:**

With the increasing number of credit card applications, banks are opting towards the use of prediction-based algorithms as opposed to manual approval methods. Data analysis has exhibited a strong correlation between several financial and personal factors of a client and the likelihood of said client complying with their respective bank's credit policies. In this paper, we propose the use of the Machine Learning algorithm to predict and grant credit cards to applicants based on the customers' activity history. We used some financial and personal factors. We predicted the resulting factors through the use of Machine Learning algorithm with an emphasis on error minimization. Using this Machine Learning model, the machine-learned which of these applicants are most likely to accumulate bad debts and granted or rejected the applications based on the prediction.

### **1.2 Purpose:**

Now a days every person needs a credit card but banks not provide credit card to everyone. Before giving the credit card the bank employee's needs to identify whether the person is fraud or good. But we are not able to analyse any one by seeing their face or look. For that we need to check his personal details like - Income, Education, Family, etc., But it is not possible to check manually now a days because population is very high and suppose if we are able to check then no one has that much of time. For time saving and growing the business we are using machine learning model. We are training the model based on historic data to check these details and after checking the details this will gives results at that time. So Machine Learning models save our time, money, energy etc.

## **2 LITERATURE SURVEY:**

### **2.1 Existing problem:**

Some of existing solution for solving this problem are:

Credit risk as the board in banks basically centers around deciding the probability of a customer's default or credit decay and how expensive it will end up being assuming it happens. It is important to consider major factors and predict beforehand the probability of consumers defaulting given their conditions. Which is where a machine learning model comes in handy and allows the banks and major financial institutions to predict whether the customer, they are giving the loan to, will default or not. This project builds a machine learning model with the best accuracy possible using python. First we load and view the dataset. The dataset has a combination of both mathematical and non-mathematical elements, that it contains values from various reaches, in addition to that it contains a few missing passages. We preprocess the dataset to guarantee the AI model we pick can make great expectations. After the information is looking great, some exploratory information examination is done to assemble our instincts. Finally, we will build a machine learning model that can predict if an individual's application for a credit card will be accepted.

-

### **2.2 Proposed solution:**

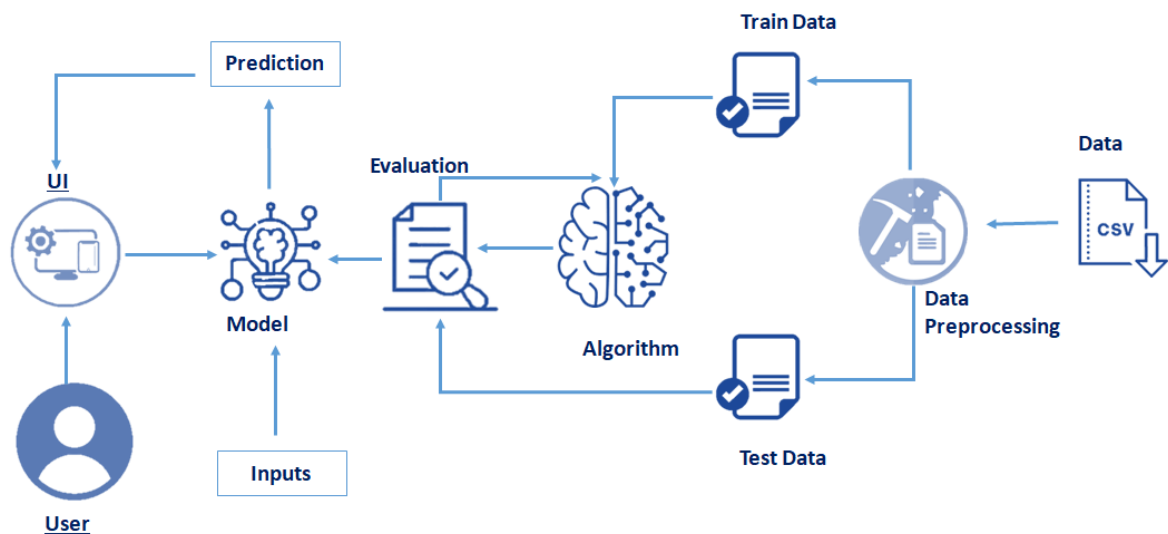
We are proposed the method credit card approval prediction using IBM Watson by machine learning:

Here we can predict whether the person will eligible for the credit card or not. For this we can use the machine learning algorithm to train, test and implementation with the help of datasets. We can use the user interface for the user interactions.

The IBM Watson is an cloud service by using this we can service the new developer to access our data. How are done this application.

