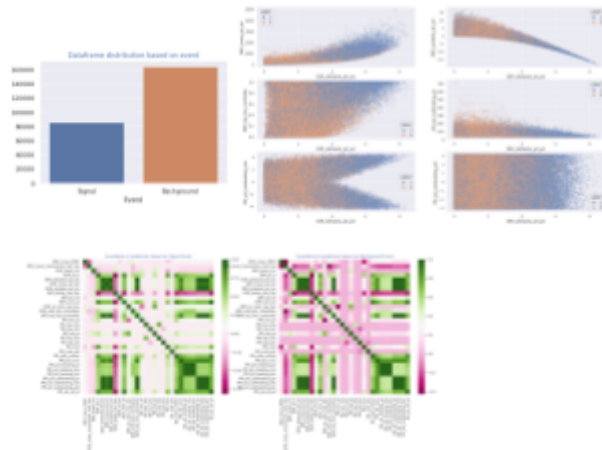


The events detected by the detector at the LHC in large majority, represent known processes(called background) that are mostly produced by the decay of exotic particles. The search for new physics provided by unknown events (called signal) is difficult because these processes are rare

Data and Labels

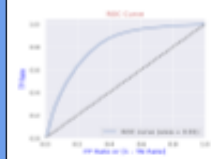


References

<https://www.kaggle.com/code/sugataghosh/higgs-boson-event-detection-part-1-eda>

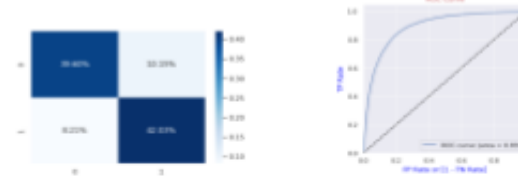
<https://jovian.ai/swarnavab98/telecom-churn-cs>

Logistic Regression (Baseline Model)

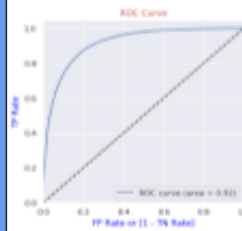


F1 score = 0.75
AUC for ROC = 0.81

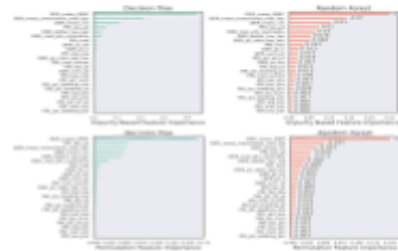
Decision Tree



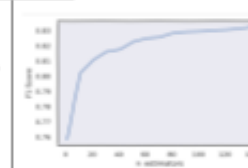
Hyperparameter Tuning gives
F1 score = 0.82



Random Forest



AdaBoost



Model Selection

Labels:

We choose Random Forest model with AUC for ROC as 0.92 as our final model for the classification of signal and background events

Weights:

We choose the Decision Tree Regressor with R2 score 0.6513 as the best model for predicting weights

Conclusion and Future Work

We can further analyse the dataset in depth and find the correlations/multicollinearity among the predictors.

We can explore more by performing EDA through Univariate/Bivariate/Multivariate analysis.

We can do RFE(recursive feature elimination), VIF(Variance Inflation Factor) to explore the importance of features.

We can also optimize the threshold value for classification problem

We can perform downscaling of the imbalance dataset and evaluate the models.

we can use XGBoost, Neural Networks(MLPs) to improve the model training.