Iteration 1:

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm name | Sn(%) | Sp(%) | Accuracy(%) |
| nb | 83.2 | 68.06723 | 75.81967 |
| ada | 80.8 | 77.31092 | 79.09836 |
| ht | 82.4 | 68.90756 | 75.81967 |
| rf | 89.6 | 66.38655 | 78.27869 |
| smo | 76 | 74.78992 | 75.40984 |
| bagg | 84.8 | 64.70588 | 75 |

Generate n-grams

Iteration 2:

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm name | Sn(%) | Sp(%) | Accuracy(%) |
| nb | 83.2 | 68.90756 | 76.22951 |
| ada | 82.4 | 74.78992 | 78.68852 |
| ht | 83.2 | 68.90756 | 76.22951 |
| rf | 91.2 | 68.06723 | 79.91803 |
| smo | 81.6 | 81.51261 | 81.55738 |
| bagg | 91.2 | 66.38655 | 79.09836 |

Generate n-grams

Iteration 3:

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm name | Sn(%) | Sp(%) | Accuracy(%) |
| nb | 88 | 69.7479 | 79.09836 |
| ada | 92.8 | 63.02521 | 78.27869 |
| ht | 90.4 | 68.90756 | 79.91803 |
| rf | 92.8 | 68.90756 | 81.14754 |
| smo | 82.4 | 75.63025 | 79.09836 |
| bagg | 88 | 64.70588 | 76.63934 |

Generated more features

Iteration 4:

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm name | Sn(%) | Sp(%) | Accuracy(%) |
| nb | 88 | 69.7479 | 79.09836 |
| ada | 92.8 | 63.02521 | 78.27869 |
| ht | 90.4 | 68.90756 | 79.91803 |
| rf | 91.2 | 67.22689 | 79.5082 |
| smo | 82.4 | 77.31092 | 79.91803 |
| bagg | 88 | 64.70588 | 76.63934 |

Generated most gapped feature

Iteration 5:

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm name | Sn(%) | Sp(%) | Accuracy(%) |
| nb | 91.2 | 75.63025 | 83.60656 |
| ada | 94.4 | 78.15126 | 86.47541 |
| ht | 92.8 | 76.47059 | 84.83607 |
| rf | 94.4 | 76.47059 | 85.65574 |
| smo | 92.8 | 88.23529 | 90.57377 |
| bagg | 93.6 | 71.42857 | 82.78689 |

Select feature using Simple / Logistic function which has positive weights for each feature on positive class.

Put emphasis building decision tree , random tree and using their nodes as feature overriding negative weights on each feature

Iteration 6:

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm name | Sn(%) | Sp(%) | Accuracy(%) |
| nb | 92 | 78.15126 | 85.2459 |
| ada | 96 | 81.51261 | 88.93443 |
| ht | 95.2 | 77.31092 | 86.47541 |
| rf | 92.8 | 73.94958 | 83.60656 |
| smo | 93.6 | 89.91597 | 91.80328 |
| bagg | 88 | 70.58824 | 79.5082 |

Clean dataset by using best first approach of feature selection and removed features which has 0% significance.