### 3. THEORITICAL ANALYSIS:

#### 3.1 Block Diagram:



#### 3.2 Hardware/software designing:

##### Hardware Requirements:

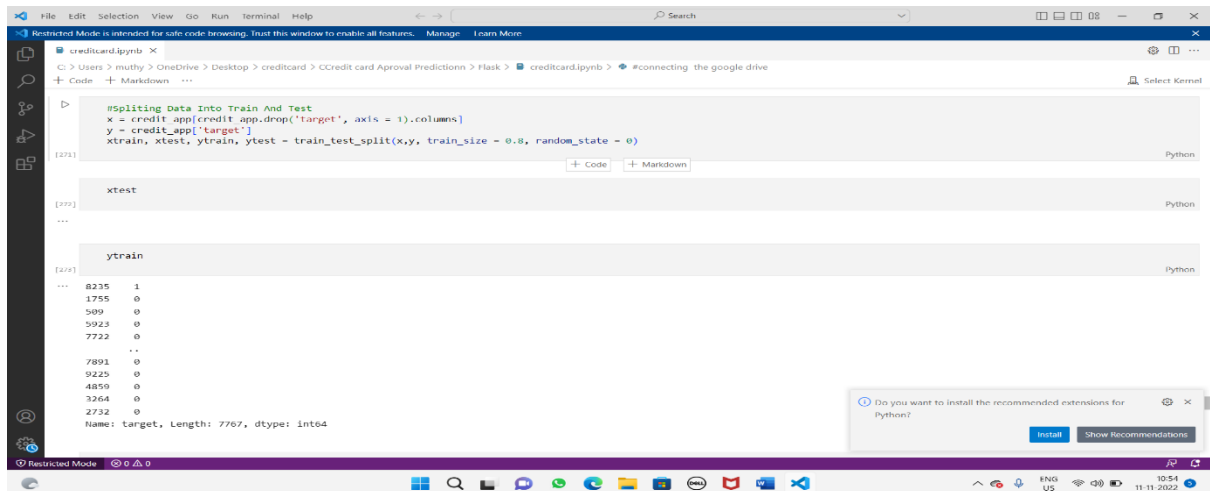
Operating system	Windows, Mac, Linux
CPU (for training)	Multi Core Processors (i3 or above/ equivalent)
GPU (for training)	NVIDIA AI Capable / Google's TPU

##### Software Requirements:

Python	V3.10.0 or Above
Python Packages	Flask, tensorflow, opencv-python, keras, nump , Pandas, virtualenv, pillow
Web Browser	Mozilla Firefox, Google Chorme or any modern web browser
IBM Cloud (for training)	Watson Studio-Model Training & Deployment as Machine Learning Instance

## 4. EXPERIMENTAL INVESTIGATIONS:

### 4.1 Training the train dataset:



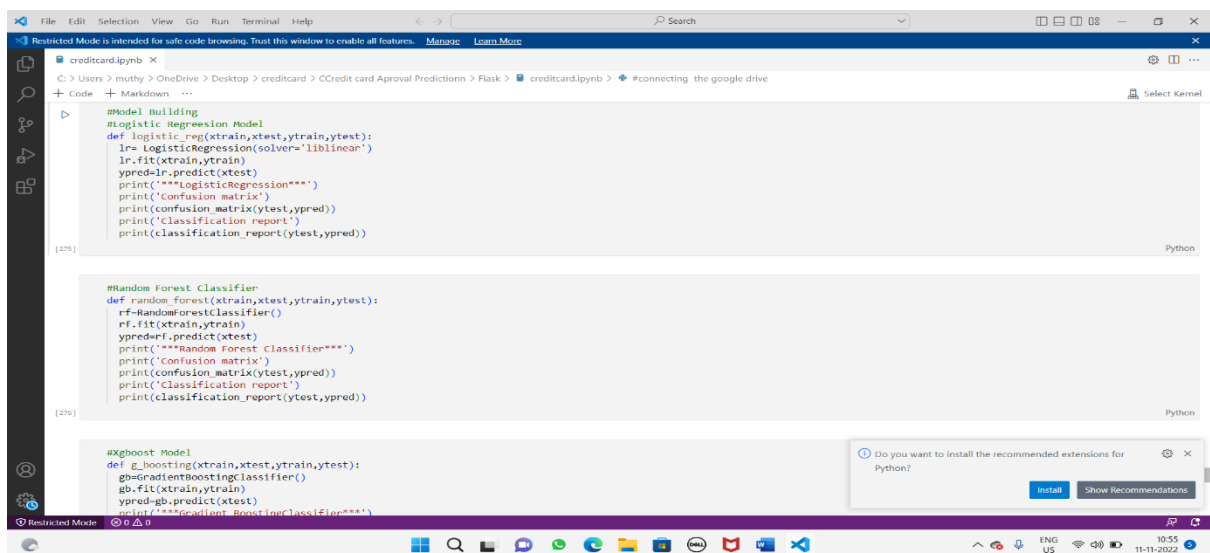
This screenshot shows the first two cells of a Jupyter Notebook. The first cell contains code to split the 'creditcard' dataset into training and testing sets using `train_test_split`. The second cell displays the resulting `xtest` and `ytrain` arrays, showing a preview of the data and its dimensions.

```
#splitting Data into Train And Test
x = credit_app[credit_app.drop('target', axis = 1).columns]
y = credit_app['target']
xtrain, xtest, ytrain, ytest = train_test_split(x,y, train_size = 0.8, random_state = 0)
```

```
xtest
```

```
ytrain
```

```
8235 1
1755 0
509 0
5923 0
7722 0
..
7891 0
9225 0
4859 0
3264 0
2732 0
Name: target, Length: 7767, dtype: int64
```

