Analysis (MT19111)

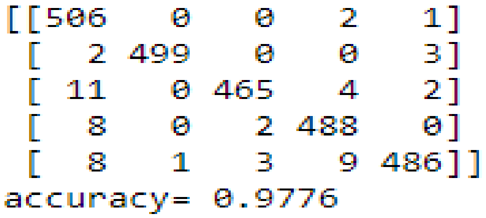
Assign 5.

Note: Here we are splitting data randomnly so according to train and test documents chosen accuracy can be different.

Data Split 50-50:

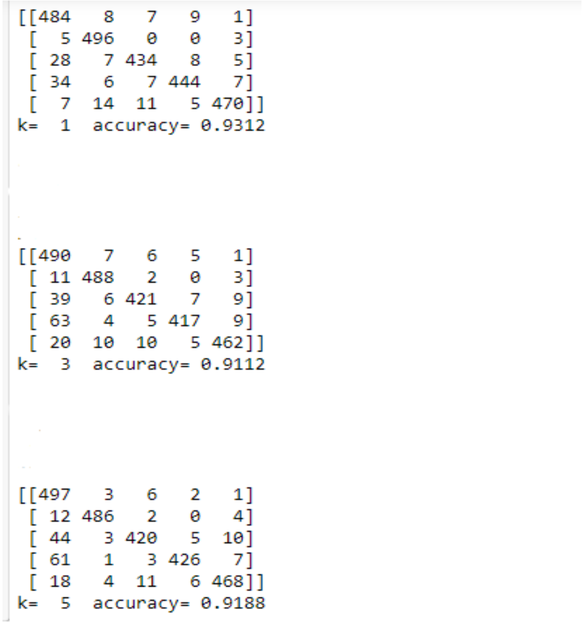
Feature selected with tfidf:

Naïve bayes:



```
[[506   0   0   2   1]
 [  2 499   0   0   3]
 [ 11   0 465   4   2]
 [  8   0   2 488   0]
 [  8   1   3   9 486]]
accuracy= 0.9776
```

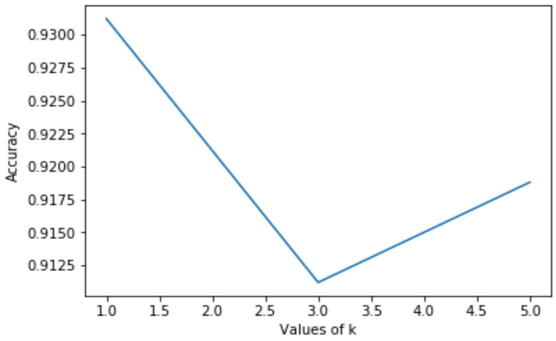
KNN:



```
[[484   8   7   9   1]
 [  5 496   0   0   3]
 [ 28   7 434   8   5]
 [ 34   6   7 444   7]
 [  7  14  11   5 470]]
k=  1  accuracy= 0.9312
```

```
[[490   7   6   5   1]
 [ 11 488   2   0   3]
 [ 39   6 421   7   9]
 [ 63   4   5 417   9]
 [ 20  10  10   5 462]]
k=  3  accuracy= 0.9112
```

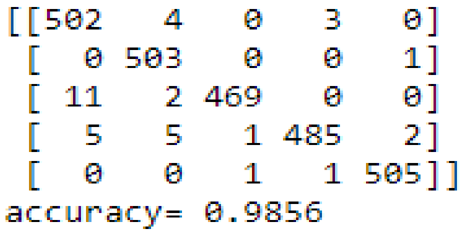
```
[[497   3   6   2   1]
 [ 12 486   2   0   4]
 [ 44   3 420   5  10]
 [ 61   1   3 426   7]
 [ 18   4  11   6 468]]
k=  5  accuracy= 0.9188
```



Here for k=1 accuracy is 0.9312 and k=3 accuracy is 0.9112 which is decreased and k=5 accuracy is 0.9188 which is increased again. (Note: this graph can change for different data splits because it depends on data we selected for training and testing.)

