

Running KNN on validation set (Using 3 topics)

Methodologies	Accuracy(%) k=1	Accuracy(%) k=3	Accuracy(%) k=5
Hamming distance	42.33	37.5	36.66
Euclidean distance	59.164	55.33	54
Cosine similarity with TF-IDF weights	53	37	34.83

KNN accuracy result for 50 iterations over test set.

From the table above, we can see that best accuracy is obtained using Euclidean distance and k=1. Results for each iteration are -

Iteration 1: 60.0
 Iteration 2: 60.0
 Iteration 3: 76.66666666666667
 Iteration 4: 53.333333333333336
 Iteration 5: 63.33333333333333
 Iteration 6: 60.0
 Iteration 7: 66.66666666666666
 Iteration 8: 60.0
 Iteration 9: 50.0
 Iteration 10: 66.66666666666666
 Iteration 11: 56.666666666666664
 Iteration 12: 66.66666666666666
 Iteration 13: 66.66666666666666
 Iteration 14: 43.333333333333336
 Iteration 15: 76.66666666666667
 Iteration 16: 66.66666666666666
 Iteration 17: 63.33333333333333
 Iteration 18: 70.0
 Iteration 19: 70.0
 Iteration 20: 63.33333333333333
 Iteration 21: 63.33333333333333
 Iteration 22: 66.66666666666666
 Iteration 23: 73.33333333333333
 Iteration 24: 83.33333333333334
 Iteration 25: 83.33333333333334
 Iteration 26: 96.66666666666667
 Iteration 27: 90.0
 Iteration 28: 90.0
 Iteration 29: 76.66666666666667
 Iteration 30: 80.0
 Iteration 31: 93.33333333333333

Iteration 32: 70.0
 Iteration 33: 80.0
 Iteration 34: 83.33333333333334
 Iteration 35: 86.66666666666667
 Iteration 36: 80.0
 Iteration 37: 90.0
 Iteration 38: 93.33333333333333
 Iteration 39: 86.66666666666667
 Iteration 40: 83.33333333333334
 Iteration 41: 80.0
 Iteration 42: 93.33333333333333
 Iteration 43: 86.66666666666667
 Iteration 44: 90.0
 Iteration 45: 93.33333333333333
 Iteration 46: 86.66666666666667
 Iteration 47: 96.66666666666667
 Iteration 48: 80.0
 Iteration 49: 93.33333333333333
 Iteration 50: 86.66666666666667

Running Naive Bayes on validation set (using 3 topics)

Smoothing factor value	Accuracy
0.005	92.83
0.05	92.67
0.1	92.67
0.2	92.33
0.3	92.33
0.5	92.167
0.7	92.167
0.9	92.167
1.0	92.0
2.0	91.83

NB accuracy result for 50 iterations over test set.

From the table above, we can see that best accuracy is obtained using smoothing factor 0.005.

Results for each iteration are -

Iteration 1: 93.33333333333333
Iteration 2: 96.66666666666667
Iteration 3: 100.0
Iteration 4: 90.0
Iteration 5: 100.0
Iteration 6: 90.0
Iteration 7: 96.66666666666667
Iteration 8: 100.0
Iteration 9: 93.33333333333333
Iteration 10: 86.66666666666667
Iteration 11: 90.0
Iteration 12: 93.33333333333333
Iteration 13: 93.33333333333333
Iteration 14: 86.66666666666667
Iteration 15: 90.0
Iteration 16: 100.0
Iteration 17: 96.66666666666667
Iteration 18: 83.33333333333334
Iteration 19: 100.0
Iteration 20: 93.33333333333333
Iteration 21: 90.0
Iteration 22: 90.0
Iteration 23: 100.0
Iteration 24: 73.33333333333333
Iteration 25: 66.66666666666666
Iteration 26: 63.33333333333333
Iteration 27: 66.66666666666666
Iteration 28: 66.66666666666666
Iteration 29: 63.33333333333333
Iteration 30: 66.66666666666666
Iteration 31: 66.66666666666666
Iteration 32: 63.33333333333333
Iteration 33: 66.66666666666666
Iteration 34: 66.66666666666666
Iteration 35: 63.33333333333333
Iteration 36: 66.66666666666666
Iteration 37: 66.66666666666666
Iteration 38: 66.66666666666666
Iteration 39: 66.66666666666666
Iteration 40: 66.66666666666666
Iteration 41: 63.33333333333333
Iteration 42: 66.66666666666666
Iteration 43: 66.66666666666666
Iteration 44: 66.66666666666666
Iteration 45: 66.66666666666666
Iteration 46: 66.66666666666666
Iteration 47: 66.66666666666666
Iteration 48: 66.66666666666666
Iteration 49: 66.66666666666666
Iteration 50: 66.66666666666666

T-test Result analysis

After calculating t-statistic what we get is this-

1. The value of statistic is 0.7906173382529886 and
2. pvalue is 0.4329773247829769.

So, for every significance level(0.005, 0.01 and 0.05) pvalue is greater than significance level. So, for each case we fail to reject the null hypothesis, which is there is no significance difference in two algorithm's accuracy.

Therefore, in this case, none of the two algorithms is performing significantly better than the other one. So, we can choose any of the two algorithms.