

Micro credit-defaulter model

Submitted by:

SHAIK ALTHAFULLA

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

References: Micro finance Institution Indonesia

Github.com

Towardsscience.com

Scikit-learn.org

Kaggle.com, etc

Research papers: The papers can be studied kaggle and github studies papers and links.

Professionals and other resources: Data trained analytics partnerand FLIPROBO Data analyst company.

Data source: Micro finance institution provide Telecom industry report

Through company is using to evaluate this project FLIPROBO Technologies.

**INTRODUCTION**

* Business Problem Framing

A micro finance institution that provides offer financial services to low income people .A Telecommunication provides the products at Lower Prices to all value conscious customers through a strategy of disruptive innovation that focuses on the subscriber. The provide low recharge amounts like 5 rupees and 10 rupees they will be payback with in time or not according to their peoples income levels of customers.

* Conceptual Background of the Domain Problem

The micro finance institution to provide loans of recharges of telecommunication they are payable or not with in time . They are collaborating with an MFI to provide micro-credit on mobile balances to be paid back in 5 days. The Consumer is believed to be defaulter if he deviates from the path of paying back the loaned amount within the time duration of 5 days. For the loan amount of 5 (in Indonesian Rupiah), payback amount should be 6 (in Indonesian Rupiah), while, for the loan amount of 10 (in Indonesian Rupiah), the payback amount should be 12 (in Indonesian Rupiah).

* Review of Literature

Many microfinance institutions (MFI), experts and donors are supporting the idea of using mobile financial services (MFS) which they feel are more convenient and efficient, and cost saving, than the traditional high-touch model used since long for the purpose of delivering microfinance services. Though, the MFI industry is primarily focusing on low income families and are very useful in such areas, the implementation of MFS has been uneven with both significant challenges and successes.

The research can be done all elements of variables what dataset include the with respect to label and target content. with complete analysis can be done.

* Motivation for the Problem Undertaken

They understand the importance of communication and how it affects a person’s life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

The mathematical, statistical and analytics modelling done during this project along with the proper justification.

Using some of libraries in python pandas, Numpy, seaborn and matplotlib and some metrics used

* Data Sources and their formats

<class 'pandas.core.frame.DataFrame'>

Int64Index: 209593 entries, 1 to 209593

Data columns (total 36 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

1. label 209593 non-null int64

1 msisdn 209593 non-null object

2 aon 209593 non-null float64

3 daily\_decr30 209593 non-null float64

4 daily\_decr90 209593 non-null float64

5 rental30 209593 non-null float64

6 rental90 209593 non-null float64

7 last\_rech\_date\_ma 209593 non-null float64

8 last\_rech\_date\_da 209593 non-null float64

9 last\_rech\_amt\_ma 209593 non-null int64

10 cnt\_ma\_rech30 209593 non-null int64

11 fr\_ma\_rech30 209593 non-null float64

12 sumamnt\_ma\_rech30 209593 non-null float64

13 medianamnt\_ma\_rech30 209593 non-null float64

14 medianmarechprebal30 209593 non-null float64

15 cnt\_ma\_rech90 209593 non-null int64

16 fr\_ma\_rech90 209593 non-null int64

17 sumamnt\_ma\_rech90 209593 non-null int64

18 medianamnt\_ma\_rech90 209593 non-null float64

19 medianmarechprebal90 209593 non-null float64

20 cnt\_da\_rech30 209593 non-null float64

21 fr\_da\_rech30 209593 non-null float64

22 cnt\_da\_rech90 209593 non-null int64

23 fr\_da\_rech90 209593 non-null int64

24 cnt\_loans30 209593 non-null int64

25 amnt\_loans30 209593 non-null int64

26 maxamnt\_loans30 209593 non-null float64

27 medianamnt\_loans30 209593 non-null float64

28 cnt\_loans90 209593 non-null float64

29 amnt\_loans90 209593 non-null int64

30 maxamnt\_loans90 209593 non-null int64

31 medianamnt\_loans90 209593 non-null float64

32 payback30 209593 non-null float64

33 payback90 209593 non-null float64

34 pcircle 209593 non-null object

35 pdate 209593 non-null object

dtypes: float64(21), int64(12), object(3)

memory usage: 59.2+ MB

* Data Preprocessing Done

In this dataset there are no null values.so EDA process can be done with simple procdure. The object oriented categorical data can be removed by using drop method. There is no effect on drop those columns. Remove outliers can be used by Z score. After analyse the describe method the statistics standard deviation , Mean and median values, correlation and skewness.

