In this machine learning churn project, we implement a churn prediction model in python using ensemble techniques.

Our learnings form the project will be as follows:

* Structuring and framing a business need into a Machine Learning problem statement
* Defining relevant metrics and setting the right expectations with business teams
* Exploratory Data Analysis - Univariate, Bivariate analysis
* Missing value and outlier treatment
* Label Encoder/One Hot Encoder and handling new categorical levels in test/production data
* Target encoding and avoiding data leakage
* Feature transforms (scaling and normalization)
* Feature engineering and Feature selection (RFE)
* Solving class imbalance
* Model explainability and interpretability through Tree visualizations and SHAP
* Hyperparameter tuning using RandomSearch and GridSearch
* Ensembling multiple models
* Error analysis
* Wrapping up code using Pipelines for production run

**Project Description**

A well-known bank has been observing a lot of customers closing their accounts or switching to competitor banks over the past couple of quarters. This has caused a huge dent in their quarterly revenues and might drastically affect annual revenues for the ongoing financial year, causing stocks to plunge and market cap to reduce significantly. The idea is to be able to predict which customers are going to churn so that necessary actions/interventions can be taken by the bank to retain such customers.

In this machine learning churn prediction project, we are provided with customer data pertaining to his past transactions with the bank and some demographic information. We use this to establish relations/associations between data features and customer's propensity to churn and build a classification model to predict whether the customer will leave the bank or not. We also go about explaining model predictions through multiple visualizations and give insight into which factor(s) are responsible for the churn of the customers.

This project walks us through a complete end-to-end cycle of a data science project in the banking industry, right from the deliberations during formation of the problem statement to making the model deployment ready.