Credit Card Lead Prediction

Approach

1. Initial study of the problem

1. Understanding the features and its importance and made some hypothesis

2.Importing Important Libraries such numpy,pandas,matplot,seaborn,sklearn,scipy,etc

3.Data Preparation:

1.Checking for Null values using isnull() in both train and test

2.Checking for Categorical and Numerical features.

3.Imputation on missing values

4.Exploratory Data Analysis

1.Univariate Analysis:

Categorical features are analysed using count plots. Numerical features are analysed using Heat map and Pair plots.

5.Encoding:

1. Label Encoder for categorical features

6. Standardization of train input features with StandardScalar

7. Train data is split into train and valid sets in the ratio of 70:30 using train_test_split.

8. Training Models:

model1 = LogisticRegression

model2 = KNeighborsClassifier

model3 = LinearDiscriminantAnalysis

model4 = GaussianNB

model5 = DecisionTreeClassifier

model6 = XGBClassifier

model7 = GradientBoostingClassifier

model8 = RandomForestClassifier

Each model is fitted with train and valid(X_train,Y_train) and probability for class 1 is predicted in valid set(X_test). ROC_AUC score for each model is found and plotted in FPR vs TPR plot(ROC Plot).

9. Final Model – Random Forest Classifier gave a better roc_auc score. The result is loaded into submission file.

Other than the mentioned models, I've also tried Bagging Classifier, Adaboost Classifier, GradientBoostingClassifier, Voting Classifier along with Cross validation techniques.