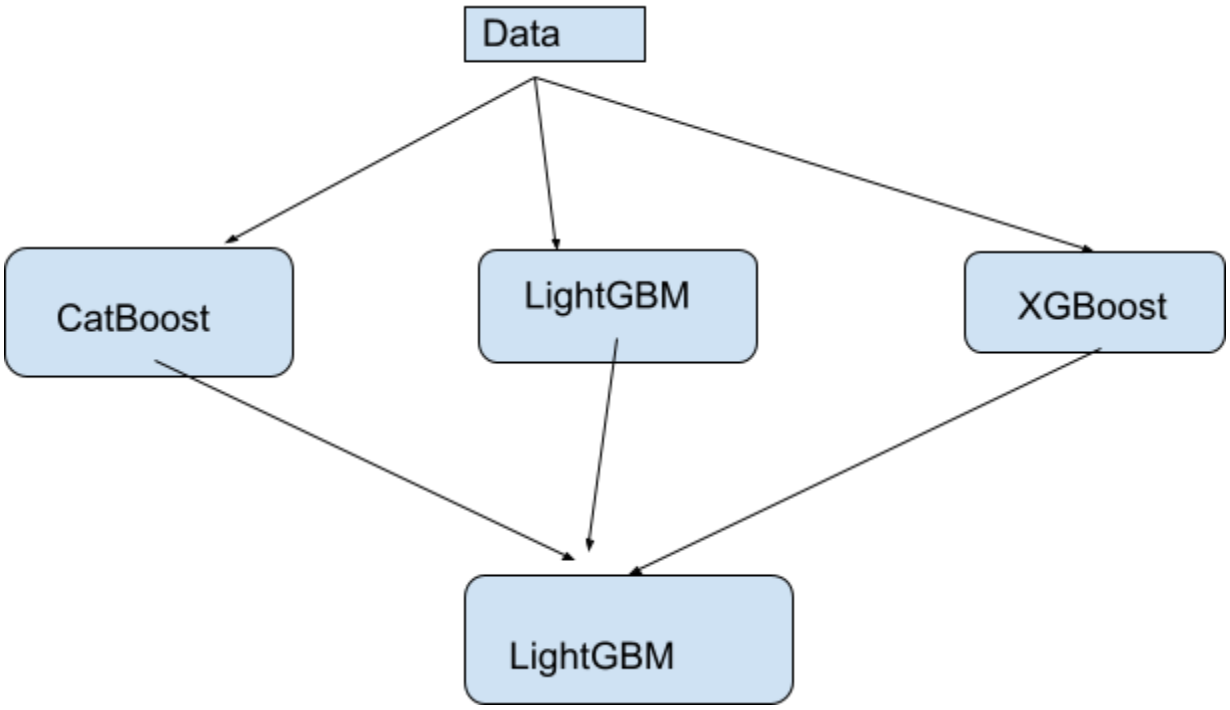
JOBATHON-2

This dataset aims to solve the problem of finding the customer who would like to take interest in taking the credit. The data contains the details of customers who all are eligible for taking credit cards

1. Architecture of Final Solution



2. Approach For the Problem

My approach for the problem was simple. First I build the baseline model with logistic regression and submit it. After that I started exploring models and checked my score for the validation set. When I reached a decent score with the Catboost model I submit the model and achieved a decent rank in the public leaderboard. Then I started exploring other models like LightGBM and XGBoost but my score was not increasing on the validation dataset. Then I did some EDA and created some more features according to my analysis and created the Catboost model and it did help me to reach the top 100 in the public leaderboard. I also did the outlier removal also but that did not work and model performance decreased then I realized sometimes in the finance sector domain outliers helps us to capture the underlying patterns in the data which rarely occurs. I applied 10 Fold Cross validation to train all my models and reached to final model.

3. Feature Engineering Ideas / Pre-Processing Ideas

I tried several preprocessing ideas, some worked but some did not work. I was trying to impute the missing values but in the later tree based there is no need to imputing the

missing values , imputing it with some random value also degrades the performance of the model so I did not finally imputed the missing value.
From EDA I figured out Gender and Avg Balance was major factors in determining the final output. So I created various features from Gender and Avg Balance and it did worked out