

Loan Predictions

Problem statement

The datasets "accounts.csv" and "transactions.csv" presents loan services details to the customers. The firm is interested in giving additional loans to its existing customers. You need to develop a machine learning model for identifying the potential customers. Potential customers are those who may opt for top up loan services in future.

The project is divided into two parts.

First part would process the static data and train machine learning models based on them. Perform data cleaning and EDA on datasets. Identify the ML model that predicts the potential customers.

In second part, develop a flask application to predict if given customer will opt for top-up loan.

Datasets

- Required datasets: accounts.csv and transactions.csv

The accounts.csv file contains variables related to basic service information of the customers. The transactions.csv file includes the behavioural and transactional attributes of the customers.

Part 1 - Feature Engineering and Model building

1. Data cleaning:

- The dataset is large and it may contain some data error.
- In order to get clean, error free data some pre-processing requires.
- Refer accounts.csv and transactions.csv.
- Perform necessary data cleansing, convert data types in appropriate data types, remove duplicates, remove inapplicable columns, change date formats.
- Visualize missing data using bar plot and heat map. Handle missing values (if and as required).
- Save the cleaned data.

2. EDA:

- Observe the variables and their distribution with EDA.
- Create new columns if required.
- Save the output data.

3. Data Modeling:

- Handle the categorical column data.
- Prepare categorical features for correlation matrix.
- Employ label encoding technique i.e. use label encoder to convert categorical values to the numerical values.

4. Model Building:

- Split the data in training and testing data set.
- Build following machine learning models.
 - Logistic Regression model
 - KNN Model
 - SVM Model
 - Decision Tree Model
 - Random Forest Model
 - XGBM Model
 - LGBM Model
 - Boosting Model
- Evaluate performance of all these modules.
- Identify the best performing model for finding potential customers.
- Save that model as "potential_customers_prediction.pkl".

Part 2 - Prediction using Flask Application

- Implement a flask-based web application for prediction.
- Load the trained model "potential_customers_prediction.pkl".
- Create an route with ("/", method = "GET") to display "potential_customers_prediction.html". This page inputs required information for prediction.
- Create another route ("/predict"). It should process inputs from user, do the prediction using ML model and display result.

Technologies/Platforms Used

- Application Type: ML Model building and Web application
- Datasets: CSV
- Technologies:
 - Python Programming
 - Machine learning & data visualization related packages
 - Flask for UI

Evaluation

- Students are expected to code in pairs. e.g. For a sub-group size of 10 students, we will evaluate 5 applications.
- If application is completed (as per requirements), your technical interview score will be verified. It should be above 60% for each partner.
- Each application will get 1 mark for the full completion (subject to 60%+ interview score) and 0 mark if not completed. At the end, we will calculate the percentage score of each group by the formula: $\text{group score} = 100 * \text{Number of complete programs} / \text{expected number or pairs in the group}$.
- The points will be rewarded to the top 3 groups.
 - First group: 10 points
 - Second group: 5 points
 - Third group: 2 points
- If your program is ready, you should test each functionality thoroughly and then inform the lab mentors. Evaluations will be done in last 1 hour.