Summary Report on Model Performance

After testing different algorithms to fit the data, such as Logistic Regression, Decision Tree, Random Forest, Catboost, LightGBM, and Neural Network, experimented with various hyperparameters combinations for each model & compared their performance based on FDR at a 3% rejection rate.

Random forest with parameters: max_depth=8, criterion = 'gini', n_estimators = 80, min_samples_split = 50, min_samples_leaf = 40, max_features = 8, was the best and final model.

The FDR scores for this model on training, testing, and OOT data were 78.52%, 78.21%, and 55.41%, respectively. (Taking number of variables =10)
Final model shows that we are not overfitting and getting good performance and we can catch 55.41% of all the Fraud by rejecting the top 3% of the application.

Performance Score (FDR @ 3%) of different models with best model parameters