**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]]