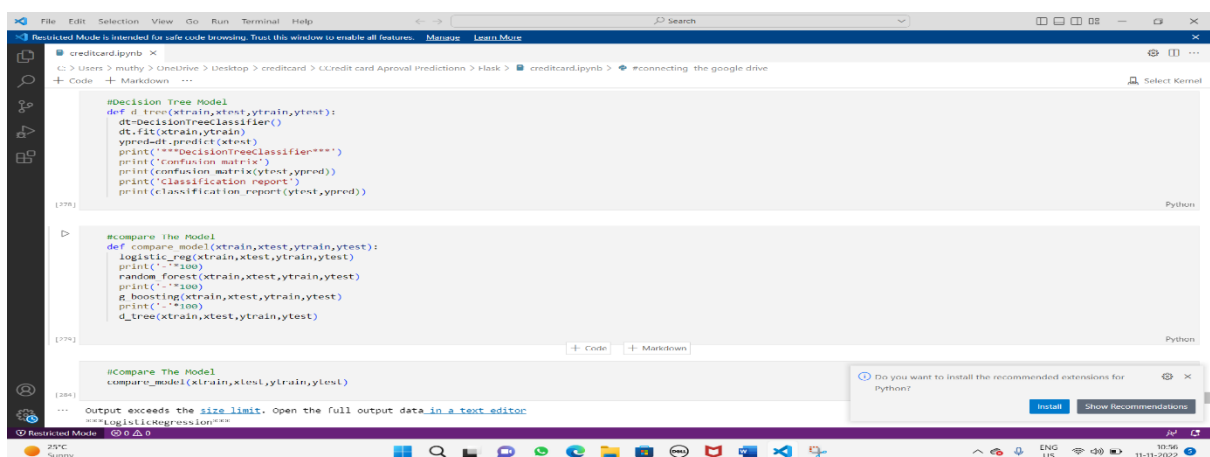


This screenshot shows the next three cells of the Jupyter Notebook. The third cell defines a `logistic_reg` function for training and testing a Logistic Regression model. The fourth cell defines a `random_forest` function for training and testing a Random Forest Classifier. The fifth cell defines a `g_boosting` function for training and testing an XGBoost model.

```
#Model Building
#Logistic Regression Model
def logistic_reg(xtrain,xtest,ytrain,ytest):
    lr= LogisticRegression(solver='liblinear')
    lr.fit(xtrain,ytrain)
    ypred=lr.predict(xtest)
    print('***LogisticRegression***')
    print('Confusion matrix')
    print(confusion_matrix(ytest,ypred))
    print('Classification report')
    print(classification_report(ytest,ypred))
```

```
#Random Forest Classifier
def random_forest(xtrain,xtest,ytrain,ytest):
    rf=RandomForestClassifier()
    rf.fit(xtrain,ytrain)
    ypred=rf.predict(xtest)
    print('***Random Forest Classifier***')
    print('Confusion matrix')
    print(confusion_matrix(ytest,ypred))
    print('Classification report')
    print(classification_report(ytest,ypred))
```

```
#Xgboost Model
def g_boosting(xtrain,xtest,ytrain,ytest):
    gb=GradientBoostingClassifier()
    gb.fit(xtrain,ytrain)
    ypred=gb.predict(xtest)
    print('***Gradient Boosting Classifier***')
```



