Approach for solving the Credit Card Lead Prediction Problem

Problem Statement: Happy Customer Bank is a mid-sized private bank that deals in all kinds of banking products, like Savings accounts, Current accounts, investment products, credit products, among other offerings.

The bank also cross-sells products to its existing customers and to do so they use different kinds of communication like tele-calling, e-mails, recommendations on net banking, mobile banking, etc.

In this case, the Happy Customer Bank wants to cross sell its credit cards to its existing customers. The bank has identified a set of customers that are eligible for taking these credit cards.

Now, the bank is looking for your help in identifying customers that could show higher intent towards a recommended credit card, given:

- Customer details (gender, age, region etc.)
- Details of his/her relationship with the bank (Channel_Code,Vintage, 'Avg_Asset_Value etc.)

Approach:

- The approach followed in solving the problem is as follows:
- Loaded the train and the test dataset into the pandas data frame
- Performed Data Inspection to get the insights about the data
- Performed Data Cleaning by checking for null and the duplicated values
- Imputed missing data by performing data analysis about the feature where there are missing values and based on the class label
- Performed Exploratory Data Analysis to understand the distribution of each feature
- Performed univariate (single feature) and bivariate (two features) analysis as part of Exploratory data analysis
- Separated the train data into train and validation sets
- Tried different machine learning algorithms like logistic regression, linear SVC, Random Forest, XGBoost and CatBoost
- Performed Hyperparameter tuning for each model and found the best hyperparameters
- Selected the best model based on the roc_auc_score on the validation data and train data