* Data Inputs- Logic- Output Relationships

In inputs are effected

|  |  |
| --- | --- |
| aon | age on cellular network in days |
| daily\_decr30 | Daily amount spent from main account, averaged over last  30 days (in Indonesian Rupiah) |
| daily\_decr90 | Daily amount spent from main account, averaged over last  90 days (in Indonesian Rupiah) |
| rental30 | Average main account balance over last 30 days |
| rental90 | Average main account balance over last 90 days |
| last\_rech\_date\_ma | Number of days till last recharge of main account |
| last\_rech\_date\_da | Number of days till last recharge of data account |
| last\_rech\_amt\_ma | Amount of last recharge of main account  (in Indonesian Rupiah) |
| cnt\_ma\_rech30 | Number of times main account got recharged in last 30 days |
| fr\_ma\_rech30 | Frequency of main account recharged in last 30 days |
| sumamnt\_ma\_rech30 | Total amount of recharge in main account over last 30 days  (in Indonesian Rupiah) |
| medianamnt\_ma\_rech30 | Median of amount of recharges done in main  account over last 30 days at user level (in Indonesian Rupiah) |
| medianmarechprebal30 | Median of main account balance just before recharge in  last 30 days at user level (in Indonesian Rupiah) |
| cnt\_ma\_rech90 | Number of times main account got recharged in last 90 days |
| fr\_ma\_rech90 | Frequency of main account recharged in last 90 days |
| sumamnt\_ma\_rech90 | Total amount of recharge in main account over last 90 days  (in Indonasian Rupiah) |
| medianamnt\_ma\_rech90 | Median of amount of recharges done in main account over last  90 days at user level (in Indonasian Rupiah) |
| medianmarechprebal90 | Median of main account balance just before recharge in last  90 days at user level (in Indonasian Rupiah) |
| cnt\_da\_rech30 | Number of times data account got recharged in last 30 days |
| fr\_da\_rech30 | Frequency of data account recharged in last 30 days |
| cnt\_da\_rech90 | Number of times data account got recharged in last 90 days |
| fr\_da\_rech90 | Frequency of data account recharged in last 90 days |
| cnt\_loans30 | Number of loans taken by user in last 30 days |
| amnt\_loans30 | Total amount of loans taken by user in last 30 days |
| maxamnt\_loans30 | maximum amount of loan taken by the user in last 30 days |
| medianamnt\_loans30 | Median of amounts of loan taken by the user in last 30 days |
| cnt\_loans90 | Number of loans taken by user in last 90 days |
| amnt\_loans90 | Total amount of loans taken by user in last 90 days |
| maxamnt\_loans90 | maximum amount of loan taken by the user in last 90 days |
| medianamnt\_loans90 | Median of amounts of loan taken by the user in last 90 days |
| payback30 | Average payback time in days over last 30 days |
| payback90 | Average payback time in days over last 90 days |

The output is effected by label

Flag indicating whether the user paid back the credit amount within 5 days of issuing the loan{1:success, 0:failure}

* Hardware and Software Requirements and Tools Used

**Python**

**Pandas libraries**

**Numpy libraries**

**Seaborn libraries**

**Matplotlib.libraries**

**Scikit learn for import metrics to do macine learning**

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

The clear cut EDA process after cleaning the data to remove outliers

Correlation check each input and label target output variables.Then plot the pairplot and analyse the data of loan taken by amount 30 days and loan taken by 90 days the pay back period of 5 days after time each and individual data can analyse.

* Testing of Identified Approaches (Algorithms)

Listing down all the algorithms used for the training and testing.

