Machine Learning Project on Gender Recognition by Voice

INSAID: GCD: Term-4 Project by S.R.Uthayanan

Problem Statement

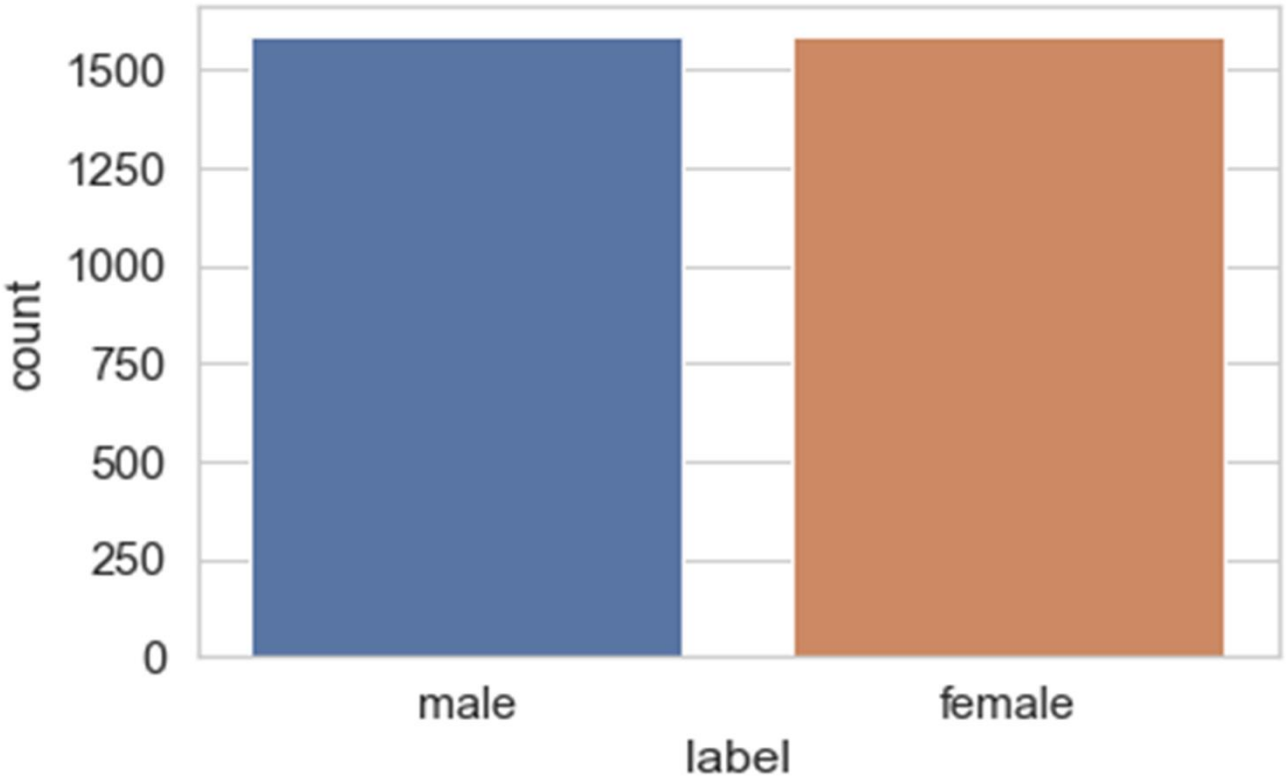
Gender Recognition by Voice and Speech Analysis

This Project aim is to identify a voice as male or female, based upon acoustic properties of the voice and speech.

