PROJECT REPORT ON:-
Micro-Credit Defaulter

SUBMITTED BY:-

NAINCY JOSHI

DATASET DESCRIBTION:-

A Microfinance Institution (MFI) is an organization that offers financial services to low income populations. MFS becomes very useful when targeting especially the unbanked poor families living in remote areas with not much sources of income. The Microfinance services (MFS) provided by MFI are Group Loans, Agricultural Loans, Individual Business Loans and so on.

We are working with one such client that is in Telecom Industry. They are a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organization based on the budget operator model, offering better products at Lower Prices to all value conscious customers through a strategy of disruptive innovation that focuses on the subscriber.

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).

FEATURE DESCRIPTION:-

THERE ARE 36 COLUMNS IN THE DATASET

1.Flag indicating whether the user paid back the credit amount within 5 days of issuing the loan{1:success, 0:failure}.This is target column i.e (LABEL)

2. mobile number of user(misisdn)

3. age on cellular network in days(aon)

4. Daily amount spent from main account, averaged over last 30 days (in Indonesian Rupiah)(daily_decr30)

5. Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah)(daily_decr90)

6. Average main account balance over last 30 days(rental30)

7. Average main account balance over last 90 days(rental90)

8. Number of days till last recharge of main account(last_rech_date_ma)

9. Number of days till last recharge of data account(last_rech_date_da)

10. Amount of last recharge of main account (in Indonesian Rupiah)(last_rech_amnt_ma)

11. Number of times main account got recharged in last 30 days(cnt_ma_rech30)

12. Frequency of main account recharged in last 30 days(fr_ma_rech30)

13. Total amount of recharge in main account over last 30 days (in Indonesian Rupiah) (sumamnt_ma_rech30)

14. Median of amount of recharges done in main account over last 30 days at user level (in Indonesian Rupiah)(medianamnt_ma_rech30)

15. **Median of main account balance just before recharge in last 30 days at user level (in Indonesian Rupiah)(medianmarechprebal30)**

16. **Number of times main account got recharged in last 90 days(cnt_ma_rech90)**

17. **Frequency of main account recharged in last 90 days(fr_ma_rech90)**

18. **Total amount of recharge in main account over last 90 days (in Indonasian Rupiah)(sumamnt_ma_rech90)**

19. **Median of amount of recharges done in main account over last 90 days at user level (in Indonasian Rupiah)(medianamnt_ma_rech90)**

20. **Median of main account balance just before recharge in last 90 days at user level (in Indonasian Rupiah)((medianmarechprebal90)**

21. **Number of times data account got recharged in last 30 days(cnt_da_rech30)**

22. **Frequency of data account recharged in last 30 days(fr_da_rech30)**

23. **Number of times data account got recharged in last 90 days(cnt_da_rech90)**

24. **Frequency of data account recharged in last 90 days(fr_da_rech90)**

25. **Number of loans taken by user in last 30 days(cnt_loans30)**

26. **Total amount of loans taken by user in last 30 days(amnt_loans30)**

27. **Maximum amount of loan taken by the user in last 30 days(maxamnt_loans30)**

28. **Median of amounts of loan taken by the user in last 30 days(medianamnt_loans30)**

29. **Number of loans taken by user in last 90 days(cnt_loans90)**

30. **Total amount of loans taken by user in last 90 days(amnt_loans90)**

31. **maximum amount of loan taken by the user in last 90 days(maxamnt_loans90)**

32. **Median of amounts of loan taken by the user in last 90 days(medianamnt_loans90)**

33. **Average payback time in days over last 30 days(payback30)**

34. **Average payback time in days over last 90 days(payback90)**

35. **telecom circle(pcircle)**

36. **date(pdate)**

Importing libraries:-

It is very important to import the necessary libraries so that the next step will run smoothely.

Data Cleaning:-

The Important step in Machine Learning is data cleaning.

➢ We will check if there is a missing value in the dataset through the code "data.isna().sum()".The missing value can be replaced by using mean,mode and the median

➢ Data.describe() will give you detail about the mean and the standard value and it also provide detail about the quantile value.If we find it illogical then we can correct it using different technique.

➢ Data.info() will provide information about the types of the data.If there is a object datatype then we have to convert the 'object' into the machine understanding language such as 'int' or 'float'.

➢ Data.drop() it will drop that column/columns which are not related with each other or with the 'target or label' column.

Data Visualization:-

Data visualization help us to analysis the data through using different plot system.we can plot Histogram,Boxplot,Pie-plot,distribution plot,scatter plot and many more.It is more easy to understand and take quick decision.

Data Selection:-

WE will select those column which are useful and have relationship with the target column and which have no relationship will be dropped.

Data Preprocessing:-

In the step we will impute the missing Value.

We will convert the 'object' datatype into ML Language through 'power transformer()' or 'Label Encoder()' or 'Binary Encoder()'.We will check if there is a skewness or the outlier in the dataset.we will correct using 'Z-score' or correcting manually.

Data Splitting:-

We will split the dataset into two 'X' as the feature and 'Y' as the target or the label column. We will split the 'X' and the 'Y' dataset as the Training and the Testing dataset through 'train_test_split'.

Model Building:-

There are different technique used to build a model will depend upon the target dataset. If we have classification dataset in the column,we will use classification model and if we have continuous target column we will use regression model.In our dataset the target column is the

binary data i.e '1' as success and '0' as the failure,so we have used classification model.the model predict the outcomes in a better way.

Hyper-parameter Tunning:-

This helps to improve or tune the model with best accuracy parameter. Hyperparameter optimization finds a tuple of hyperparameters that yields an optimal model which minimizes a predefined loss function on given test data. The objective function takes a tuple of hyperparameters and returns the associated loss There are basic techniques such as 'Grid SearchCV', 'Random Search'.