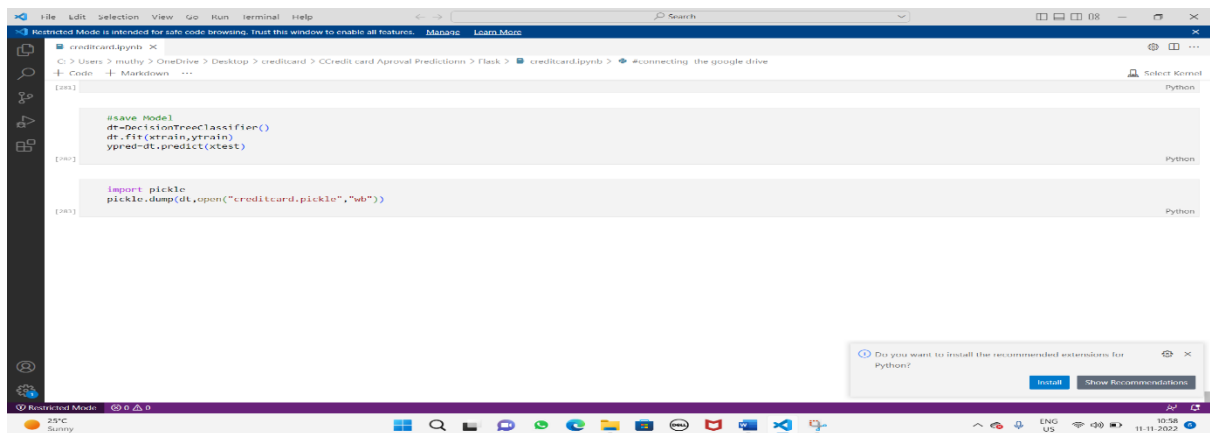
This screenshot shows the final two cells of the Jupyter Notebook. The sixth cell defines a `dt` function for training and testing a Decision Tree Classifier. The seventh cell defines a `compare_model` function that trains and tests all three models (Logistic Regression, Random Forest, and XGBoost) and prints their performance metrics.

```
#Decision Tree Model
def dt(xtrain,xtest,ytrain,ytest):
    dt=DecisionTreeClassifier()
    dt.fit(xtrain,ytrain)
    ypred=dt.predict(xtest)
    print('***DecisionTreeClassifier***')
    print('Confusion matrix')
    print(confusion_matrix(ytest,ypred))
    print('Classification report')
    print(classification_report(ytest,ypred))
```

```
#compare the Model
def compare_model(xtrain,xtest,ytrain,ytest):
    logistic_reg(xtrain,xtest,ytrain,ytest)
    print('***100')
    random_forest(xtrain,xtest,ytrain,ytest)
    print('***100')
    g_boosting(xtrain,xtest,ytrain,ytest)
    print('***100')
    dtree(xtrain,xtest,ytrain,ytest)
```

```
#Compare The Model
compare_model(xtrain,xtest,ytrain,ytest)
```

Output exceeds the size limit. open the full output data in a text editor



The screenshot shows a Jupyter Notebook window with the following code:

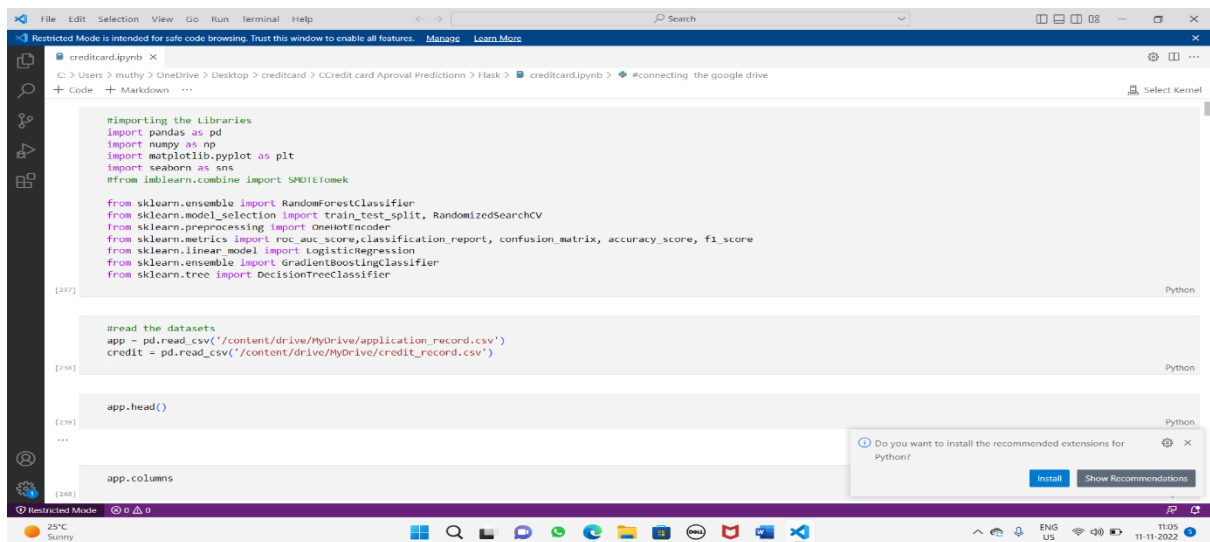
```
[265]: #save Model
dt=DecisionTreeClassifier()
dt.fit(xtrain,ytrain)
ypred=dt.predict(xtest)

[267]:

[268]: import pickle
pickle.dump(dt,open("creditcard.pickle","wb"))
```

