

Assignment 3

Predictive Analytics

Name: Inavamsi B Enaganti

NetID: ibe214

Prof. Name: Bari Anasse

- **Confusion Matrix**
- **Brief report**
- **Sample Output**

Confusion matrix for Normal KNN.

K=3,5,7 same answer.

Distance Metric : Cosine Similarity

	Predicted Cluster 1	Predicted Cluster 4	Predicted Cluster 7	
Actual Cluster 1	5	0	0	TC1=5
Actual Cluster 4	0	3	0	TC4=3
Actual Cluster 7	0	0	2	TC7=2
	TP1=5	TP4=3	TP7=2	

Recall = $TP_i / (TP_i + FN_i)$ = TruePositive(i)/Actual size of cluster i

Precision = $TP_i / (TP_i + FP_i)$ = TruePositive(i)/Total predicted as i

Precision=1

Recall =1

F-measure = $2PR / (P+R)$ =1

We generated the confusion Matrix for the following Output

unknown01.txt belongs to Cluster : 1
unknown02.txt belongs to Cluster : 1
unknown03.txt belongs to Cluster : 1
unknown04.txt belongs to Cluster : 1
unknown05.txt belongs to Cluster : 4
unknown06.txt belongs to Cluster : 4
unknown07.txt belongs to Cluster : 7
unknown08.txt belongs to Cluster : 7
unknown09.txt belongs to Cluster : 4
unknown10.txt belongs to Cluster : 1

By looking at the documents the correct classification is

unknown01.txt belongs to Cluster : 1
unknown02.txt belongs to Cluster : 1
unknown03.txt belongs to Cluster : 1
unknown04.txt belongs to Cluster : 1
unknown05.txt belongs to Cluster : 4
unknown06.txt belongs to Cluster : 4
unknown07.txt belongs to Cluster : 7
unknown08.txt belongs to Cluster : 7
unknown09.txt belongs to Cluster : 4 (1 with a low %)
unknown10.txt belongs to Cluster : 1 or 7

Fuzzy KNN has also been implemented with the following Output for k=7.

unknown01.txt :	C1: 100 % ,	C4: 0 % ,	C7: 0
unknown02.txt :	C1: 100 % ,	C4: 0 % ,	C7: 0
unknown03.txt :	C1: 60 % ,	C4: 20 % ,	C7: 20
unknown04.txt :	C1: 100 % ,	C4: 0 % ,	C7: 0
unknown05.txt :	C1: 0 % ,	C4: 100 % ,	C7: 0
unknown06.txt :	C1: 0 % ,	C4: 100 % ,	C7: 0
unknown07.txt :	C1: 0 % ,	C4: 0 % ,	C7: 100
unknown08.txt :	C1: 0 % ,	C4: 0 % ,	C7: 100
unknown09.txt :	C1: 20 % ,	C4: 80 % ,	C7: 0
unknown10.txt :	C1: 60 % ,	C4: 40 % ,	C7: 0