Summary

Started by removing the null values. The various columns were studied and the columns with very high number of null values (>30%) were removed. After that, the columns with smaller percentage of null values had their null values handled by different strategies. After the null values were handled. EDA was performed to gain understanding of the data. After the EDA was complete, the data had to be prepared for modelling which included two steps. The null values were already handled. So only converting object type columns into numerical form was left. Hence creating dummies for the categorical dummies was done and we were left with very clean data which was non null and all numerical. After that Model was initialised and rfe was used to select top 20 features for modelling. After we got the top 20 features, model was trained using those features. The model performed well on both train and test datasets. The accuracy, precision, recall as well area under the curve score were really high. Over all the model performance was very satisfactory as there was no sign of overfitting or underfitting.