Train/Test Report

...

Training objective:
Binary Classification

Multi-Model Approach

Logistic Regression

Considering Binary classification problem Logistic Regression was our first go to model

- Easy to implement
- Fast

XGBoost

Discovering that our data had lot of missing values and was fairly imbalanced, we considered XGB

- handles missing values very well
- gradient boosted trees for imbalanced data

MLP (Keras)

MLP was more sort of a experimental approach that we considered

- Can give exceptional results
- Data heavy

Logistic Regression

Accuracy Score:

Train : 87.51%

Test : 86.87%

Logistic Regression - Confusion Matrix

Train

No. of Samples	Predicted Failure	Predicted Successfull
Actual Failures	22614	17
Actual Successfull	3211	2

Test

No. of Samples	Predicted Failure	Predicted Successfull
Actual Failures	5613	9
Actual Successfull	839	1

XGBoost Classifier

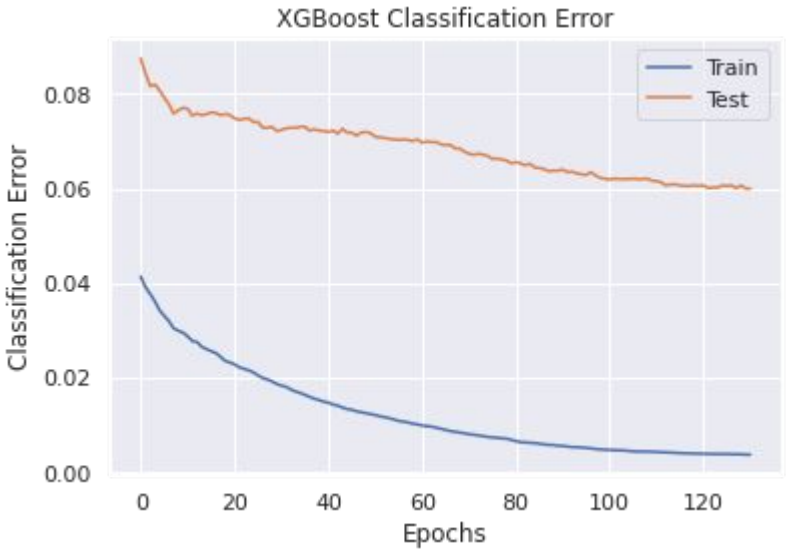
Accuracy Score:

Train : 99.60%

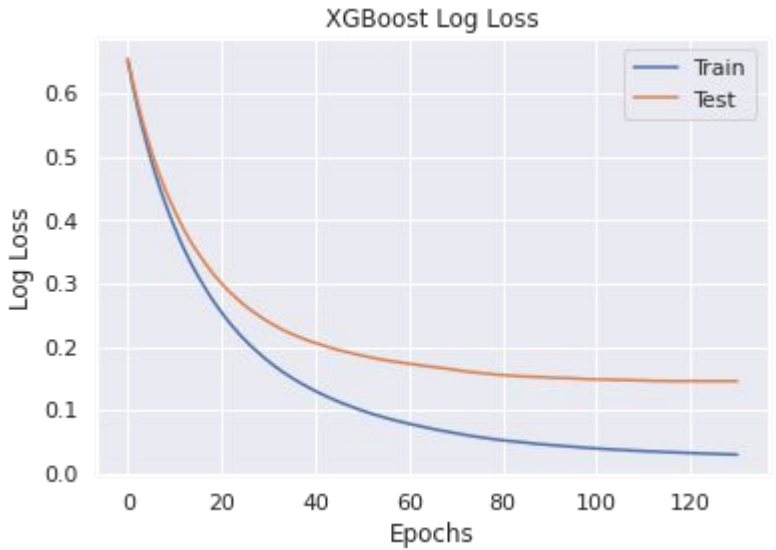
Test : 93.94%

XGBoost Classifier - Training Scores

Classification Error



Log Loss



XGBoost Classifier - Confusion Matrix

Train

No. of Samples	Predicted Failure	Predicted Successfull
Actual Failures	31701	10
Actual Successfull	134	3944

Test

No. of Samples	Predicted Failure	Predicted Successfull
Actual Failures	7702	177
Actual Successfull	365	704

MLP Keras Classifier

Accuracy Score:

Baseline Model : 87.45%
(2 Layers)

Deep Model : 87.57%
(4 layers)

Model Selection

Findings

XGBoost Outperforms Logistic Regression as well as our current MLP model

Conclusions:

- XGBoost to be solely solely for current of InvestRight Predictor
- Work on better Neural Network Models