HACKTHON PROJECT LEVEL-2

- BHARATESHA N S

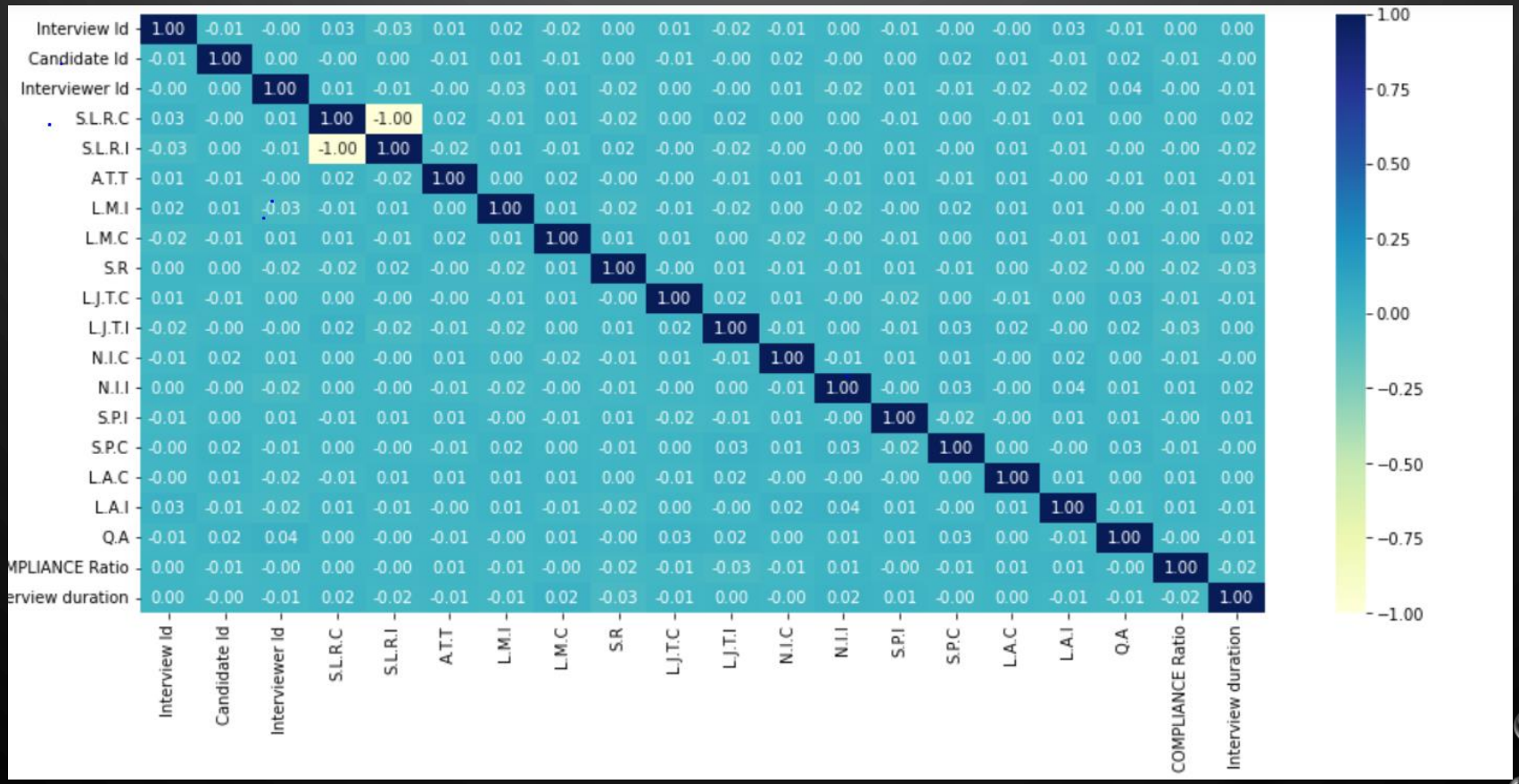
PROBLEM STATEMENT:

- An MNC company conducting an interview hiring more than 10k candidate .
- Based on each candidate's performance they divided into three category
    i. Consider
    ii. May consider
    iii. Not consider
- Now build a model which gives accuracy based on the hiring status .

SOLUTION APPROACHES

1.Checking the Correlation by plotting heap map.

Removing the Highly negatively Correlated Columns.



