Credit Card customers- Predict Churning customers

Domain Background :

Customer Churn is one of the most important and challenging problems for businesses such as Credit Card companies(Banks), Telecommunication service providers etc.. Customer Churning is relatively applicable to most customer facing sectors. Customer churn is directly proportional to Company outlook and it is a vital process to understand customers better.

Churn prediction aims to identify subscribers who are about to transfer their business to a competitor. Since the cost associated with customer acquisition is much greater than the cost of customer retention, churn prediction has emerged as a crucial Business Intelligence (BI) application for modern Business.

Kaggle Dataset that is associated with Credit Card Customer Churn Prediction is used all along for Nanodegree Program.

Reference :
https://www.kaggle.com/sakshigoyal7/credit-card-customers
https://epubs.siam.org/doi/abs/10.1137/1.9781611972801.64

Problem Statement :

A manager at the bank is disturbed with more and more customers leaving their credit card services. They would really appreciate it if one could predict for those who are getting churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction.

Reference : https://www.kaggle.com/sakshigoyal7/credit-card-customers

Dataset:

Dataset consists of 10,000 customers mentioning their age, salary, marital_status, credit card limit, credit card category, etc. There are nearly 23 features.We have only 16.07% of customers who have churned. Dataset Source: https://leaps.analyttica.com/home

Solution statement:

This is a classic labeled supervised classification task hence algorithms that are related to supervised classification machine learning algorithms such as Logistic Regression,Random Forest,SVM Classifier could be adopted.

Benchmarking:

Scikit-Learn models such as Gradient Boost , Random Forest,Light-BGM ,XGBosst will be used for Benchmarking. Dataset will be benchmarked across the mentioned models.
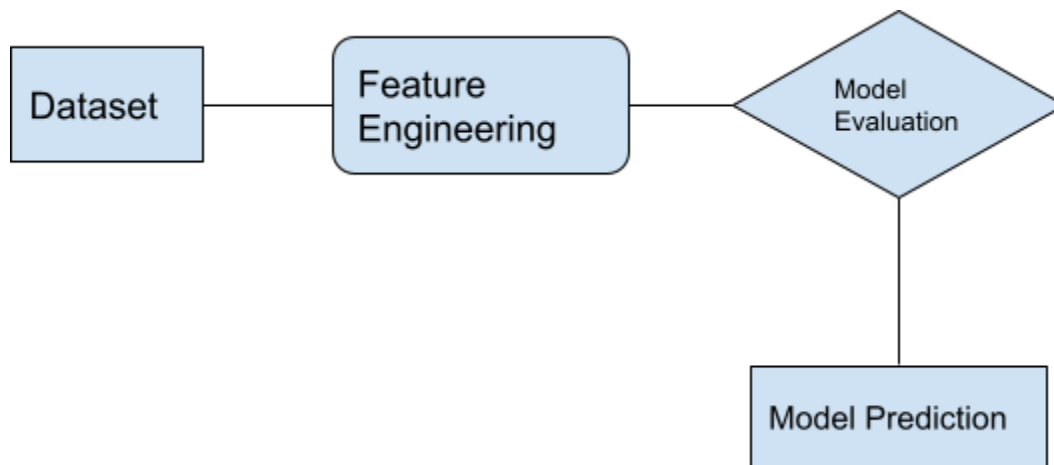
Evaluation Metrics:

Models would be evaluated across Precision ,Accuracy , Recall. Model with a higher Recall rate is preferred since non-churning customers predicted as churned would not harm the business.

$$Recall = (TP/TP+FN)$$

Model thus with higher Recall is the evaluation criteria

Project Design :

Dataset :

Data Source is in csv format, Pandas or Vaex would be used to read the dataset

Feature Engineering :

Feature Engineering includes Cleansing of dataset,EDA,Preprocessing of Dataset ,Transformation of Dataset , Feature selection and Feature Importance .

Model Evaluation :

Cleansed Dataset with qualified features would be split into Train,Validate,Test using Cross-Validation Technique and would be subject to various Model such as Logistic Regression, Random Forest ,SVM etc.. and one with higher Recall score would be subjected for Model Prediction.