Description of Dataset

Dataset info:	
Number of variables:	21
Number of observations:	3168
Variables types:	
Numeric :	20
Categorical :	1

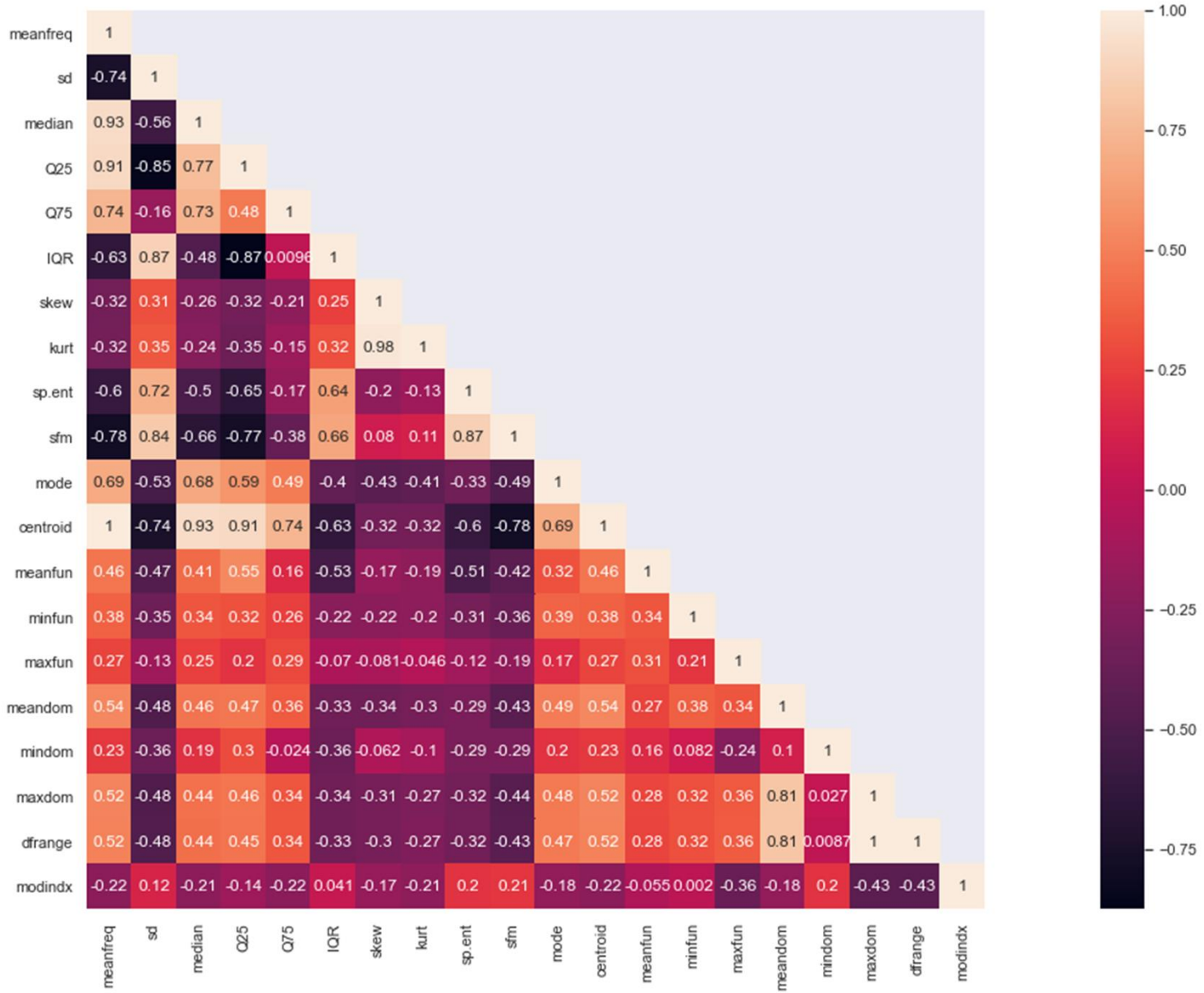
Balanced Dataset



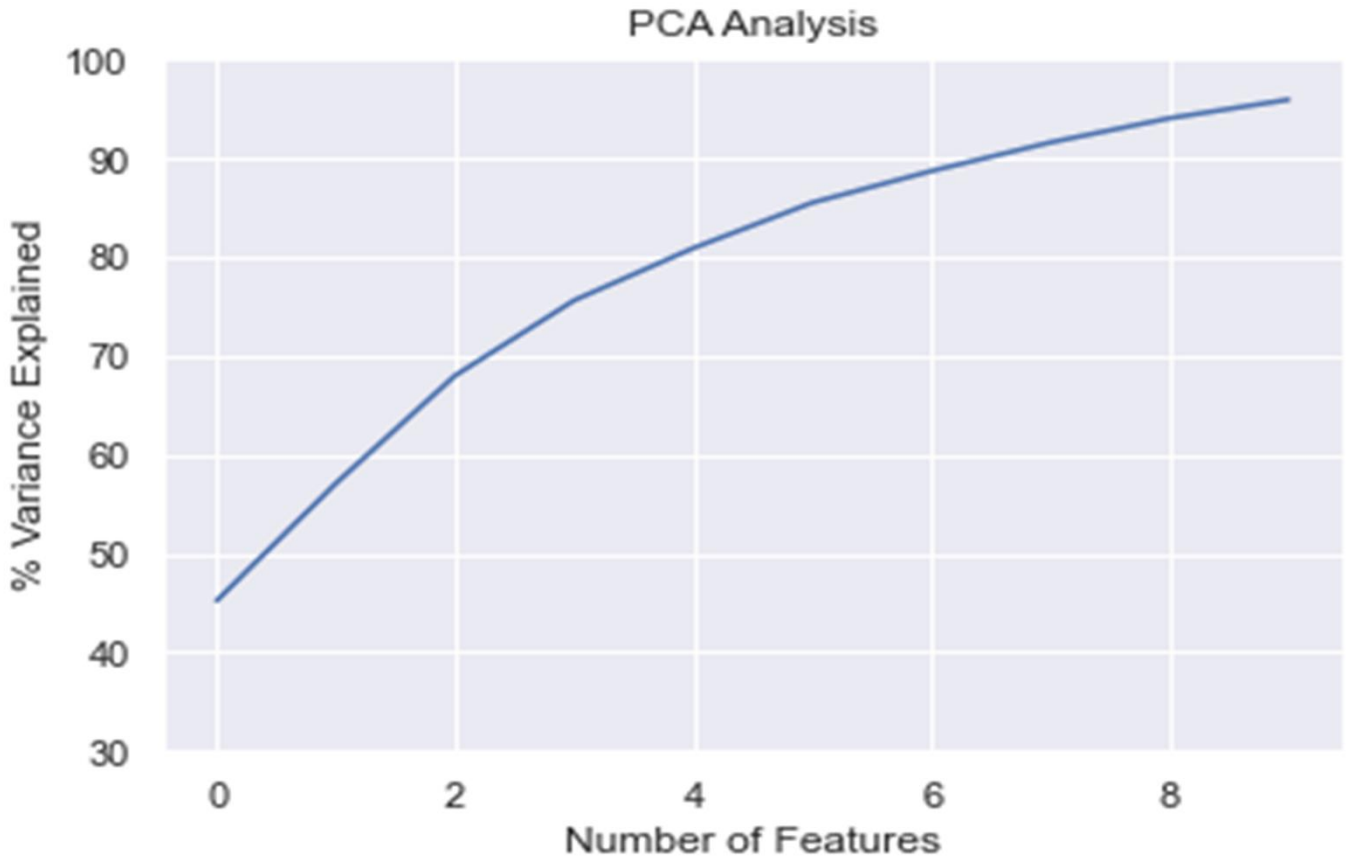
