Micro Credit Defaulter Project

Submitted by:

Sagnik Das

ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to (Datatrained) also to my SME (Mr Sajid Choudhary). Who gave me the golden opportunity to do this wonderful project on the topic (Micro Credit Defaulter Project), which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to tall of them.

INTRODUCTION

- Business Problem Framing

  This problem helps you understand the field of 'Micro Credit' and about the financial background of the people availing them.

- Conceptual Background of the Domain Problem

  The problem has been divided by EDA steps, then PCA analysis, confusion matrix and logistic regression, ROC and AUC curve, decision tree, Gradient boosting, knn classifier, Random Forest classifier, hyperparameter optimization.

- Review of Literature

  This is a comprehensive summary of the research done on the topic of "Micro Credit'. It also shows about the financial background of the people availing them. People who can return the loan in time and about the people who will be defaulter list.

- Motivation for the Problem Undertaken

  To find out about the group of people who can return the loan in time and about the people who will be defaulter list.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

PCA Standardization

$$z = \frac{value - mean}{standard\ deviation}$$

Covariance Matrix computation

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

Covariance Matrix for 3-Dimensional Data

Linear Regresion

$$Y = a + bX$$

ROC curve is a graph with: The x-axis showing 1 – specificity (= false positive fraction = FP/(FP+TN)) The y-axis showing sensitivity (= true positive fraction = TP/(TP+FN))

- Data Sources and their formats

Data Provided by Flip Robo Technologies

- Data Preprocessing Done
- Remove duplicate or irrelevant observations
- Fix structural errors
- Filter unwanted outliers
- Handle missing data
- Changed strings into integers
- Hyperparameter optimization using grid search cv
- Validate
-

- Data Inputs- Logic- Output Relationships

Micro credit defaulter model

EDA steps, then PCA analysis, confusion matrix and logistic regression, roc and auc curve, knn and hyperparameter optimization using grid search cv. This process helps us in isolating low-income families, high-income families and their money usage pattern. This also helps

microfinancing organizations to pinpoint the families that can and cannot repay loan amount within specified time.

- State the set of assumptions (if any) related to the problem under consideration

  Dropped down the columns and created an updated datafile.

- Hardware and Software Requirements and Tools Used

  Hardware – Laptop
  Software –
  import pandas as pd, import os, import csv, import sklearn, import numpy as np, import matplotlib.pyplot as plt, import seaborn as sns, %matplotlib inline, from sklearn.preprocessing import MinMaxScaler, PCA, knn, ROC and AUC curve, Hyperparameter optimization using grid search cv.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  EDA Method, PCA analysis, Covariance, Confusion Matrix, Linear Regression and prediction, PCA, knn, ROC and AUC curve, Hyperparameter optimization using grid search cv.

- Testing of Identified Approaches (Algorithms)

  EDA Method, PCA analysis, Covariance, Confusion Matrix, knn, ROC and AUC curve, Hyperparameter optimization using grid search cv.

- Run and Evaluate selected models

  Linear Regression, prediction, knn, ROC and AUC curve, Hyperparameter optimization using grid search cv.

- Key Metrics for success in solving problem under consideration

  Classification Matrix and Confusion Matrix.

- Visualizations

  Contour, Heat map, pair plot, box plot, histogram, gridsearchcv.

- ## Interpretation of the Results

  This process helps us in isolating low-income families, high-income families and their money usage pattern. Families taken or not taken loans. This also helps microfinancing organizations to pinpoint the families that can and cannot repay loan amount within specified time.

CONCLUSION

- ## Key Findings and Conclusions of the Study

  This process helps us in isolating low-income families, high-income families and their money usage pattern. Families taken or not taken loans. This also helps microfinancing organizations to pinpoint the families that can and cannot repay loan amount within specified time. This project also helps us to learn how big the world of microfinancing is and how it helps in eradicating poverty.

- ## Learning Outcomes of the Study in respect of Data Science

  EDA and PCA algorithm helped in cleaning excess data and helped it in putting it in visualised matter. Hyperparameter optimization using grid search cv helped to create a model.

- ## Limitations of this work and Scope for Future Work

Excessive unfiltered data.

Can help microfinancing organization identify needy families and can assist them.