CS 6375

ASSIGNMENT   3

Names of students in your group:

Ashwani Kumar Kashyap (AXK190033)

Anshul Pardhi (ARP180012)

Number of free late days used: _____0_____

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

- Lecture Slides
- reference Jaccard Distance: http://en.wikipedia.org/wiki/Jaccard_index
- Dataset url: https://archive.ics.uci.edu/ml/datasets/Health+News+in+Twitter

ASSIGNMENT 3

Experiment Analysis on "bbchealth" dataset

Value of K	SSE	Size of each clusters
3	3462.671	1: 1511 tweets
		2: 809 tweets
		3: 1609 tweets
4	3334.074	1: 1017 tweets
		2: 1166 tweets
		3: 816 tweets
		4: 930 tweets
5	3335.843	1: 1526 tweets
		2: 579 tweets
		3: 720 tweets
		4: 758 tweets
		5: 346 tweets
6	3300.087	1: 1317 tweets
		2: 470 tweets
		3: 781 tweets
		4: 447 tweets
		5: 447 tweets
		6: 467 tweets
7	3239.233	1: 605 tweets
		2: 452 tweets
		3: 675 tweets
		4: 536 tweets
		5: 762 tweets
		6: 510 tweets
		7: 389 tweets