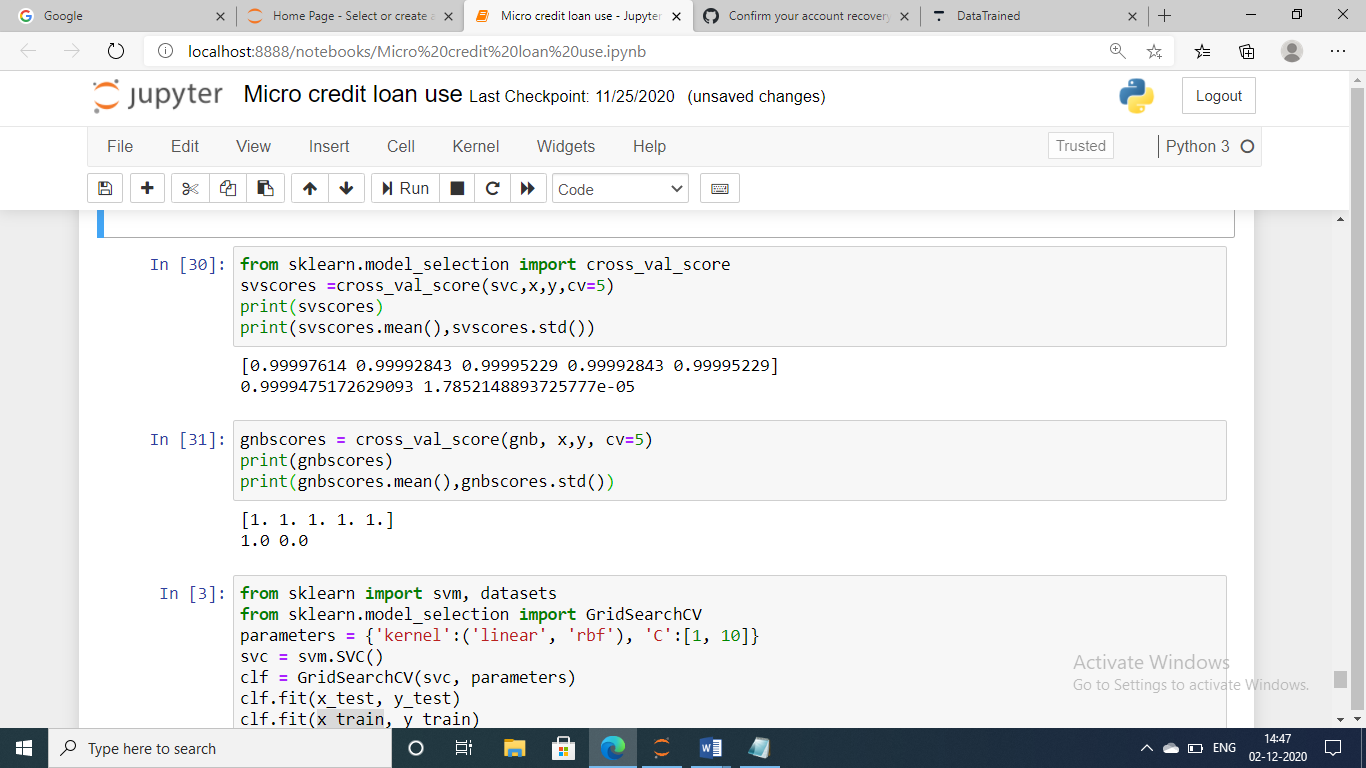
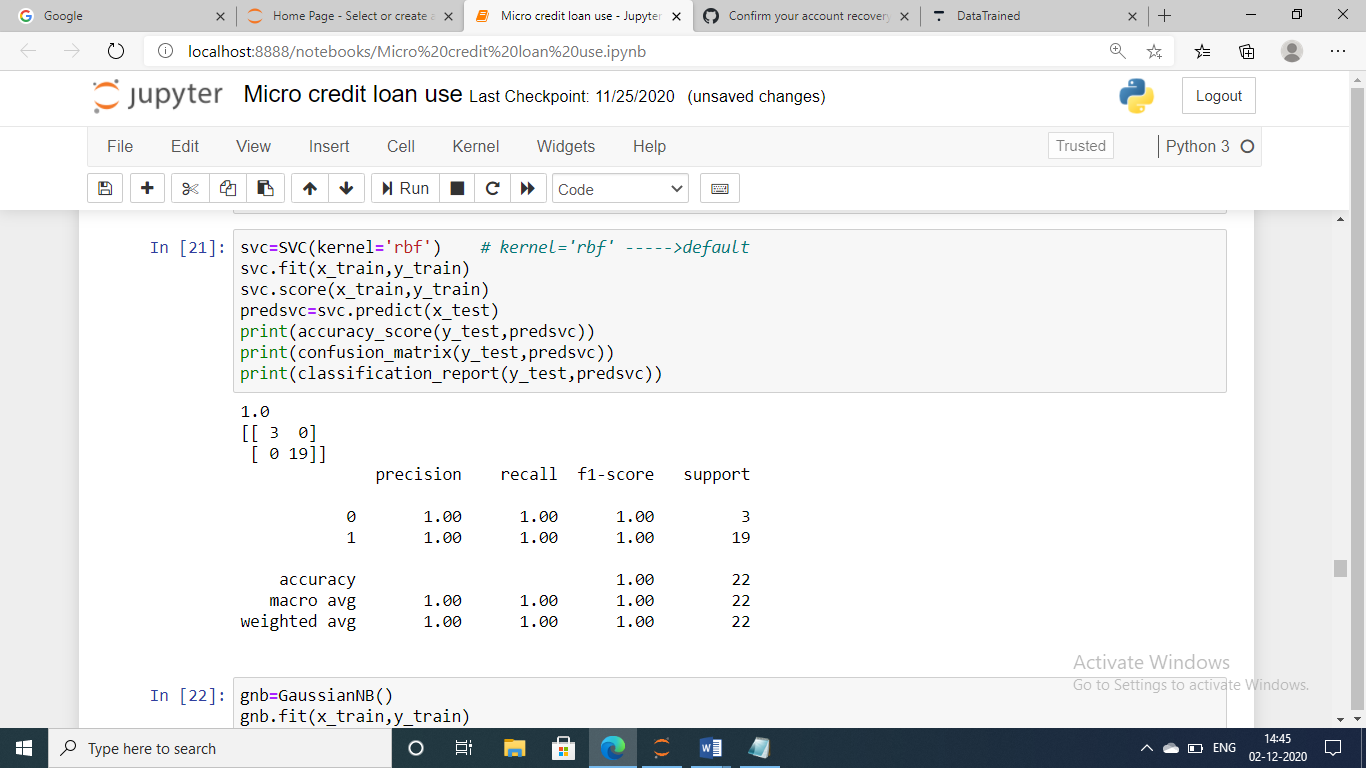
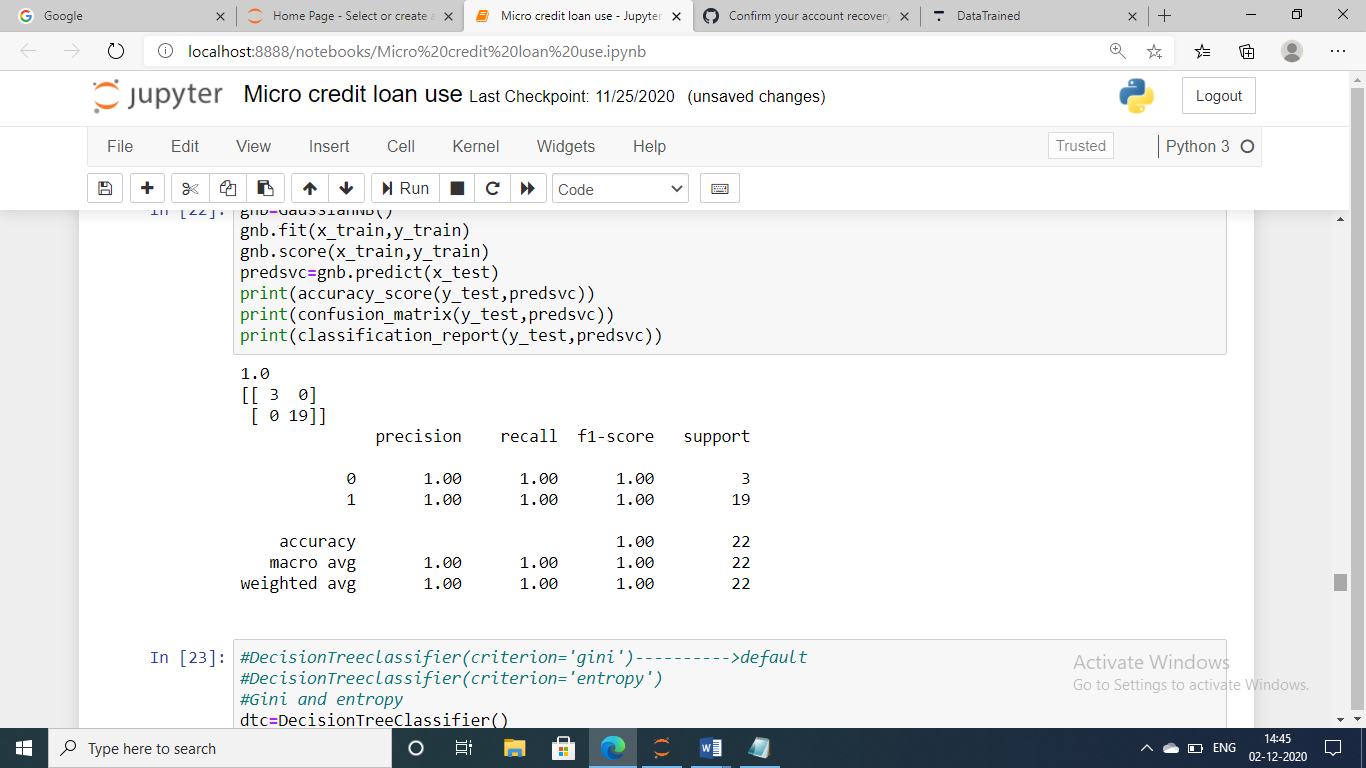
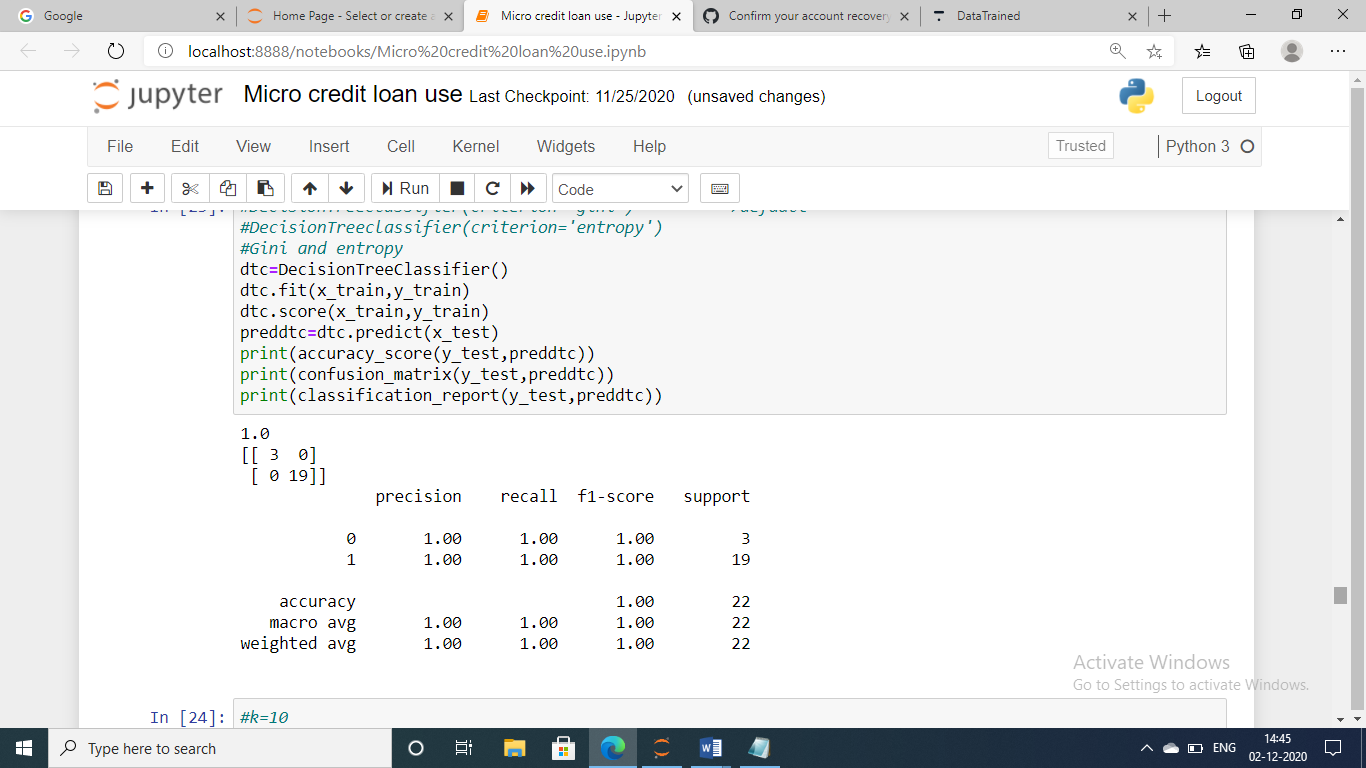
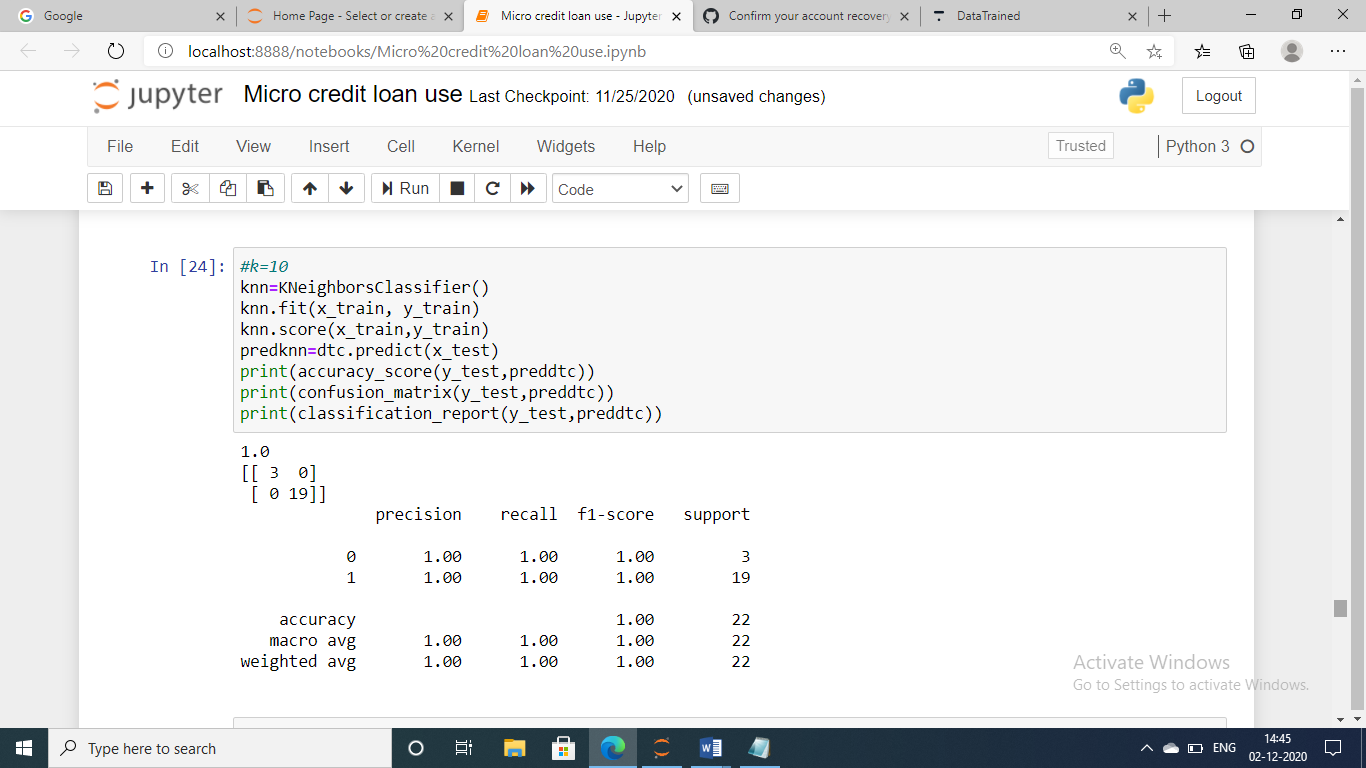