A notification box at the bottom right asks: "Do you want to install the recommended extensions for Python?" with buttons for "Install" and "Show Recommendations".

## Testing the test dataset:



The screenshot shows a Jupyter Notebook window with the following code:

```
[267]: #importing the Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#from imblearn.combine import SMOTETomek

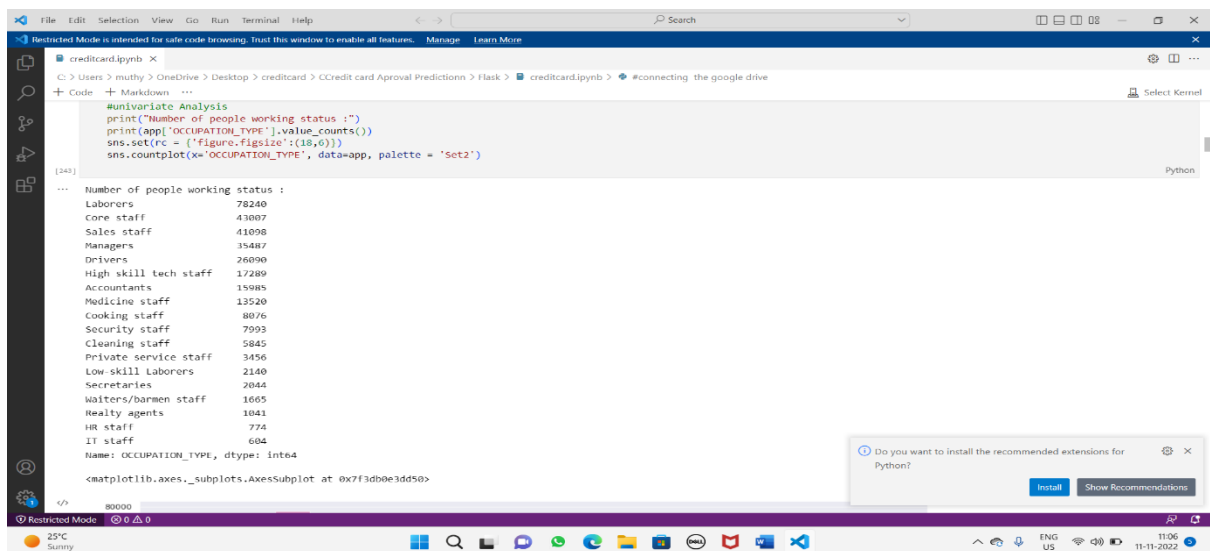
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split, RandomizedSearchCV
from sklearn.preprocessing import OneHotEncoder
from sklearn.metrics import roc_auc_score, classification_report, confusion_matrix, accuracy_score, f1_score
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier

[268]: #read the datasets
app = pd.read_csv('/content/drive/MyDrive/application_record.csv')
credit = pd.read_csv('/content/drive/MyDrive/credit_record.csv')

[269]: app.head()

[270]: app.columns
```

A notification box at the bottom right asks: "Do you want to install the recommended extensions for Python?" with buttons for "Install" and "Show Recommendations".



The screenshot shows a Jupyter Notebook window with the following code:

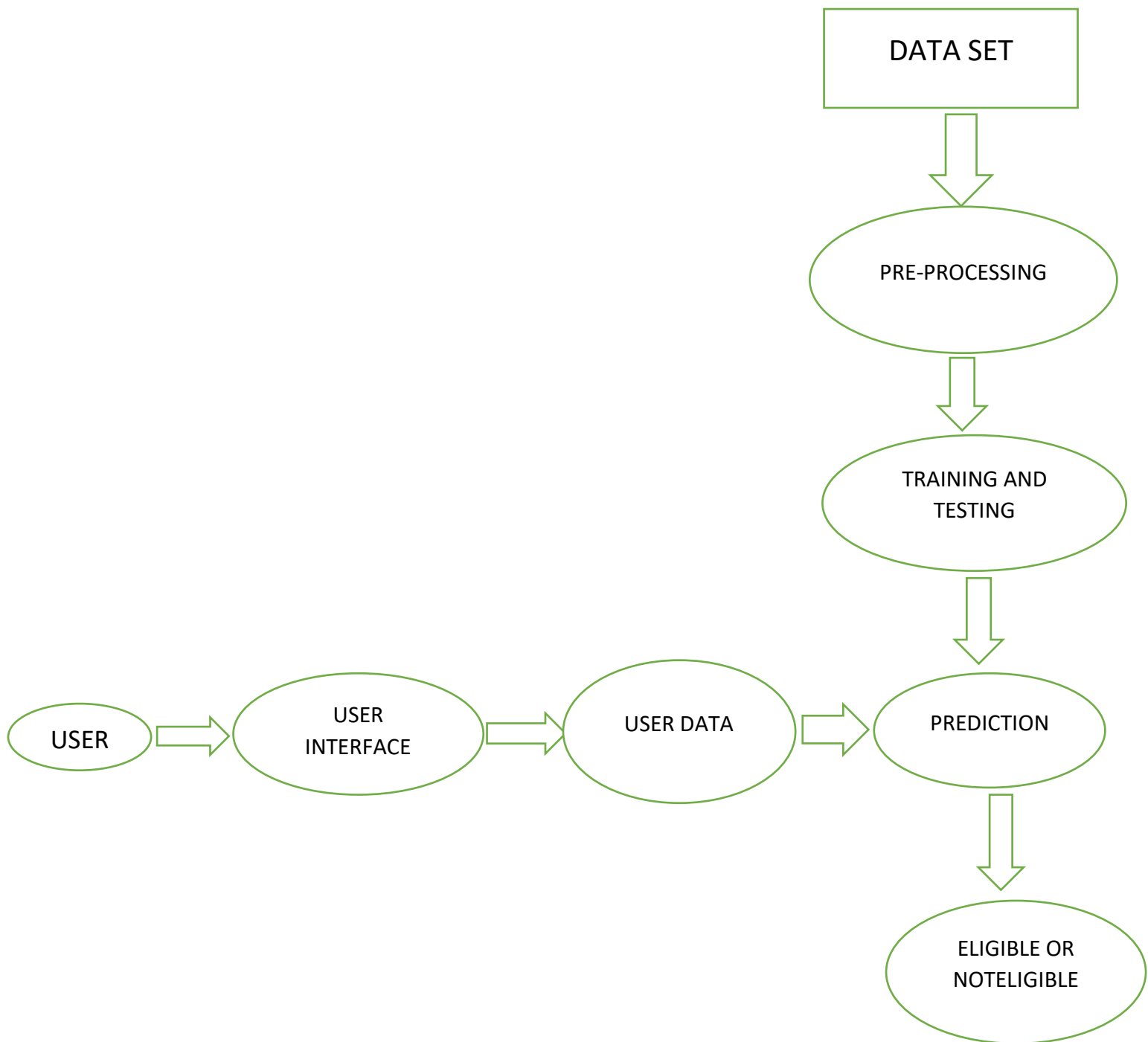
```
[242]: #univariate Analysis
print("Number of people working status :")
print(app['OCCUPATION_TYPE'].value_counts())
sns.set(rc = {'figure.figsize':(18,6)})
sns.countplot(x='OCCUPATION_TYPE', data=app, palette = 'Set2')
```

The output of the code is a bar chart showing the number of people working status for each occupation type. The x-axis is labeled 'OCCUPATION\_TYPE' and the y-axis is labeled 'count'. The bars are colored according to the 'Set2' palette.

OCCUPATION_TYPE	count
Laborers	78240
Core staff	43007
Sales staff	41098
Managers	35487
Drivers	26090
High skill tech staff	17289
Accountants	15985
Medicine staff	13520
Cooking staff	8076
Security staff	7993
Cleaning staff	5845
Private service staff	3456
Low-skill Laborers	2140
Secretaries	2044
Walters/barmen staff	1665
Health agents	1041
HR staff	774
IT staff	604

A notification box at the bottom right asks: "Do you want to install the recommended extensions for Python?" with buttons for "Install" and "Show Recommendations".