We have equal no of observations for the 'males' and the 'females'.

Hence it is a balanced dataset

Correlations between Features



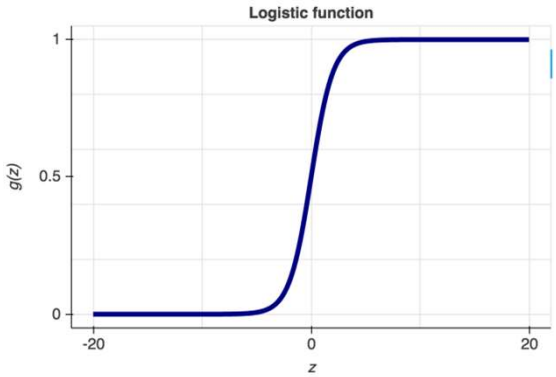
Dimensionality Reduction



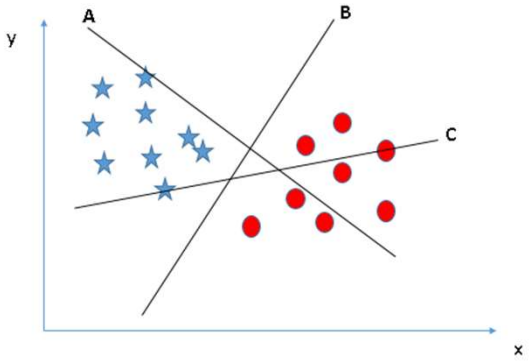
10 components explain 95% of the variation in data

