**RESULTS**

Count of negative and posititve values in Train Data:

['neg', 'pos']

[19666 333]

Count of negative and posititve values in Test Data:

['neg', 'pos']

[15625 375]

Training 1 logistic regression Started

Training 1 logistic regression Complete

Accuracy score for LR prediction on given data: 0.9884375

F1 score for LR prediction on given data: 0.734576757532281

Confusion matrix for LR prediction on given data:

[[15559 66]

[ 119 256]]

Accuracy score for SVM prediction on given data: 0.9765625

Training 1 Gaussion NB Started

Training 1 Gaussion NB Complete

Accuracy score for Gaussion NB prediction on given data: 0.96975

F1 score for Gaussion NB prediction on given data: 0.5798611111111112

Confusion matrix for Gaussion NB prediction on given data:

[[15182 443]

[ 41 334]]

Training 1 KNN Started

Training 1 KNN Complete

Accuracy score for KNN prediction on given data: 0.9813125

F1 score for KNN prediction on given data: 0.4432029795158287

Confusion matrix for KNN prediction on given data:

[[15582 43]

[ 256 119]]

Training 1 ANN Started

Training 1 ANN Complete

Accuracy score for ANN prediction on given data: 0.980125

F1 score for ANN prediction on given data: 0.4218181818181818

Confusion matrix for ANN prediction on given data:

[[15566 59]

[ 259 116]]

PCA: 85

TEST\_PCA: 85

Training 2 logistic regression Started

Training 2 logistic regression Complete

Accuracy score for LR prediction on normalized and PCA data: 0.9769375

F1 score for LR prediction on normalized and PCA data: 0.44677661169415295

Confusion matrix for LR prediction on normalized and PCA data:

[[15482 143]

[ 226 149]]

Training 2 SVM Started

Training 2 SVM Complete

Accuracy score for SVM prediction on normalized and PCA data: 0.9765625

Training 2 Gaussion NB Started

Training 2 Gaussion NB Complete

Accuracy score for Gaussion NB prediction on normalized and PCA data: 0.9655625

F1 score for Gaussion NB prediction on normalized and PCA data: 0.527038626609442

Confusion matrix for Gaussion NB prediction on normalized and PCA data:

[[15142 483]

[ 68 307]]

Training 2 KNN Started

Training 2 KNN Complete

Accuracy score for KNN prediction on normalized and PCA data: 0.97775

F1 score for KNN prediction on normalized and PCA data: 0.2192982456140351

Confusion matrix for KNN prediction on normalized and PCA data:

[[15594 31]

[ 325 50]]

Training 2 ANN Started

Training 2 ANN Complete

Accuracy score for ANN prediction on normalized and PCA data: 0.980125

F1 score for ANN prediction on normalized and PCA data: 0.4218181818181818

Confusion matrix for ANN prediction on normalized and PCA data:

[[ 5381 10244]

[ 84 291]]

Validation split

Number of data samples in Training: 15999

Number of data samples in Validation: 4000

Number of positive and negative indices in data = 267 15732

After under sampling:

Percentage Neg after under sampling: 0.5

Percentage Pos after under sampling: 0.5

Total number of data points: 534

Training 3 logistic regression Started

Training 3 logistic regression Completed

Accuracy score for LR prediction on Undersampled, normalized and PCA data: 0.860875

F1 score for LR prediction on Undersampled, normalized and PCA data: 0.24285714285714285

Confusion matrix for LR prediction on Undersampled, normalized and PCA data:

[[13417 2208]

[ 18 357]]

Training 3 SVM Started

Training 3 SVM Complete

Accuracy score for SVM prediction on Undersampled, normalized and PCA data: 0.946125

F1 score for SVM prediction on Undersampled, normalized and PCA data: 0.45443037974683537

Confusion matrix for SVM prediction on Undersampled, normalized and PCA data:

[[14779 846]

[ 16 359]]

Training 3 Gaussion NB Started

Training 3 Gaussion NB Complete

Accuracy score for Gaussion NB prediction on Undersampled, normalized and PCA data: 0.963625

F1 score for Gaussion NB prediction Undersampled, normalized and PCA data: 0.5182119205298014

Confusion matrix for Gaussion NB prediction on Undersampled, normalized and PCA data:

[[15105 520]

[ 62 313]]

Training 3 KNN Started

Training 3 KNN Complete

Accuracy score for KNN prediction on Undersampled, normalized and PCA data: 0.976625

F1 score for KNN prediction on Undersampled, normalized and PCA data: 0.6120331950207468

Confusion matrix for KNN prediction on Undersampled, normalized and PCA data:

[[15331 294]

[ 80 295]]

Training 3 ANN Started

Training 3 ANN Complete

Accuracy score for ANN prediction on Undersampled, normalized and PCA data: 0.9333125

F1 score for ANN prediction on Undersampled, normalized and PCA data: 0.3474006116207951

Confusion matrix for ANN prediction on Undersampled, normalized and PCA data:

[[14649 976]

[ 91 284]]

Over Sampled Data Size: 27536

Pos Count: 13768

Negative in Oversampled data set: 13768

Ratio of positive in oversampled data: 0.5

Ratio of negative in oversampled data: 0.5

(27536, 85) (27536, 1)

Training 4 logistic regression Started

Training 4 logistic regression Complete

Accuracy score for LR prediction on Oversampled, normalized and PCA data: 0.9525

F1 score for LR prediction on Oversampled, normalized and PCA data: 0.3436960276338515

Confusion matrix for LR prediction on Oversampled, normalized and PCA data:

[[13417 2208]

[ 18 357]]

Training 4 SVM Started

Training 4 SVM Complete

Accuracy score for SVM prediction on Oversampled, normalized and PCA data: 0.9508125

F1 score for SVM prediction on Oversampled, normalized and PCA data: 0.45536332179930794

Confusion matrix for SVM prediction on Oversampled, normalized and PCA data:

[[14884 741]

[ 46 329]]

Training 4 Gaussion NB Started

Training 4 Gaussion NB Complete

Accuracy score for Gaussion NB prediction on Oversampled, normalized and PCA data: 0.962625

F1 score for Gaussion NB prediction on Oversampled, normalized and PCA data: 0.5106382978723405

Confusion matrix for Gaussion NB prediction on Oversampled, normalized and PCA data:

[[15090 535]

[ 63 312]]

Training 4 KNN Started

Training 4 KNN Complete

Accuracy score for KNN prediction on Oversampled, normalized and PCA data: 0.95925

F1 score for KNN prediction on Oversampled, normalized and PCA data: 0.44082332761578047

Confusion matrix for KNN prediction on Oversampled, normalized and PCA data:

[[15091 534]

[ 118 257]]

Training 4 ANN Started

Training 4 ANN Complete

Accuracy score for ANN prediction on Oversampled, normalized and PCA data: 0.94675

F1 score for ANN prediction on Oversampled, normalized and PCA data: 0.32594936708860756

Confusion matrix for ANN prediction on Oversampled, normalized and PCA data:

[[14942 683]

[ 169 206]]