**MODEL COMPARISON**

**Note:** The original data set has not been split up into training and testing sets because of the limited number of available data points

* Data Visualisation has been accomplished using Principal Component Analysis to first find the data in the direction of maximum Variances and the plotting those two dimensions.
* It is totally possible to do that for 3 Dimensions as well, but not for 4 dimensions
* The structure/shape of data in these two dimensions/score space(s) looks as is shown below:
* Chart, scatter chart

  Description automatically generated
* The different colors here show the data points that belong to different classes, signifying this is a multi-class classification problem

**COMAPRISONS:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classifier** | **Training Time** | **Testing Time** | **F1 Score** | **Accuracy** |
| K Nearest Neighbors | 0.60402512 | 0.00256466 | [1., 0.95, 0.94] | 96.67% |
| Gaussian Naïve Bayes | 0.01688599 | 0.00028204 | [1., 0.94, 0.94] | 96.0% |
| SVM (Linear) | 0.09331822 | 0.00018405 | [1., 0.96, 0.97] | 98.0% |
| SVM (RBF) | 0.37127685 | 0.00092315 | [1., 0.96, 0.97] | 98.0% |

**CONFUSION MATRICES:**

**K Nearest Neighbors:**

[[50 0 0]

[ 0 49 1]

[ 0 4 46]]

**Gaussian Naïve Bayes:**

[[50 0 0]

[ 0 47 3]

[ 0 3 47]]

**Support Vector Machine:**

[[50 0 0]

[ 0 48 2]

[ 0 1 49]]

1. It is quite clear from the data represented above that the better model is SVM’s, it doesn’t matter if its RBF or Linear because both perform at the same level
2. But it can also be seen that these perform comparatively slower when compared to Gaussian Naïve Bayes and KNN which partly has to do with the Quadratic optimization problem in SVM
3. Now these differences are not that discernible owing to smaller data sets, but these might change a lot when the size of data set is significant and composition complex
4. Time taken is almost definitely going to increase a lot in case of SVM because the QOP because that much more difficult and time consuming to solve
5. Confusion Matrix and F1 score also paint the same picture as we have been able to see from the plot ourselves, one of the classes is clearly much more separable and hence the value on off diagonals is zero for that class in case of every single classifier/value of first index is 1.0 in each F1 score
6. The other two classes do have a higher overlap and hence could not be separated completely and hence have almost but not perfect F1 scores/Confusion Matrices
7. Since the error rates on SVM are the lowest that’s why it only makes sense for it to have the total number of false calls on the off diagonals of the confusion matrix to be the lowest of all and same foes for the F1 score as well