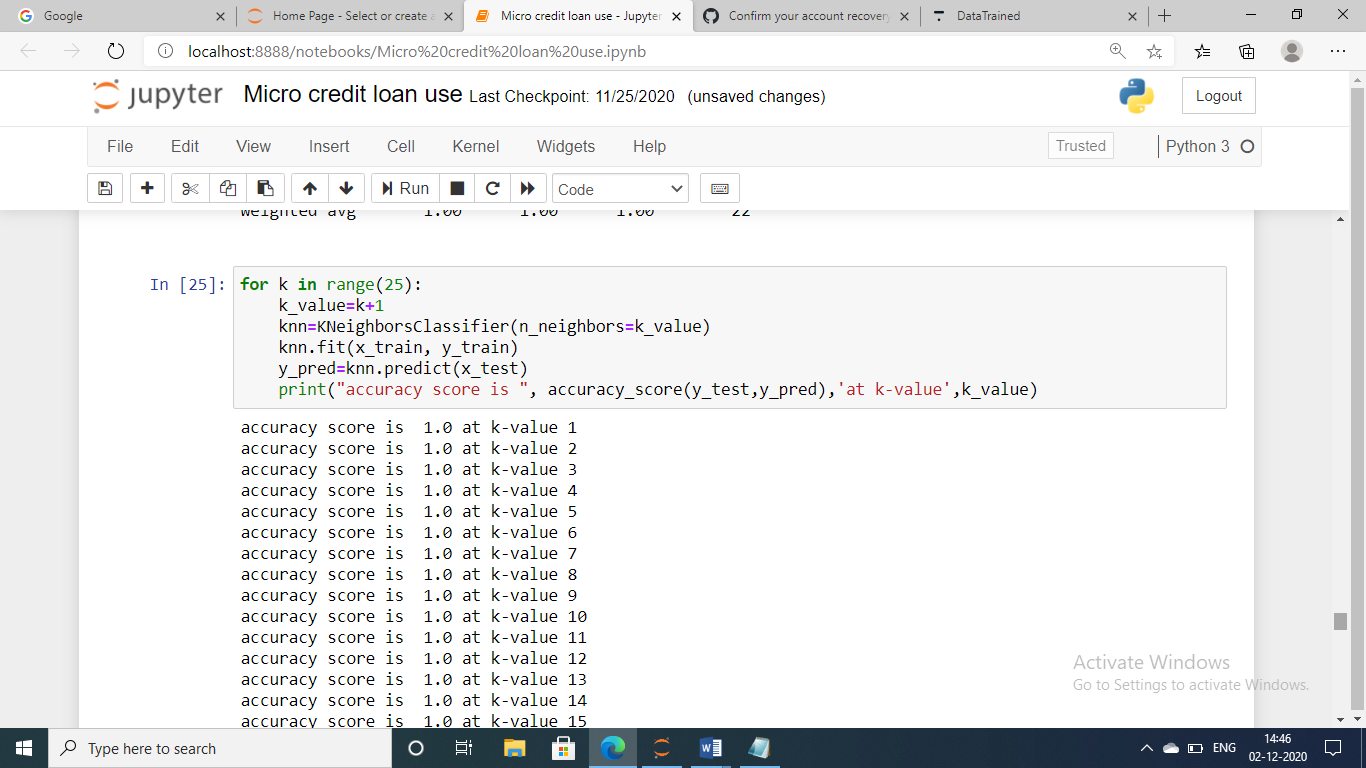
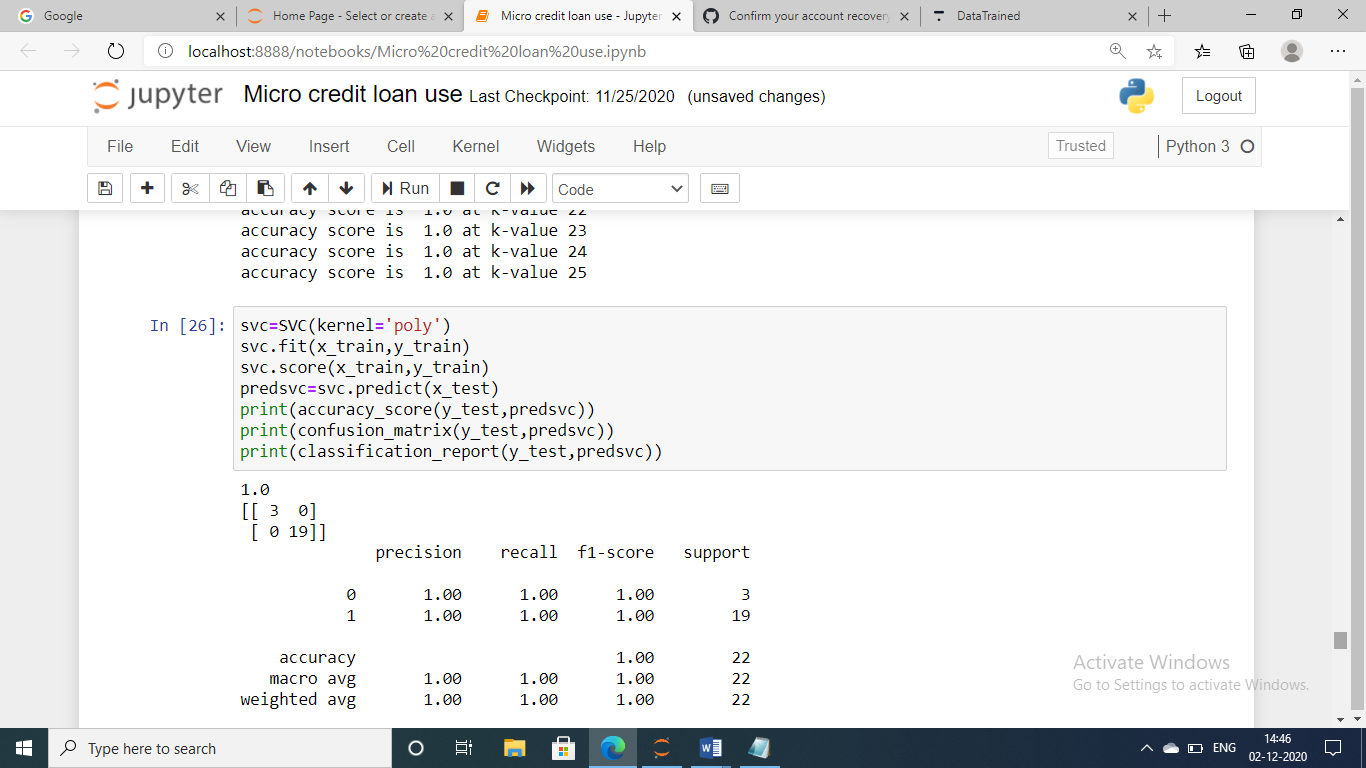
Decision Tree classifier

GuassianNB

Support vector machine

Knn Neighbours

* Run and Evaluate selected models



* Key Metrics for success in solving problem under consideration

Classification Metrics (**accuracy**, precision, recall, F1-score, ROC, AUC

* Interpretation of the Results

The summary of what results were interpreted from the visualizations, preprocessing and modelling.The target is payable accuracy is more so the loan amount can be paid maximum level of customers with in 5 days.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Some of the data imbalanced.

There is no null values in data.

The every input useful to predict output the loan is payable or not.

* Learning Outcomes of the Study in respect of Data Science

The data can be algorithms is suitable to the guassian Nb,decision tree classifier,knn,support vector machine models can be used.

* Limitations of this work and Scope for Future Work

The data is increases the lot of assumptions and visualization is different

In algorithms to improve accuracy.