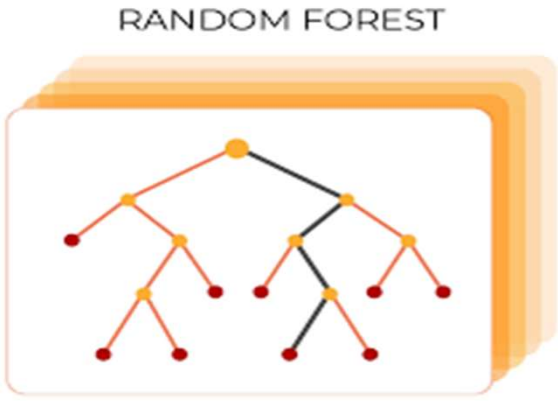
Classification Algorithms used for Modelling

Logistic Regression



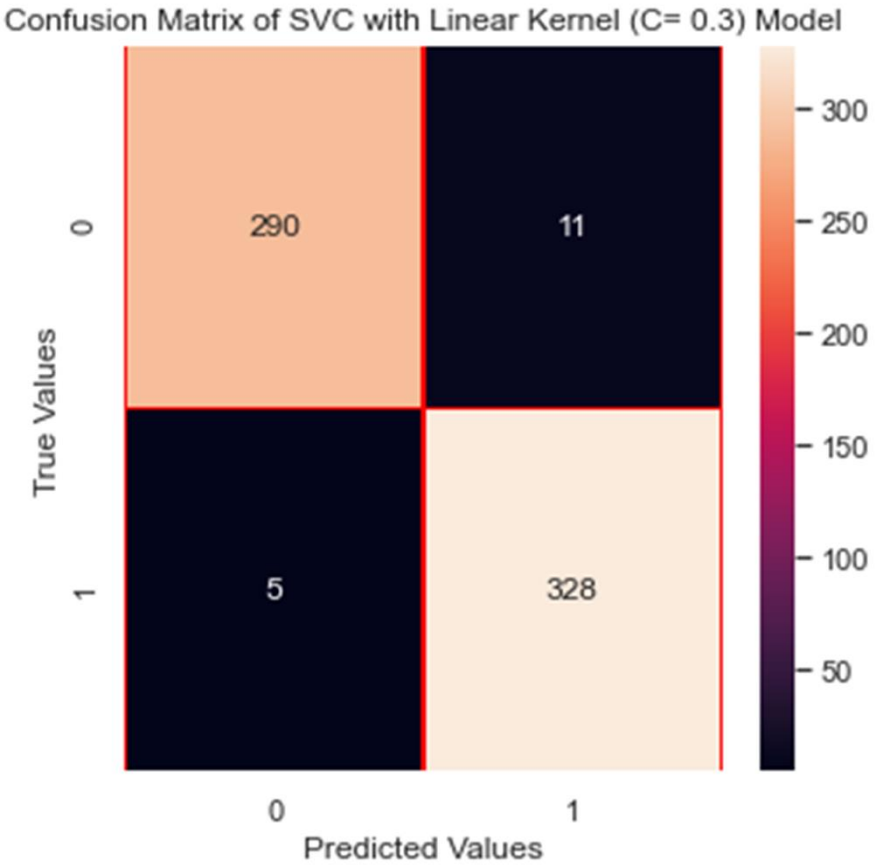
Support Vector Machines





Models are evaluated using Accuracy Score

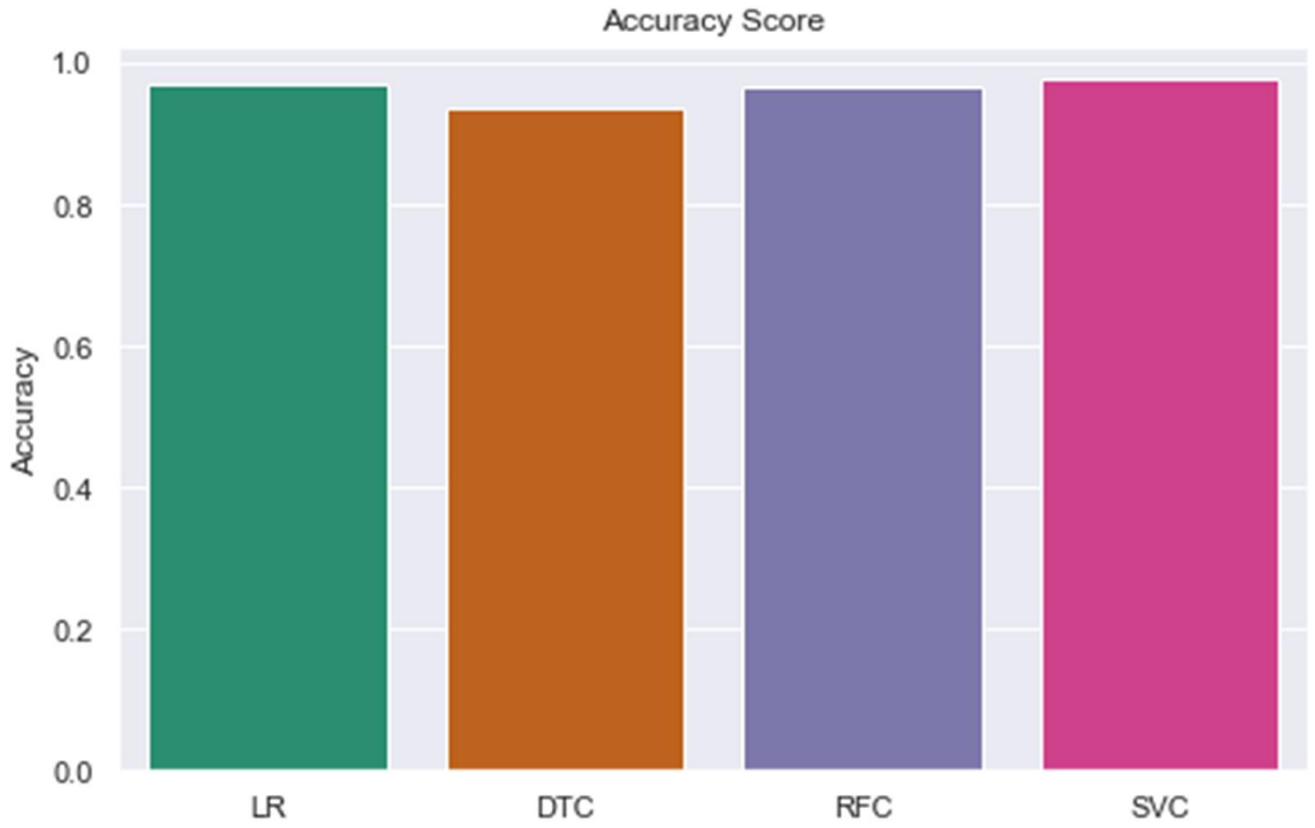
Model Evaluation



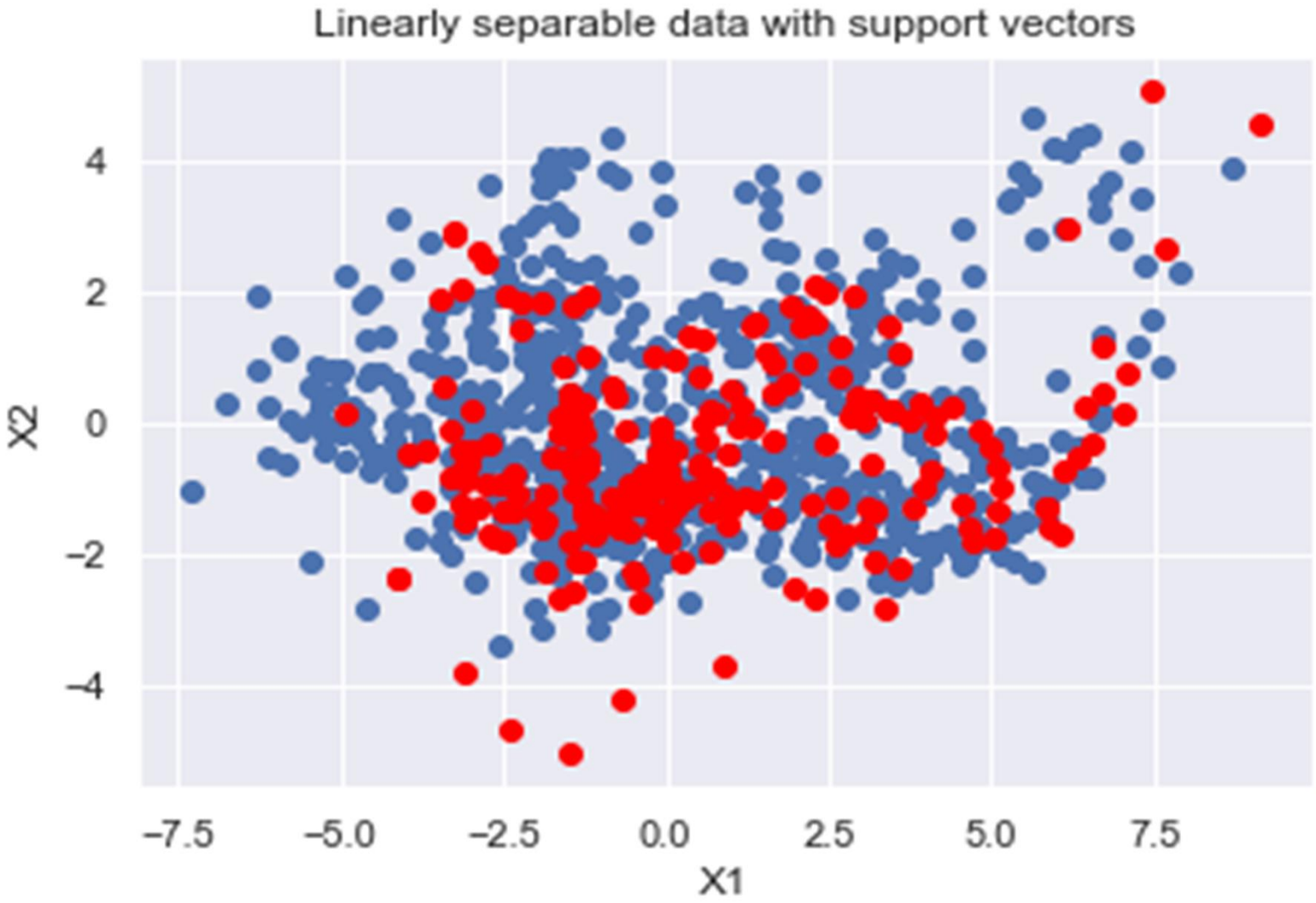
Comparison of Performance Parameters

Algorithms	Accuracy Score	Precision Score	Recall Score	F1_Score
LR	0.968454	0.956268	0.984985	0.970414
DTC	0.933754	0.939577	0.933934	0.936747
RFC	0.963722	0.958580	0.972973	0.965723
RFC_CV	0.951104	0.944118	0.963964	0.953938
SVC_Linear	**0.974777**	0.967552	0.984985	0.976190
SVC_Rbf	0.971609	0.970149	0.975976	0.973054
SVC_Poly	0.906796	0.967552	0.975976	0.964497

Accuracy Score



Plotting Predictions



Conclusion

- The main contribution is to find out the best algorithm for the classification Gender from the datasets.
- We have compared the performance of these models using Accuracy, Recall, Precision and F1 Score.
- The result points out that the algorithm with the highest accuracy is SVM with Linear Kernel(C=0.3). So the best algorithm for Voice Based Gender Classification is SVM with Linear Kernel.

Thank You