**Credit Card Lead Prediction**

**JOB-A-THON MAY 2021**

The objective of this competition was to predict whether the customer is interested in buying Credit Card or not given their details and relationship with the bank.

I tried the following things for data preprocessing and predictive modelling:

* First, I built baselines by Label Encoding all categorical columns. I filled missing values by the majority value in that column and used XGBoost, LightGBM, Random Forest Classifiers with default parameters. XGBoost and LightGBM performed better out of all, so, I used them for further improvement.
* After getting the baselines, I tried different combinations of Label Encoding, Target Encoding, Ordinal Encoding, Imputing missing values with KNN Imputer. These gave slightly better scores than the baseline models.
* Since Boosted Tree Algorithms are good at handling missing values, I decided not to impute the values (Only Label Encoded Categorical Values) and let the models handle them. This gave a better score than before.
* Finally, the best model that I was able to come up with was a blend (average) of predictions from XGBoost and LightGBM (Label Encoded Categorical Features and Leaving Null Values as it is) which were tuned using Randomized Search with 5-Fold Cross Validation.

Things I tried but didn’t work:

* Taking Log transform of Continuous Variables.
* Removing Feature with Low scores according to the models.
* Removing Correlated Features.