

## 2. Clustering text with k-means

5. Prepare a table that shows the performance of vanilla k-means, random restart and seeding for 30, 50, 100, and 500 word lexicons.

	30 words	50 words	100 words	500 words
Vanilla K-Means	Homogeneity: [1.0, 1.0]	Homogeneity: [1.0, 1.0]	Homogeneity: [0.77, 1.0]	Homogeneity: [0.53, 0.67]
	Completeness: [1.0, 1.0]	Completeness: [1.0, 1.0]	Completeness: [0.7, 1.0]	Completeness: [0.8, 0.9]
Random Restart	Homogeneity: [1.0, 1.0]	Homogeneity: [1.0, 1.0]	Homogeneity: [1.0, 1.0]	Homogeneity: [0.56, 1.0]
	Completeness: [1.0, 1.0]	Completeness: [1.0, 1.0]	Completeness: [1.0, 1.0]	Completeness: [1.0, 0.8]
Seeding	Homogeneity: [1.0, 1.0]	Homogeneity: [1.0, 1.0]	Homogeneity: [1.0, 1.0]	Homogeneity: [0.53, 1.0]
	Completeness: [1.0, 1.0]	Completeness: [1.0, 1.0]	Completeness: [1.0, 1.0]	Completeness: [1.0, 0.9]

## 3. Naive Bayes

- a. Compare your classifier to ZeroR on the breast cancer data, using five-fold cross-validation for each, and include a table in your PDF showing the results.

	k = 1	k = 2	k = 3	k =4	k =5
Naive Bayes	0.44	0.32	0.57	0.52	0.41
ZeroR	0.0	0.0	0.0	0.0	0.0