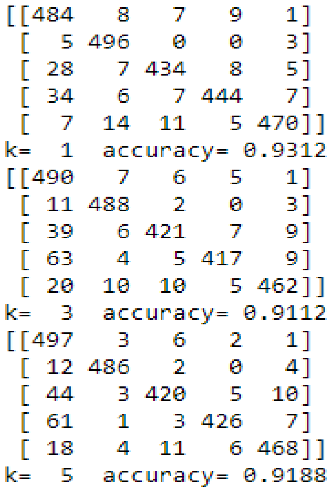
•*Feature selected with MI.*

Naïve Bayes.

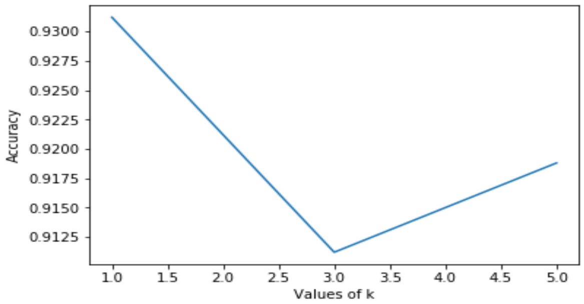


```
[[502   4   0   3   0]
 [  0 503   0   0   1]
 [ 11   2 469   0   0]
 [  5   5   1 485   2]
 [  0   0   1   1 505]]
accuracy= 0.9856
```

KNN:



```
[[484   8   7   9   1]
 [  5 496   0   0   3]
 [ 28   7 434   8   5]
 [ 34   6   7 444   7]
 [  7  14  11   5 470]]
k=  1  accuracy= 0.9312
[[490   7   6   5   1]
 [ 11 488   2   0   3]
 [ 39   6 421   7   9]
 [ 63   4   5 417   9]
 [ 20  10  10   5 462]]
k=  3  accuracy= 0.9112
[[497   3   6   2   1]
 [ 12 486   2   0   4]
 [ 44   3 420   5  10]
 [ 61   1   3 426   7]
 [ 18   4  11   6 468]]
k=  5  accuracy= 0.9188
```

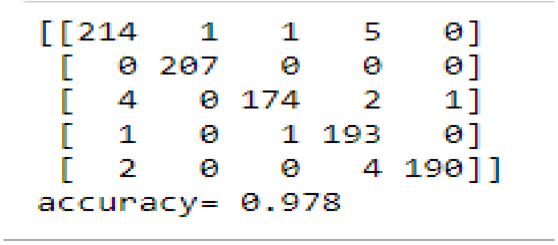


Here for k=1 accuracy is 0.9312 and k=3 accuracy is 0.9112 which is decreased and k=5 accuracy is 0.9188 which is increased again.

Data Split 80-20:

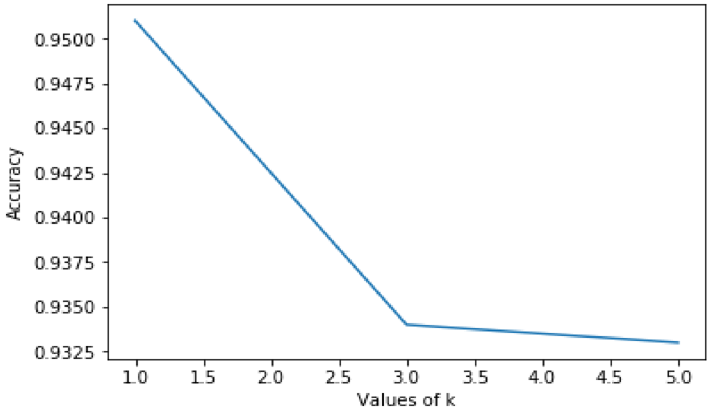
Feature selected with tfidf:

Naïve bayes:



```
[[214   1   1   5   0]
 [  0 207   0   0   0]
 [  4   0 174   2   1]
 [  1   0   1 193   0]
 [  2   0   0   4 190]]
accuracy= 0.978
```

KNN:

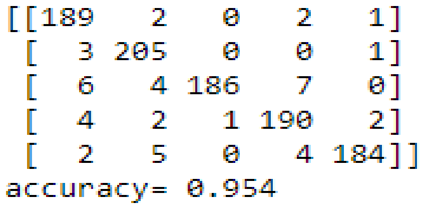


Here Accuracy goes from0.95 to 0.93.

Naïve bayes works better that knn for this split.

