Approach Followed High Level

1. Download and load the dataset in google colab
2. Performed Basic Data Exploratory analysis
3. Imputing the missing (NAN’s) in the dataset
4. Feature Engineering – created and added few new features from the existing data.
5. Normalize / Scale the data (Used both MinMaxScaler and StandardScaler)
6. Model building (models tried)
   1. Logistic Regression
   2. DecisionTreeClassifier
   3. Naïve Bayes - GaussianNB
   4. RandomforestClassifier – bagging model
   5. XGBoostClassifier – Boosting model
   6. LightGBMClassifier
   7. CatBoostClassifier
7. Used Stratified Fold Cross Validation and RandomizedSearchCV technique.
8. CatBoostClassifier Gave best Results followed by LGBM and XGB.
9. Finally tried simple average on multiple model predictions (CB, LGBM, XGB) and Rank Average
10. The best score I got is 0.6617113498