**IMPLEMENTATION AND ANALYSIS OF PROJECT**

Business Idea: - Predict Customers which default in paying back loan so considered as bad loan.

Supervised Binary Classification Project

Domain – Banking Domain

**Implementation**

Data Shape – (20000,15)

Data Cleaning

Column wise data cleaning

1. id - drop

2. grade - perform label encoding

3. annual\_inc - performed outlier handling

4.short\_emp - drop as most values are 0

5.emp\_length – As it is

6.home\_ownership - handled missing values and applied one hot encoding

7. dti - Handle missing values

8. purpose - Replaced few used categories with 'other' and applied one hot encoding

9. term

10. last\_delinq\_none - As it is (final)

11. last\_major\_derog\_none - Drop as maximum values are missing

12. revol\_util - perform outlier handling

13. total\_rec\_late\_fee - removed as most of the values are 0

14. od\_ratio - remain as it is

15. bad\_loan - target variable

New Data Shape – (45211,12)

**Data Encoding**

Nominal Variables – One Hot Encoding and drop the originals

Ordinal Variables – Label Encoding

Numerical Variables – Do Nothing

**Imbalance Data**

Perform Oversampling so as to handle imbalance data

**MODELLING**

Logistic Regression

Decision Tree

Random Forest Classifier

**MODEL ANALYSIS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MODEL** | **ACCURACY** | **PRECISION** | **RECALL** | **F1-SCORE** |
| **LOGISTIC**  **REGRESSION** | 0.78 | FOR 0:-0.77 FOR 1:-0.79 | FOR 0:-0.80 FOR 1:-0.76 | FOR 0:-0.79  FOR 1:-0.78 |
| **DECISION TREE**  **CLASSIFIER** | 0.97 | FOR 0:-1.00  FOR 1:-0.95 | FOR 0:-0.94 FOR 1:-1.00 | FOR 0:-0.97 FOR 1:-0.97 |
| **RANDOMFOREST CLASSIFIER** | 0.89 | FOR 0:-1.00 FOR 1:-0.95 | FOR 0:0.95  FOR 1:-1.00 | FOR 0:-0.97 FOR 1:-0.97 |

**CODE**

Link for code in Google Colab :

https://colab.research.google.com/drive/1dfm7zZ1JsMoie-Dt3NePN48\_PlFbxb4H?usp=sharing

**INFERENCE**

According to me best model is Random Forest Classifier is best as accuracy and recall both are high for default loan case and any probability to predict persons which can default in loans in future is high.