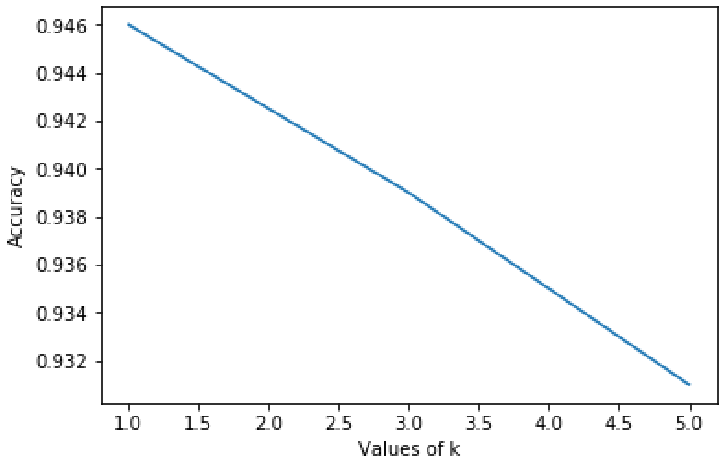
•*Feature selected with MI.*

Naïve Bayes:



```
[[189   2   0   2   1]
 [  3 205   0   0   1]
 [  6   4 186   7   0]
 [  4   2   1 190   2]
 [  2   5   0   4 184]]
accuracy= 0.954
```

KNN:



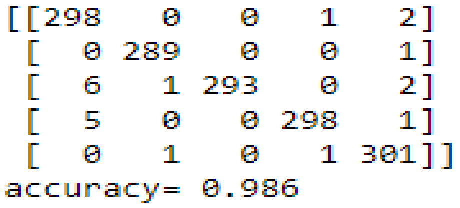
Here accuracy goes from 0.946 to 0.932.

Here too Accuracy of Naïve bayes is better than KNN.

Data Split 70-30:

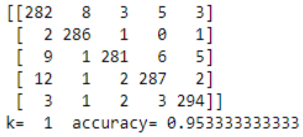
Feature selected with tfidf:

Naïve bayes:

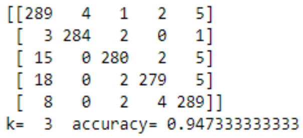


```
[[298   0   0   1   2]
 [  0 289   0   0   1]
 [  6   1 293   0   2]
 [  5   0   0 298   1]
 [  0   1   0   1 301]]
accuracy= 0.986
```

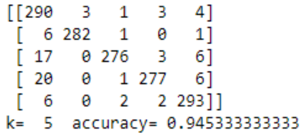
KNN:



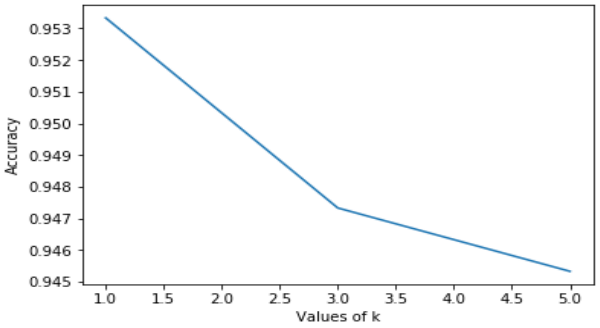
```
[[282   8   3   5   3]
 [  2 286   1   0   1]
 [  9   1 281   6   5]
 [ 12   1   2 287   2]
 [  3   1   2   3 294]]
k=  1  accuracy= 0.953333333333
```



```
[[289   4   1   2   5]
 [  3 284   2   0   1]
 [ 15   0 280   2   5]
 [ 18   0   2 279   5]
 [  8   0   2   4 289]]
k=  3  accuracy= 0.947333333333
```



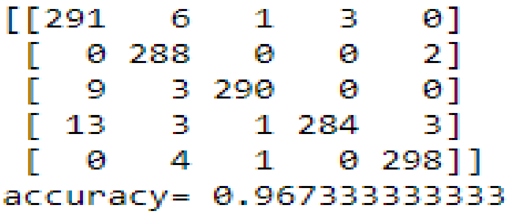
```
[[290   3   1   3   4]
 [  6 282   1   0   1]
 [ 17   0 276   3   6]
 [ 20   0   1 277   6]
 [  6   0   2   2 293]]
k=  5  accuracy= 0.945333333333
```



Here accuracy goes from 0.95 to 0.94.

