Implementation and result:

To demonstrate the effectiveness of the used algorithms in predicting Parkinson’s disease effectively the non-biased speech dataset has been trained on different models and provided us with some significant results. In the following segment, the comparative study and analysis of results provided by different Machine Learning models are presented. Note that all models are trained after pre-processing and preparing the data.

KNN classifier:

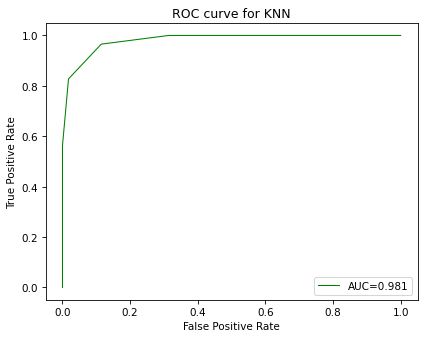
A knn model is implemented with the default number of neighbours i.e. 3. The model provides the following outputs:

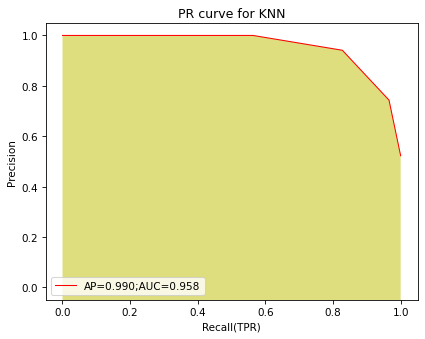
Confusion matrix:

|  |  |  |
| --- | --- | --- |
| Actual | True | False |
| Predicted |
| True | 16 | 2 |
| False | 3 | 55 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| 0 | 0.84 | 0.89 | 0.86 | 18 |
| 1 | 0.96 | 0.95 | 0.96 | 58 |

The accuracy of the model is = 93.42105263157895%





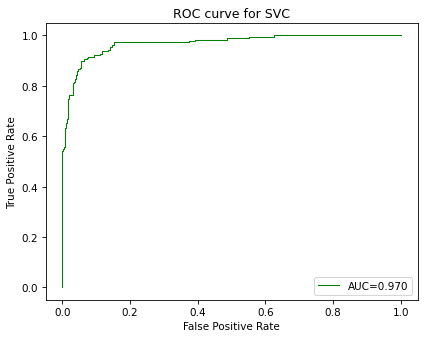
Support Vector Classifier:  
The support vector model is trained on the processed data with c=1.0 and kernel = “rbf” which are the default values for the parameters. The results obtained are as follows:

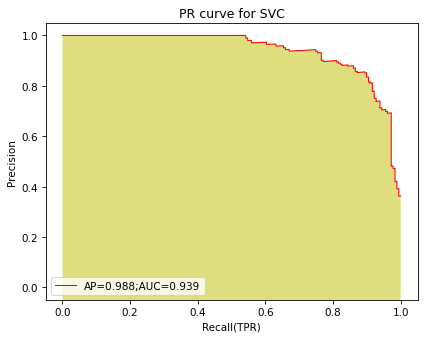
Confusion matrix:

|  |  |  |
| --- | --- | --- |
| Actual | True | False |
| Predicted |
| True | 7 | 6 |
| False | 2 | 61 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| 0 | 0.78 | 0.54 | 0.64 | 13 |
| 1 | 0.91 | 0.97 | 0.94 | 63 |

The accuracy of the model is = 89.47368421052632%





Decision Tree Classifier:

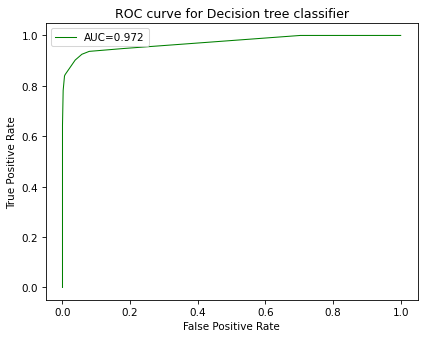
Working on the decision-making principle the decision tree classifier is trained with criterion= “gini” and max\_depth=7 which produced the following results:

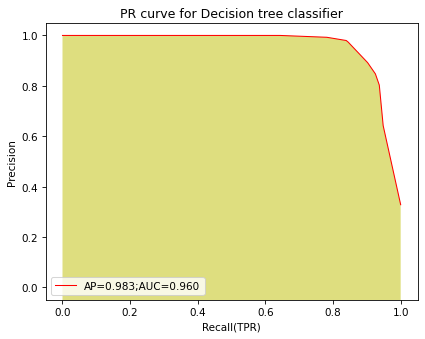
Confusion matrix:

|  |  |  |
| --- | --- | --- |
| Actual | True | False |
| Predicted |
| True | 12 | 7 |
| False | 4 | 53 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| 0 | 0.75 | 0.63 | 0.69 | 19 |
| 1 | 0.88 | 0.93 | 0.91 | 57 |

