Model Performance Report

Handling null values in each column:

1. As column9 has 9,75,990 null values out of 10,46,845 it is better to drop the column.
2. we have discrete value for column3 so, we will fill the Nan values with mode of distribution.
3. After performing log transformation on column 5, filling the null values of column5 with median as there are outliers in the data.
4. After performing log transformation on column 6 we are getting a normal distribution so, we will fill the missing values of column6 by mean value.
5. In column 8 we will choose median value for skewed and outlier data set.
6. Filling null values of column0, column14 and column15 with mode, because while plotting the histplot we are getting discrete values with very high frequency.

Model Building:

1. First we performed standardization of the data, which basically redistribute data such as mean of data becomes zero and standard deviation becomes 1
2. Then we starting testing our model performance on various algorithms like , Logistic Regression, Decision Tree classifier, and AdaBoost classifier.
3. Out of all these AdaBoost is performing the best so, we decided to choose this model.
4. Then we performed fine tuning with the help of Grid search cv , which gives us the best parameters to use in order to get best performances

Model Evaluation:

After fine tuning our model is giving us these scores:

1. Accuracy:0.9763518677019013
2. Roc Auc score :0.9638379760545387
3. Precision: {class 0:0.99} and {class 1:0.83}
4. Recall: {class 0:0.98} and {class 1:0.95}
5. f1-score:{class 0:0.99} and {class 1:0.88}
6. Our True Positives and True negatives are significant in numbers
7. Our model is providing us least number of False positives