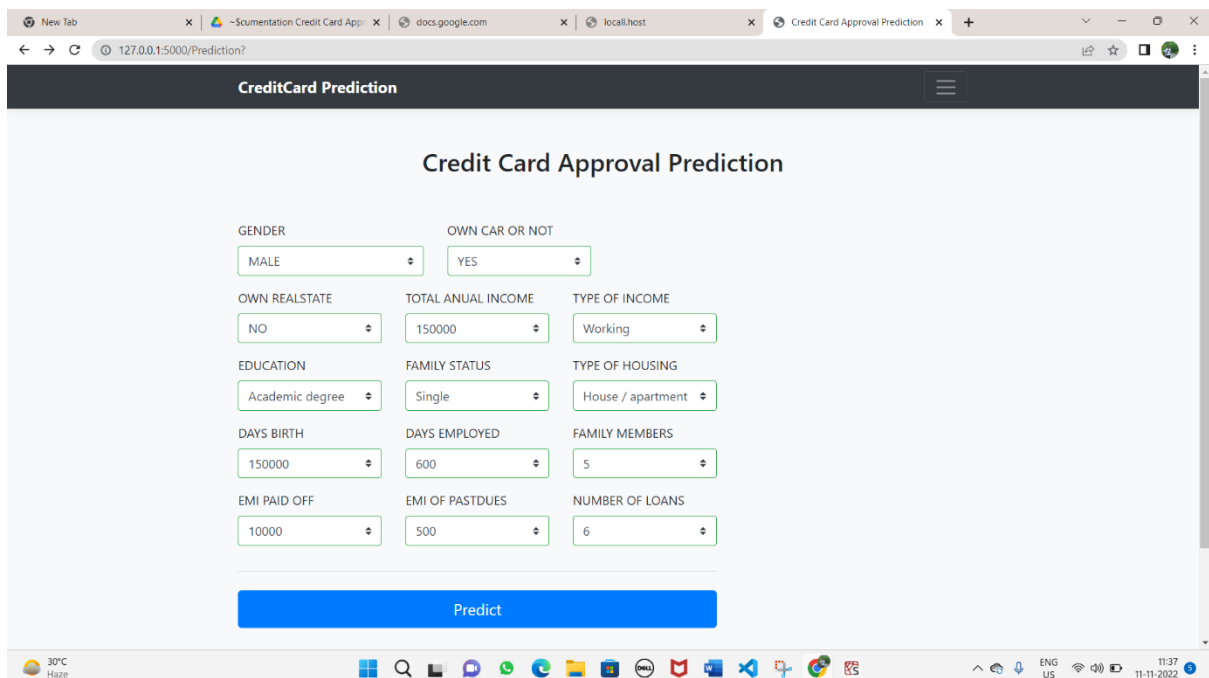
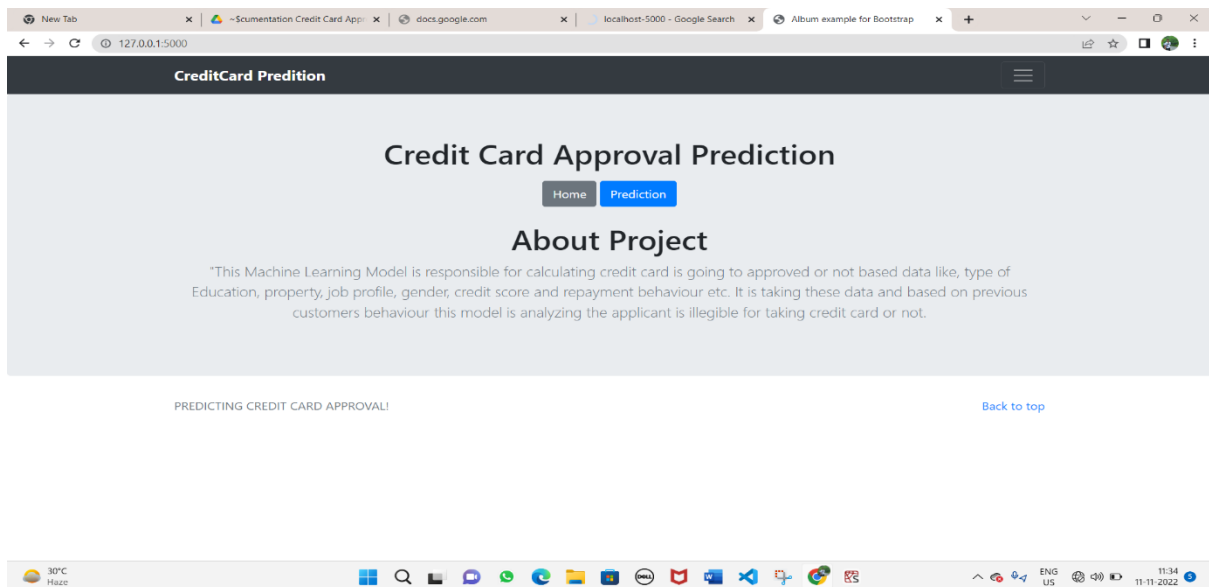
## 5 FLOWCHART:

