**CHAPTER 7**

**PERFORMANCE ANALYSIS**

## 7.1 PERFORMANCE METRICS

## ROC-AUC SCORE AND CURVE

The ROC (Receiver Operating Characteristic) curve is a graphical representation of the performance of a binary classifier system as its discrimination threshold is varied¹. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two groups (Edge present /(Edge Absent). AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0.

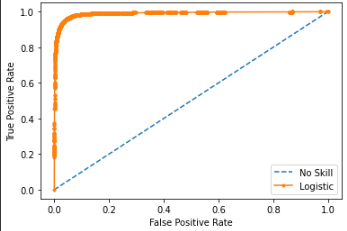


Figure 7.1 ROC-AUC Curve for Logistic Regression

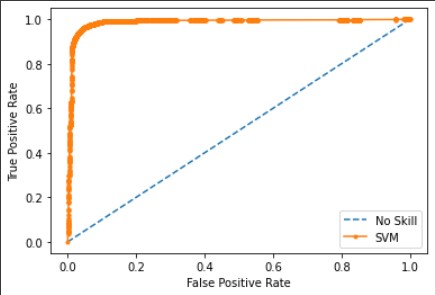
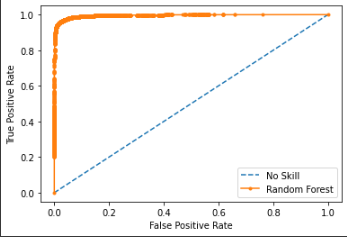


Figure 7.2 ROC-AUC Curve for Support Vector Machine



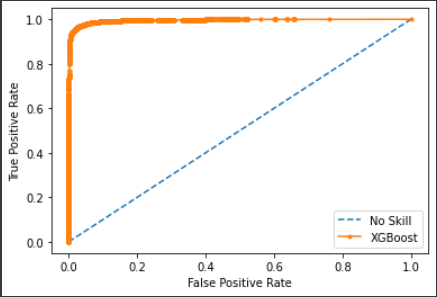


Figure 7.3 ROC-AUC Curve for Random Forest

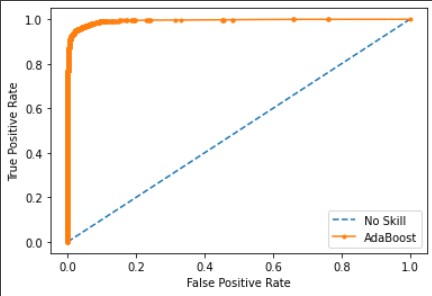


Figure 7.4 ROC-AUC Curve for AdaBoost

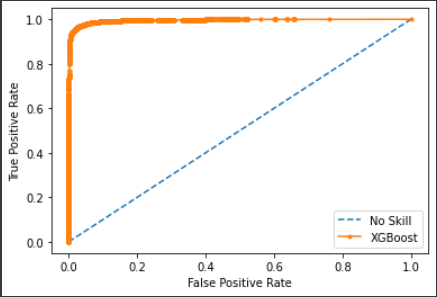


Figure 7.5 ROC-AUC Curve for XGBoost

Table 7.1 AUC SCORE for Machine Learning Models

|  |  |
| --- | --- |
| Machine Learning Model | AUC SCORE |
| Logistic Regression | 0.988 |
| Support Vector Machine | 0.985 |
| Random Forest | 0.995 |
| AdaBoost | 0.994 |
| XGBoost | 0.995 |

## CLASSIFICATION REPORT

## 7.1.2.1 PRECISION

It is the number of correct positive results divided by the number of positive results predicted by the classifier.



Figure 7.6 Precision Formula

## ****7.1.2.2 RECALL****

It is the number of correct positive results divided by the number of **all**relevant samples (all samples that should have been identified as positive).

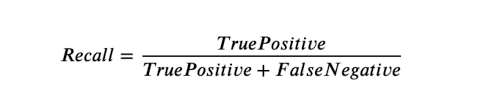


Figure 7.7 Recall Formula

## ****7.1.2.3**** F1 SCORE

F1 Score is the Harmonic Mean between precision and recall. The range for F1 Score is [0, 1]. It tells you how precise your classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances)

.

Figure 7.8 F1 Score Formula

## ****7.1.2.4**** ACCURACY

Classification Accuracy is what we usually mean, when we use the term accuracy. It is the ratio of number of correct predictions to the total number of input samples.