2. HANDLING MISSIG VALUES

- Some categorical features had missing values .
- To resolve this problem of missing values treat with MODE

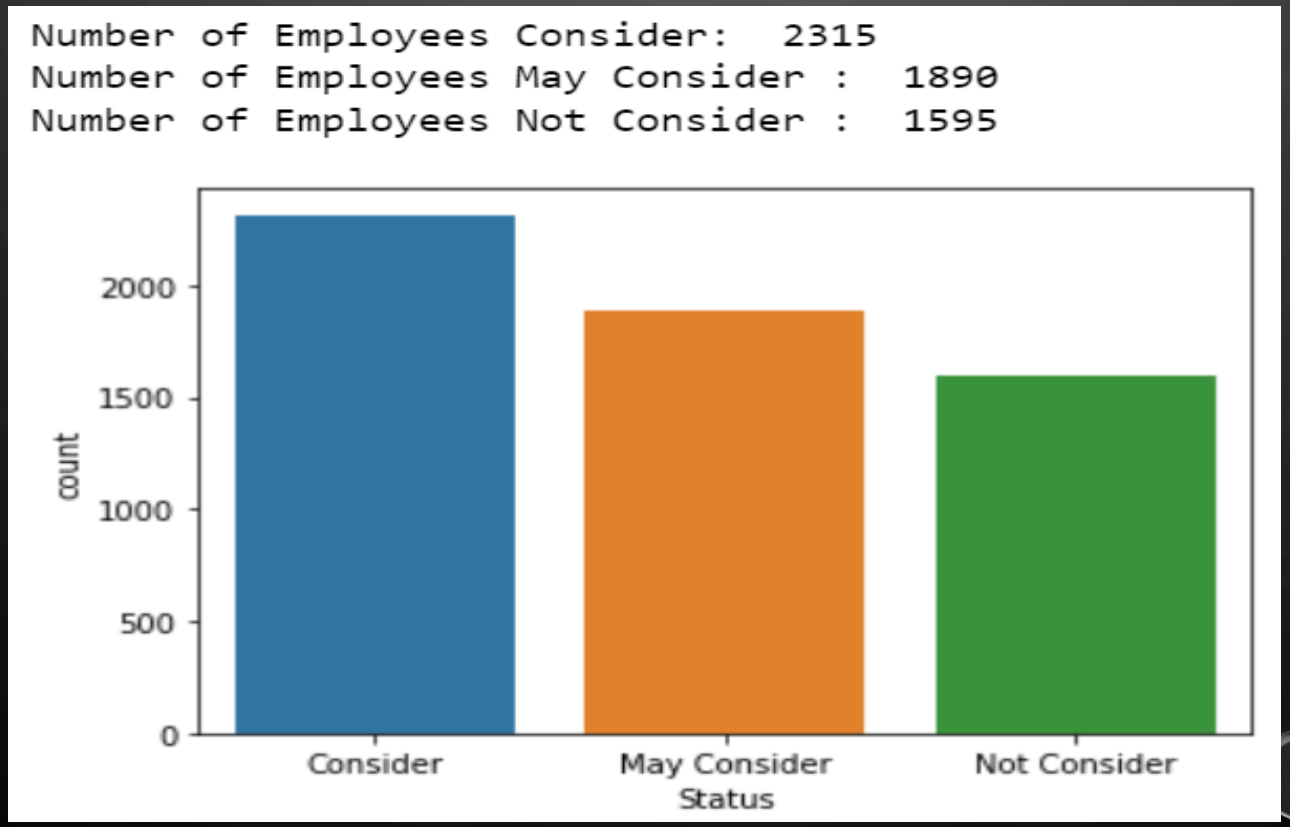
3.ENCODING FOR CATEGORICAL FEATURES

- Label encoding for “ target ” variable
- The features that had nominal data is converted into binary features by doing “ One-Hot Encoding ”.

4.CHECKING WHETHER TARGET DATA IS BALANCED OR NOT

- Here some how target data is balanced, if in case the data is imbalanced we need to do oversampling.

```
Number of Employees Consider:  2315
Number of Employees May Consider :  1890
Number of Employees Not Consider :  1595
```



