Health Insurance Lead Prediction

Steps Performed:

1. Imported the Datasets
2. Explored the Train and Test Datasets
3. Removed the Duplicates in the dataset if any.
4. Handled Missing values in the dataset using Forward Fill Technique for 3 features which have missing values.
5. There are multiple steps performed in feature engineering technique. They are :
    a. Replaced "Is_Spouse"column with 0 and 1 and made it a binary categorical variable.
    b. Replaced "Accomodation_Type"column with 0 and 1 and made it a binary categorical variable.
    c. Replaced "Reco_Insurance_Type"column with 0 and 1 and made it a binary categorical variable.
    d. Replaced "14+"category in the feature named "Holding_Policy_Duration" to "15".
    e. Removed "Lower_Age" feature because two Age related features are highly correlated.
6. Converted all the datatypes to Numerical columns because the algorithm reads only numbers.
7. Performed Feature Scaling
8. Trained Individual Models of Decision Tree Classifier, Random Forest Classifier, XG Boost Classifier, Light GBM Classifier, Cat Boost Classifier.
9. Generated Predicitions for Individual Models.
10. Since the target column is a binary categorical variable, I have considered Voting Classifier by including the models Random Forest, XG Boost, Cat Boost, Light GBM and generated the final prediction.