**Data processing:**

After analysing the data it was found that lot of columns had missing values. Thus it was necessary to fill these values to train the machine learning model. In order to deal with missing values the following approach was taken.

* Find the columns which had 70 % or more fields as null. Delete these columns, as more than half of its field had to be filled. Even if these columns are kept after filling missing columns it would not help in predicting the target value as most of these are not real data.
* Fill the columns were percentage of missing values was less than the 70% using interpolation.
* Better approach would be to impute these values using KNNimputer or missforest. But due to hardware constrains, next best approach was taken.

Percentage of missing values for a column= (sum of fields as null in that column /total number of fields)\*100

**Feature scaling:**

Since the data ware in different scales, standard scaler was used to make sure that the features were in the same scale. (It is also necessary to perform feature scaling as Gaussian kernel is used)

**Prediction:**

For prediction support vector machine classifier with Gaussian kernel was used.

* Since all the value was numeric and the data was nonlinear, SVM would be better as it would easily generalize to new data.
* Gaussian kernel was chosen as the number of features where was small compared to the number of examples