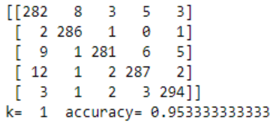
•*Feature selected with MI.*

Naïve Bayes:

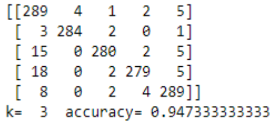


```
[[291   6   1   3   0]
 [  0 288   0   0   2]
 [  9   3 290   0   0]
 [ 13   3   1 284   3]
 [  0   4   1   0 298]]
accuracy= 0.967333333333
```

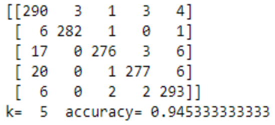
KNN:



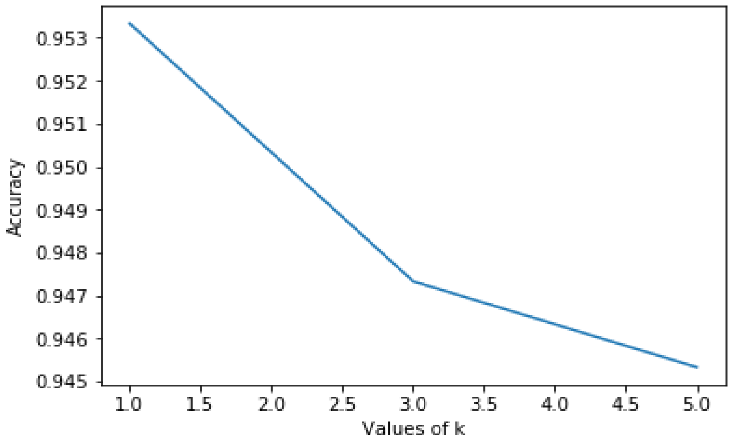
```
[[282   8   3   5   3]
 [  2 286   1   0   1]
 [  9   1 281   6   5]
 [ 12   1   2 287   2]
 [  3   1   2   3 294]]
k=  1  accuracy= 0.953333333333
```



```
[[289   4   1   2   5]
 [  3 284   2   0   1]
 [ 15   0 280   2   5]
 [ 18   0   2 279   5]
 [  8   0   2   4 289]]
k=  3  accuracy= 0.947333333333
```

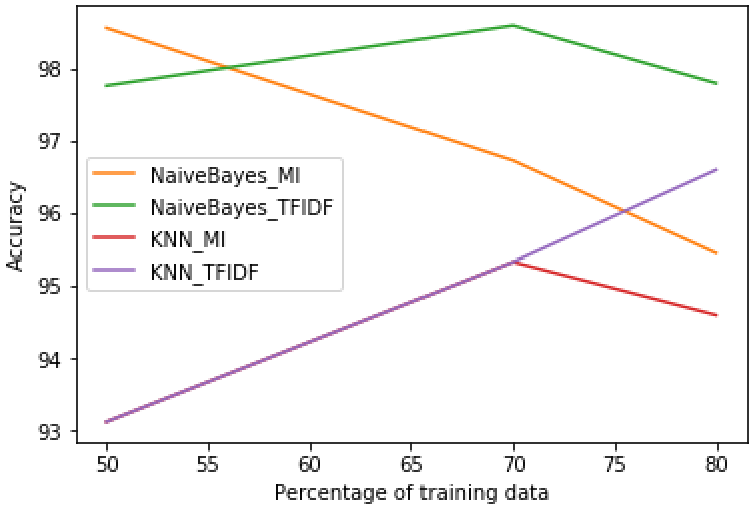


```
[[290   3   1   3   4]
 [  6 282   1   0   1]
 [ 17   0 276   3   6]
 [ 20   0   1 277   6]
 [  6   0   2   2 293]]
k=  5  accuracy= 0.945333333333
```



Here accuracy goes from 0.95 to 0.94.

Comparison Between Naïve Bayes and KNN using both feture selection Algorithm.



Here in this case Naïve bayes with Tf-IDF is performing better, followed by naïve bayes with MI, followed by KNN-tfidf and at last KNN_MI.

(Note: Here in some cases MI performs better and in some cases tf-idf performs better. This all will depend upon the data we have chosen for train and for test.)