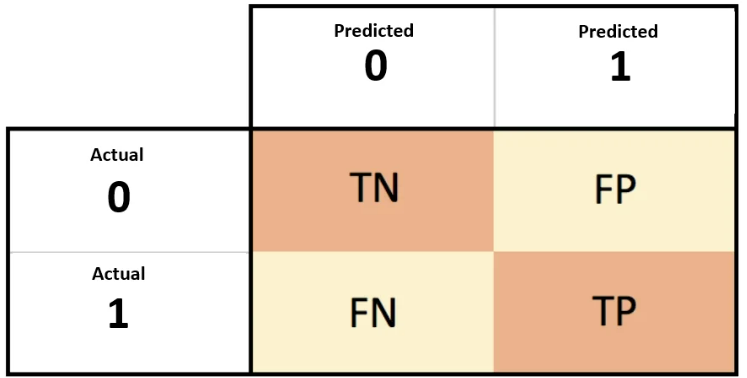
Figure 7.9 Accuracy Formula

## CONFUSION MATRIX

## A confusion matrix is a table that is used to define the performance of a [classification algorithm](https://www.sciencedirect.com/topics/engineering/classification-algorithm). A confusion matrix visualizes and summarizes the performance of a classification algorithm.

## **true positives**: the cases in which we predicted yes and the actual output was also yes.

* **true negatives**: the cases in which we predicted no and the actual output was no.
* **false positives**: the cases in which we predicted yes and the actual output was no.
* **false negatives**: the cases in which we predicted no and the actual output was yes.



## Fig 7.10 Confusion Matrix

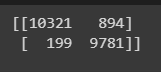


Fig 7.11 Confusion Matrix for Logistic Regression

https://lh4.googleusercontent.com/BTfzfRfSy7eOg5HvU9oJchR-LG7b7Ulmbg5bpk_n43aYGCq0DerWO9oxXUCts9Yv_2XhvG3_NshWKqokjPmqVyIKQl44VuCQ9ZOh2n0voCKDcBE3llchmZAwc6utq0jO4ZepCOMOzpFVghJ35YSzj7U

Fig 7.12 Confusion Matrix for Random Forest

https://lh4.googleusercontent.com/yQtCAbtPCxIrUNs56yFDcGmAc2NgeEPTUY_AVlZvEX7XVYOfUi_YzH0ETJ8xHZwNANosIu_FIHK99jmZQIIo3EIlfb8IeSVznznQE33RgdRw8trHaEzD-QvE0LYWG4BupqOPC2CD-0THXQqLQ2mAdQI

Fig 7.13 Confusion Matrix for Random Forest Support Vector Machine

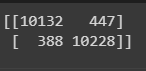


Fig 7.14 Confusion Matrix for AdaBoost

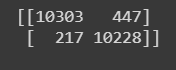


Fig 7.15 Confusion Matrix for XGBoost

## RESULTS AND EVALUATION

We observed very good accuracy for all the models, with Random Forest, XGBoost and AdaBoost performing the best. The following table shows various classification metrics for each of the machine learning models.

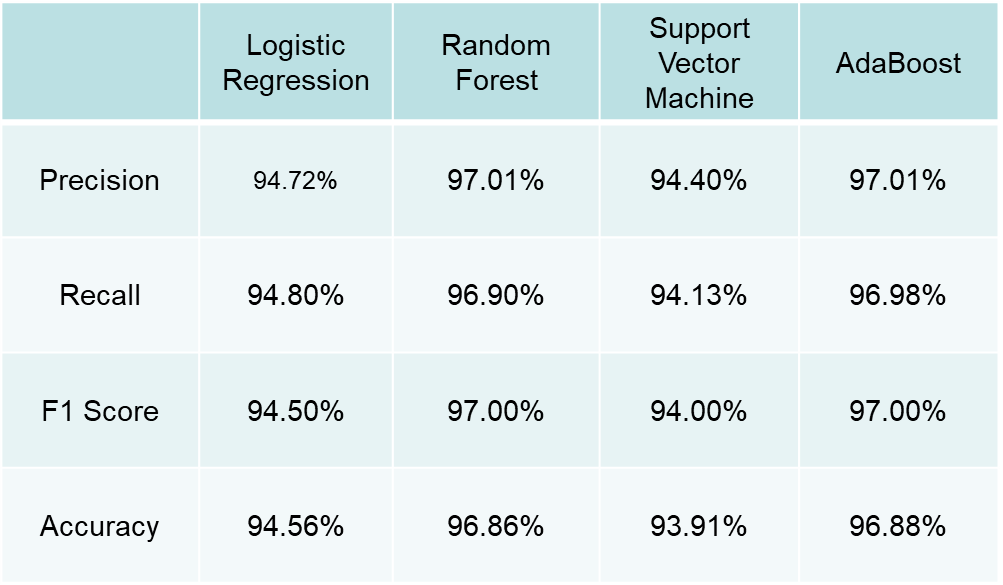
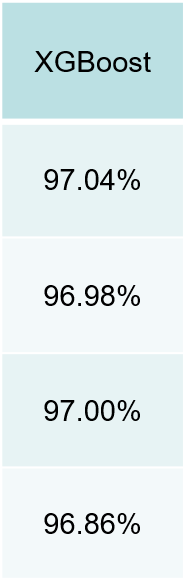


Fig 7.16 ML Model Performance Metrics