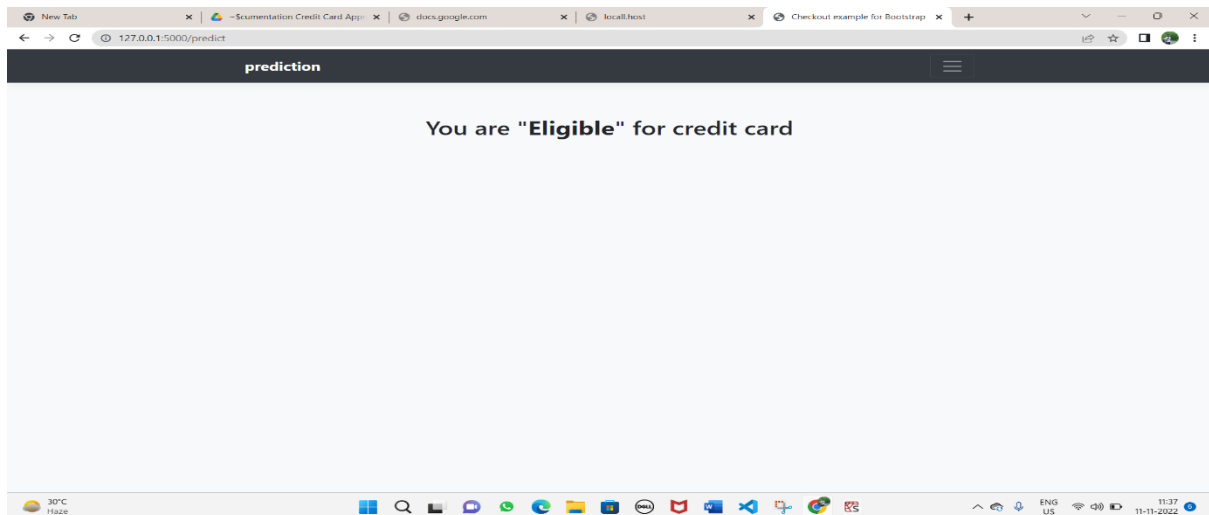


## 6 RESULT:

The proposed procedure was implemented and tested with set of Databases. The set of data is Read, pre-process, model building, training, testing and executed. The result is present in below

The output of Credit Card Prediction is provided below:





## 7 ADVANTAGES AND DISADVANTAGES:

### Advantages:

1. We can save the applicant time to whether the credit card approval is ACCEPTED OR REJECTED by using our UI.
2. Here the prediction is done Automatically without human errors.
3. The applicant will learn how much salary, job, loans get the credit card by using this application.
4. The applicant time is consumed.

### Disadvantage:

- 1 We can use the user interface with the internet connection.
- 2 The single error in data set can change the entire data.
- 3 Minimum due trap
- 4 Easy to overuse
- 5 High interest rate



## 8 APPLICATIONS:

- Authorized user for the Credit card.
- Remove stress button to bank employees.

## 9 CONCLUSION:

In this project, we will be using regression algorithms such as Decision tree, Random forest, KNN, and XGBoost. We will train and test the data with these algorithms. From this the best model is selected and saved in pickle format. We will be doing flask integration and IBM deployment.

This feature can predict whether the applicant will eligible for the credit card or not. By the data which is given by the applicant.

## 10 FUTURE SCOPE:

- In the future more and more customers will join with the banking industry, so immense amount for the will be generated handling which can be quiet an impossible task.
- So, in order to tackle with this situation this task can be automated with the power of machine learning which pretty much every bank does so nowadays.
- In the future this application can build as an app which can be present in the play store and every person will used it with the free of cost.
- We can extend this application with how much the limit of the card.

## 11 BIBILOGRAPHY:

- 1 Kaggle data set: <https://www.kaggle.com/namphuengauawatcharo/credit-card-approval-prediction/data>
- 2 Split data into train and test: <https://www.geeksforgeeks.org/how-to-split-a-dataset-into-train-and-test-sets-using-python/>
- 3 Register for IBM Cloud: <https://www.ibm.com/academic/home>
- 4 Login to IBM Cloud: <https://cloud.ibm.com/login>
- 5 The Train Model on IBM Cloud: <https://youtu.be/TysuP3KgSzc>