Predictive Modeling for Customer Churn

Problem Statement: build a predictive model that can predict customer churn for a given company

Approach:

- **Reading Dataset:** first Importing train dataset into pandas data frame

- **Ask Basic Questions:** asking some basic question with that dataset like how big is the data ? how does the data look like ? what are the data types of columns ? are there any missing values ? are there duplicates values in dataset ? how does data look like mathematically ? how is the correlation between columns ? etc.

- **Exploratory Data Analysis:** getting some answer of questions then apply EDA on columns in that predicted column is imbalanced data and also independent columns are most of have imbalanced data and numerical column also have lots of outliers and right skewed data

- **Extract Columns:** extract independent columns and dependent column from train dataset and split training and testing dataset 80:20 ratio to check model performance

- **Pre-processing:** apply some preprocessing techniques on training dataset like ordinal encoding on ranked columns, one hot encoding on not ranked columns using column transformer and after that apply standard scaler on all column to better prediction of models

- **Model Selection:** apply various algorithms on training dataset like decision tree, logistic regression, svm, knn, random forest and gradient boosting in that algorithms best perform model is gradient boosting getting best result compare to others algorithm

- **Hyperparameter Tuning:** selecting gradient boosting classifier then turn the hyperparameter like max_depth, max_features, leaning rate, n_estimators etc. I get best result of training dataset of from the algorithm

Results of Algorithms

-	accuracy_score	precision_score	recall_score	f1_score	roc auc score
Decision Tree Classifier	0.883	0.427	0.484	0.454	0.706
Logistic Regression	0.913	0.656	0.275	0.386	0.629
Support Vector Classifier	0.913	0.643	0.297	0.406	0.639
K Neighbors Classifier	0.898	0.48	0.132	0.207	0.558
Random Forest Classifier	0.910	0.619	0.286	0.391	0.633
Gradient Boosting Classifier	0.916	0.6	0.495	0.542	0.729