ANALYTICS VIDHYA JOB-A-THON-v2

I used a catboost classifier to build the model and I found out that the most important features were the Credit_Product, Channel_Code, Age, Vintage and Avg_Account_Balance.

Most of my Feature Engineering Steps were based on these features.
I used a groupby on the Important Categorical Columns (Credit_Product and Channel_Code to get some new features based on these columns.

I filled missing values in Credit_Product with -999. I tried other methods as well but this gave the best score.

Approach::
1. Emulating the saying “Look Before you Leap” I Performed Exploratory Data Analysis (EDA) to understand how the data are distributed and if i had to deal with any irregularities

2. Merged the train and test data together to enable easy feature engineering and fix all irregularities

3. Filled missing values with arbitrary number (-999) to enable the model understand the point has a odd one and deal with it appropriately

4. Did non-visual analysis focusing on the target (Is_Lead) to save time and have a general understanding of the data

5. Dropped columns like ['ID' , Is_Active, 'Gender', 'Region_Code', 'Occupation'] as they had no real impact on the prediction performance of the model

6. Specified Cat_features in cat boost hyperparameter tuning cat_features=['Credit_Product', 'Channel_Code']

7. Did feature engineering based on my understanding of some features
    a. Specifically used groupby on some categorical columns and compute the aggregate (mean, std) on some numerical columns

8. Used CatBoostClassifier for modelling while specifying blind/baseline hyperparameters
    a. Used a 5-fold stratifiedKFold cross validation method to have a more generalizable model
    b. Specified categorical columns to be used as objects. This help improve the performance of my model a lot
    c. Hyperparameters used include random_state= 221,n_estimators= 3000, max_depth= 7 , verbose= 500 , learning_rate= 0.0102 , eval_metric= 'AUC'

9. To Improve the Model 5 fold catboost was stacked with 5 fold xgboost and it improved my score.

Link to code here

A big Thank you to the organizers (Analytics Vidhya) of this wonderful hackathon
Thank You Abubakar Alaro