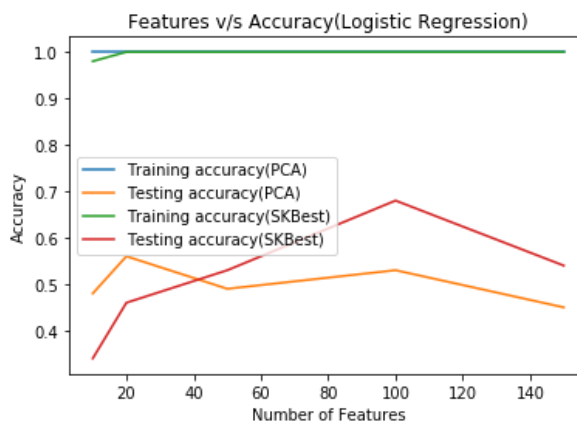
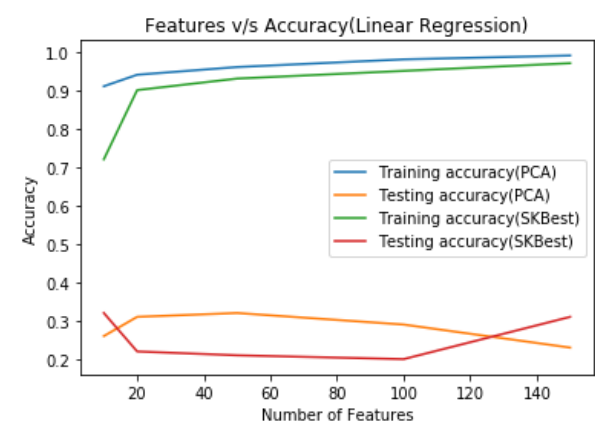
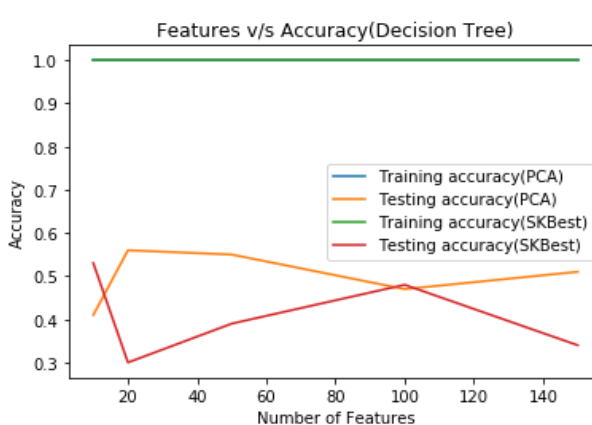
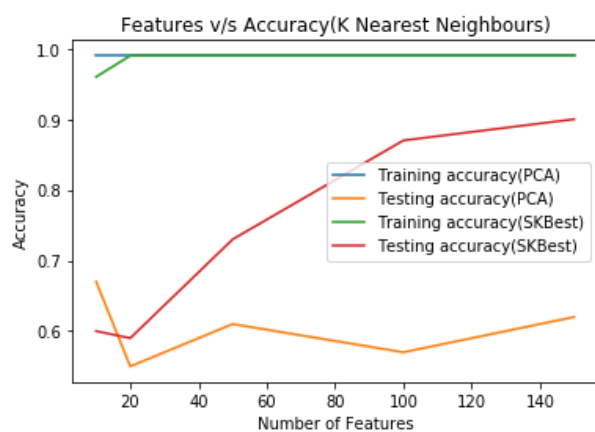
