APPROACH FOR AV-CONTEST

STEPS:

1. Importing both train and test data.
2. Checking null values and imputing with new class ‘no_card’ instead of ‘nan’ in CREDIT_PRODUCT.
3. Separating categorical and numerical columns.
4. Checking any new region code in test data.
5. EDA--- Distribution plot for numerical values to check any possibility to change to normal –Log transformation for ‘Avg_Account_Balance’.
6. EDA----Boxplot for checking outliers in numerical data with respect to ‘Is_Lead’.
7. EDA—Categorical count plot for all classes with respect to Is_Lead .
8. Problem interpreted as Imbalance classification model.
9. Label encoding categorical data for both train and test.
10. Applying first catboost model for with just label encoding.
11. One-hot coding on categorical data.
12. Selecting categorical features using chi2 feature selection method with threshold 500.
13. Applying second catboost model with selected categorical and all other numerical columns (This model gave best).
14. Predicting test data.