A REPORT ON FEATURED EXTRACTED WITH PRINCIPAL COMPONENT ANALYSIS

**Principal component analysis** (**PCA**) is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. The principal component used are 2, 3, 4, 5, 6, 7, and 8 these featured extracted were classified by the following classifier; Linear Support Vector Classifier, Support Vector Classifier, Logistic Regression, Random Forest Classifier, AdaBoost Classifier and Gradient Boosting Classifier.

Principal Component of 2 Accuracy and Receiver Operating Characteristic (ROC)

Table 1: Accuracy of each classifier with 2 principal components.

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset accuracy | Test dataset accuracy |
| LinearSVC | 0.5053 | 0.5 |
| SVC | 0.5285 | 0.4875 |
| Logistic Regression | 0.5089 | 0.5125 |
| Random Forest | 1 | 0.45 |
| AdaBoost Classifier | 0.6303 | 0.475 |
| Gradient Boost Classifier | 0.841 | 0.45 |

Table 2: Receiver Operating Characteristic (ROC)

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset AUC score | Test dataset AUC score |
| LinearSVC | - | - |
| SVC | 0.47 | 0.49 |
| Logistic Regression | 0.51 | 0.50 |
| Random Forest | 1 | 0.45 |
| AdaBoost Classifier | 0.71 | 0.45 |
| Gradient Boost Classifier | 0.94 | 0.47 |

|  |  |
| --- | --- |
|  |  |

Figure 1: ROC curve of Support vector classifier. Figure 2: ROC curve of Logistic regression

|  |  |
| --- | --- |
| Figure 3: ROC curve of random forest | Figure 4: ROC curve of AdaBoost Classifier |

Principal Component of 3 Accuracy and Receiver Operating Characteristic (ROC)

Table 3: Accuracy of each classifier with 3 principal components.

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset accuracy | Test dataset accuracy |
| LinearSVC | 0.4928 | 0.4937 |
| SVC | 0.5232 | 0.506 |
| Logistic Regression | 0.5071 | 0.4812 |
| Random Forest | 1 | 0.5687 |
| AdaBoost Classifier | 0.6678 | 0.5125 |
| Gradient Boost Classifier | 0.8732 | 0.4562 |

Table 4: Receiver Operating Characteristic (ROC)

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset AUC score | Test dataset AUC score |
| LinearSVC | - | - |
| SVC | 0.44 | 0.49 |
| Logistic Regression | 0.50 | 0.49 |
| Random Forest | 1 | 0.53 |
| AdaBoost Classifier | 0.73 | 0.49 |
| Gradient Boost Classifier | 0.96 | 0.46 |

|  |  |
| --- | --- |
| Figure 5: ROC curve of SVC | Figure 6: ROC curve for Logistic Regression |
| Figure 7: ROC curve Random Forest | Figure 8: ROC curve AdaBoost classifier |

Principal Component of 4 Accuracy and Receiver Operating Characteristic (ROC)

Table 5: Accuracy of each classifier with 4 principal components.

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset accuracy | Test dataset accuracy |
| LinearSVC | 0.4964 | 0.4875 |
| SVC | 0.5375 | 0.49375 |
| Logistic Regression | 0.4785 | 0.475 |
| Random Forest | 1 | 0.4687 |
| AdaBoost Classifier | 0.6625 | 0.50625 |
| Gradient Boost Classifier | 0.8660 | 0.4625 |

Table 6: Receiver Operating Characteristic (ROC)

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset AUC score | Test dataset AUC score |
| LinearSVC | - | - |
| SVC | 0.42 | 0.49 |
| Logistic Regression | 0.50 | 0.50 |
| Random Forest | 1 | 0.47 |
| AdaBoost Classifier | 0.74 | 0.49 |
| Gradient Boost Classifier | 0.95 | 0.49 |

Principal Component of 5 Accuracy and Receiver Operating Characteristic (ROC)

Table 7: Accuracy of each classifier with 5 principal components.

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset accuracy | Test dataset accuracy |
| LinearSVC | 0.5053 | 0.4937 |
| SVC | 0.5500 | 0.5062 |
| Logistic Regression | 0.4964 | 0.475 |
| Random Forest | 1 | 0.4562 |
| AdaBoost Classifier | 0.6696 | 0.5062 |
| Gradient Boost Classifier | 0.8607 | 0.4437 |

Table 8: Receiver Operating Characteristic (ROC)

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset AUC score | Test dataset AUC score |
| LinearSVC | - | - |
| SVC | 0.42 | 0.49 |
| Logistic Regression | 0.51 | 0.50 |
| Random Forest | 1 | 0.44 |
| AdaBoost Classifier | 0.75 | 0.48 |
| Gradient Boost Classifier | 0.95 | 0.48 |

Principal Component of 6 Accuracy and Receiver Operating Characteristic (ROC)

Table 9: Accuracy of each classifier with 6 principal components.

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset accuracy | Test dataset accuracy |
| LinearSVC | 0.5053 | 0.525 |
| SVC | 0.5696 | 0.5000 |
| Logistic Regression | 0.4982 | 0.5062 |
| Random Forest | 1 | 0.4500 |
| AdaBoost Classifier | 0.6839 | 0.5062 |
| Gradient Boost Classifier | 0.8714 | 0.4812 |

Table 10: Receiver Operating Characteristic (ROC)

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset AUC score | Test dataset AUC score |
| LinearSVC | - | - |
| SVC | 0.39 | 0.49 |
| Logistic Regression | 0.52 | 0.49 |
| Random Forest | 1 | 0.47 |
| AdaBoost Classifier | 0.76 | 0.49 |
| Gradient Boost Classifier | 0.96 | 0.46 |

Principal Component of 7 Accuracy and Receiver Operating Characteristic (ROC)

Table 11: Accuracy of each classifier with 7 principal components.

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset accuracy | Test dataset accuracy |
| LinearSVC | 0.5142 | 0.4812 |
| SVC | 0.5714 | 0.5062 |
| Logistic Regression | 0.5107 | 0.4875 |
| Random Forest | 1 | 0.4500 |
| AdaBoost Classifier | 0.7000 | 0.475 |
| Gradient Boost Classifier | 0.9339 | 0.4687 |

Table 12: Receiver Operating Characteristic (ROC)

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset AUC score | Test dataset AUC score |
| LinearSVC | - | - |
| SVC | 0.39 | 0.49 |
| Logistic Regression | 0.53 | 0.50 |
| Random Forest | 1 | 0.46 |
| AdaBoost Classifier | 0.78 | 0.49 |
| Gradient Boost Classifier | 0.99 | 0.46 |

Principal Component of 8 Accuracy and Receiver Operating Characteristic (ROC)

Table 13: Accuracy of each classifier with 8 principal components.

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset accuracy | Test dataset accuracy |
| LinearSVC | 0.5214 | 0.4937 |
| SVC | 0.5732 | 0.5125 |
| Logistic Regression | 0.5035 | 0.4937 |
| Random Forest | 1 | 0.4312 |
| AdaBoost Classifier | 0.7053 | 0.4875 |
| Gradient Boost Classifier | 0.9071 | 0.4750 |

Table 14: Receiver Operating Characteristic (ROC)

|  |  |  |
| --- | --- | --- |
| Classifier | Train dataset AUC score | Test dataset AUC score |
| LinearSVC | - | - |
| SVC | 0.39 | 0.49 |
| Logistic Regression | 0.53 | 0.50 |
| Random Forest | 1 | 0.44 |
| AdaBoost Classifier | 0.78 | 0.49 |
| Gradient Boost Classifier | 0.97 | 0.48 |