The accuracy of the model is = 85.52631578947368%





Random Forest classifier:

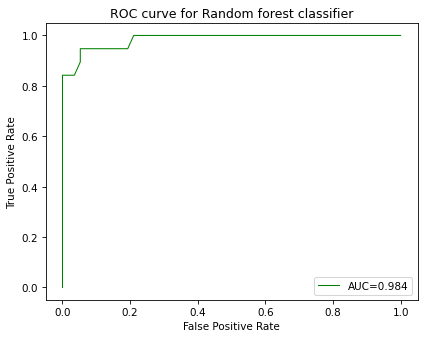
The Random forest classifier is trained using the default parameters i.e. n\_estimators=100 and criterion = “gini”. The processed data were then fit into the model and the following results were obtained:

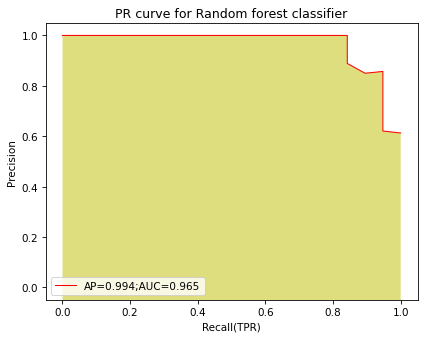
Confusion matrix:

|  |  |  |
| --- | --- | --- |
| Actual | True | False |
| Predicted |
| True | 12 | 7 |
| False | 0 | 57 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| 0 | 1 | 0.63 | 0.77 | 19 |
| 1 | 0.89 | 1 | 0.94 | 57 |

The accuracy of the model is = 90.78947368421053%





Naïve Bayes Classifier:

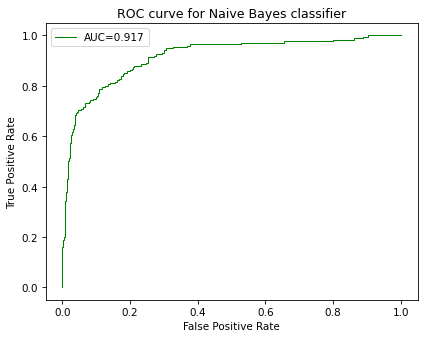
The probabilistic algorithm here using the Gaussian Naïve Bayes technique is trained using the training sample on default parameters producing the following results:

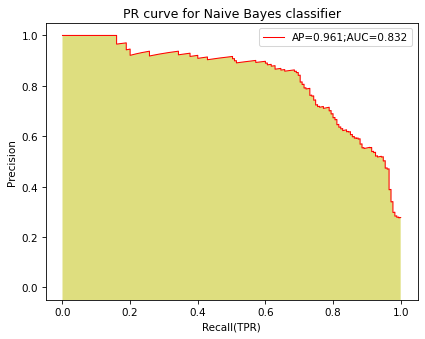
Confusion matrix:

|  |  |  |
| --- | --- | --- |
| Actual | True | False |
| Predicted |
| True | 12 | 5 |
| False | 3 | 56 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| 0 | 0.80 | 0.71 | 0.75 | 17 |
| 1 | 0.92 | 0.95 | 0.93 | 59 |

The accuracy of the model is = 89.47368421052632%





Logistic regression:

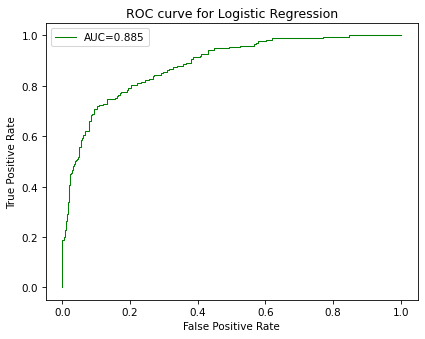
Logistic Regression is used for the classification problem where the penalty function used is the L2 norm. The results thus obtained are as follows:

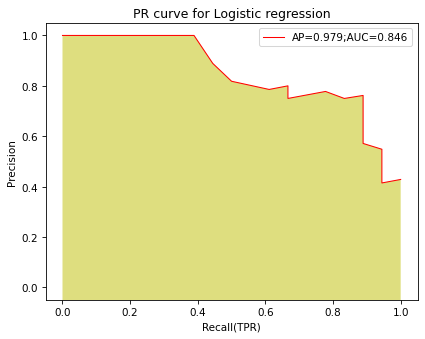
Confusion matrix:

|  |  |  |
| --- | --- | --- |
| Actual | True | False |
| Predicted |
| True | 8 | 10 |
| False | 10 | 48 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| 0 | 0.44 | 0.44 | 0.44 | 18 |
| 1 | 0.83 | 0.83 | 0.83 | 58 |

The accuracy of the model is = 85.52631578947368%





XGBoost classifier:

93.42105263157895