5. HANDLING OUTLIERS

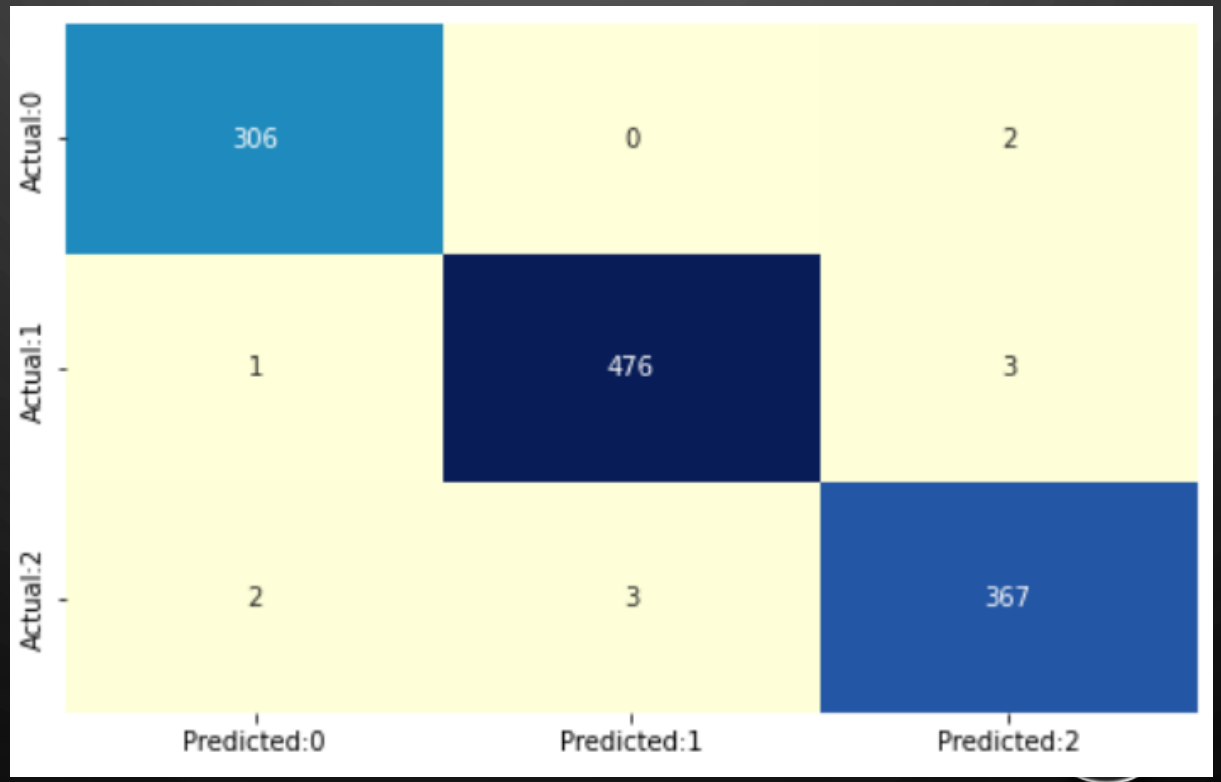
- “ Box-plotting ” is done to check whether outliers are present or not.
- If found remove the outliers.

6. FEATURE SCALING

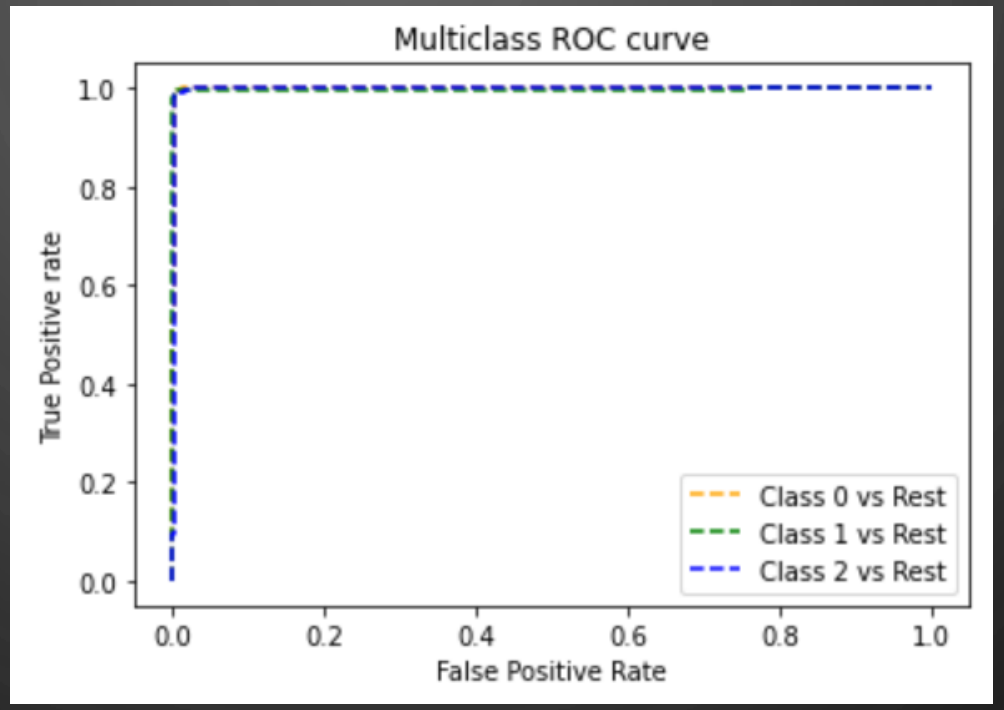
- Standard Scaler is used for to convert the values having high variance into the range between [0 -1].

7.BULIDING MODEL

- For this classification problem I used “ Logistic Regression” method to build the model .
- Then “ Confusion matrix ” is used to know how the model has performed.



- The a “ ROC – Curve ” is plotted to show graphically the probability of the outcome.



- Finally the accuracy of the model can be known using “ Classification report ”.

THANK YOU