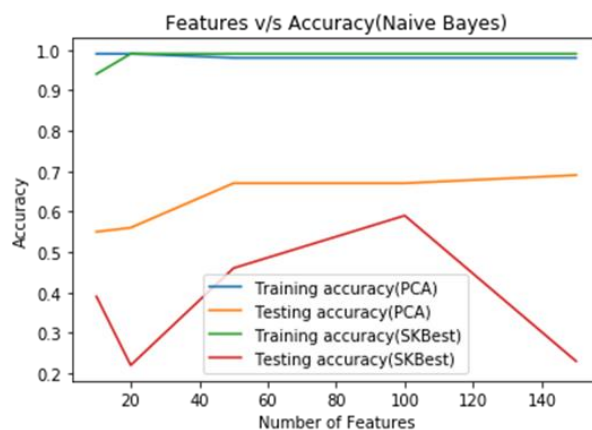
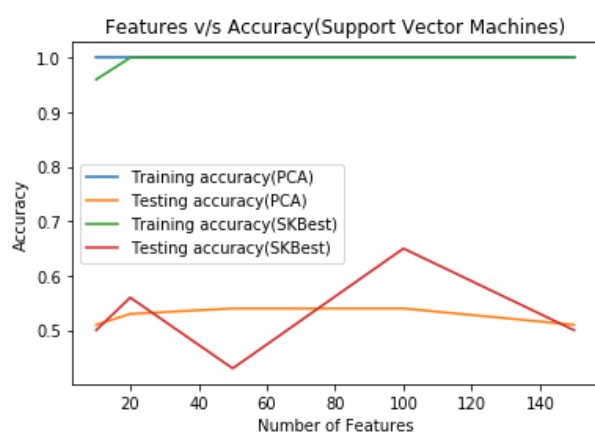
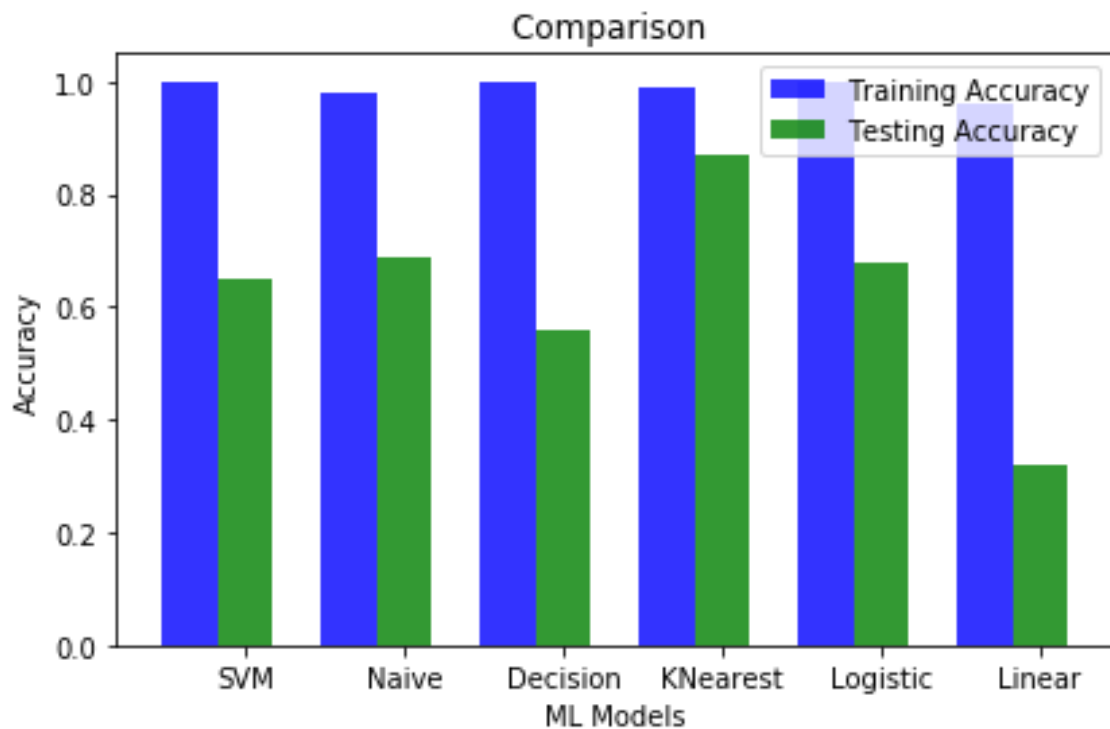


