Data mining

Assignment 3

Vivek Pandit

Metrics for the various models implemented directly, without caret package

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Knn.3 | Knn.5 | svmLinear | svmRadial | Rforest | Nnet3 | Nnet5 |
| Accuracy | 0.966 | 0.9611 | 0.9507 | 0.9507 | 0.9444 | 0.9514 | 0.9544 |
| Kappa | 0.9546 | 0.9481 | 0.9343 | 0.9343 | 0.9259 | 0.9352 | 0.9497 |
| Sensitivity |  |  |  |  |  |  |  |
| Class1 | 0.9694 | 0.9583 | 0.9444 | 0.9444 | 0.9500 | 0.9500 | 0.9556 0 |
| Class2 | 0.9556 | 0.9611 | 0.9500 | 0.9500 | 0.9389 | 0.9472 0 | 0.9417 |
| Class3 | 0.9583 | 0.9611 | 0.9528 | 0.9528 | 0.9389 | 0.9389 | 0.9528 |
| Class4 | 0.9806 | 0.9639 | 0.9556 | 0.9556 | 0.9500 | 0.9694 | 0.9556 |
| Specificity |  |  |  |  |  |  |  |
| Class1 | 0.9944 | 0.9926 | 0.9870 | 0.9870 | 0.9870 | 0.9852 | 0.9880 |
| Class2 | 0.9843 | 0.9833 | 0.9787 | 0.9787 | 0.9806 | 0.9824 | 0.9824 |
| Class3 | 0.9861 | 0.9806 | 0.9759 | 0.9759 | 0.9713 | 0.9787 | 0.9750 |
| Class4 | 0.9898 | 0.9917 | 0.9926 | 0.9926 | 0.9870 | 0.9889 | 0.9898 |

The model of Knn.3 appears to be best model for the given data because primarily its accuracy is high. Also other values are comparable with other methods.

Metrics after using train control using 10 fold validation

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Knn.3 | Knn.5 | svmLinear | svmRadial | Rforest | Nnet3 | Nnet5 |
| Accuracy | 0.9535 | 0.9535 | 0.9507 | 0.9514 | 1 | 0.9514 | 0.95 |
| Kappa | 0.9381 | 0.938 | 0.9343 | 0.9352 | 1 | 0.9352 | 0.9333 |
| Sensitivity |  |  |  |  |  |  |  |
| Class1 | 0.9528 | 0.9639 | 0.9556 | 0.9472 | 1 | 0.9528 | 0.9667 |
| Class2 | 0.9417 | 0.9389 | 0.9472 | 0.9500 | 1 | 0.9444 | 0.9444 |
| Class3 | 0.9444 | 0.9528 | 0.9500 | 0.9528 | 1 | 0.9472 | 0.9250 |
| Class4 | 0.9750 | 0.9583 | 0.9500 | 0.9556 | 1 | 0.9611 | 0.9639 |
| Specificity |  |  |  |  |  |  |  |
| Class1 | 0.9880 | 0.9870 | 0.9861 | 0.9880 | 1 | 0.9852 | 0.9843 |
| Class2 | 0.9833 | 0.9852 | 0.9815 | 0.9796 | 1 | 0.9833 | 0.9806 |
| Class3 | 0.9796 | 0.9769 | 0.9741 | 0.9750 | 1 | 0.9769 | 0.9815 |
| Class4 | 0.9870 | 0.9889 | 0.9926 | 0.9926 | 1 | 0.9898 | 0.9870 |

Random forest appears to be the best method. This is because it converges efficiently giving rise to high accuracy along with superior classification method.