# Feature Analysis

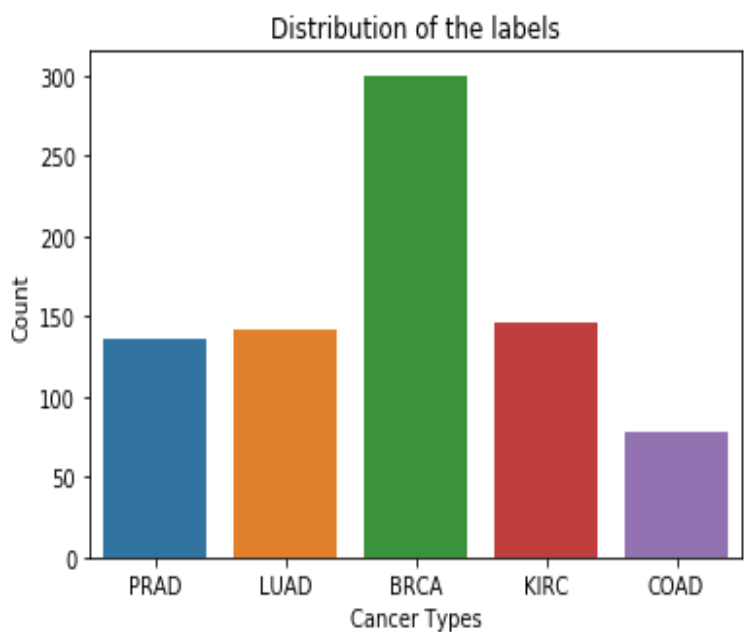
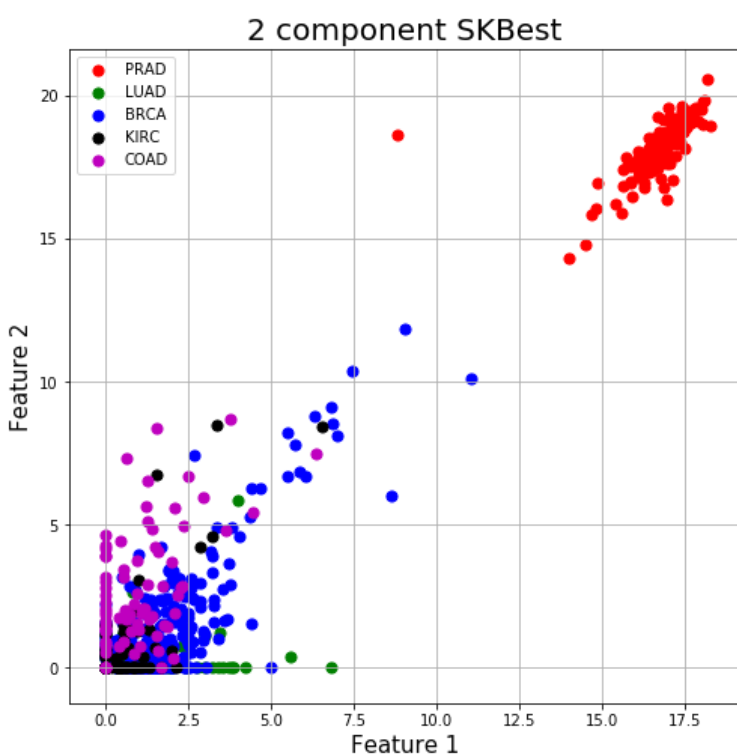
- For each machine learning model, analysis of training and testing accuracy with respect to different feature set.



- Comparison among different models by considering the best feature set for each model.



- Distribution of initial data in 2 dimensions



➤ **Approaches tried to overcome overfitting**

- Feature selection methods- LDA, PCA and Annova (with 10-200 features)
- Different classifiers- SVM, Naïve Bayes, Decision Tree and K Nearest Neighbours (with k=5)
- Cross validation using 3,5,10 folds
- GridSearchCV for SVM
- Normalisation of dataset
- Regularisation
- Manually creating a balanced subset using original dataset