Clustering Result

Homogeneity (H) - each cluster contains only members of a single class.

Completeness (C) - all members of a given class are assigned to the same cluster

V-measure (V) - is an entropy-based measure which explicitly measures how successfully the criteria of homogeneity and completeness have been satisfied.

Rand-index (R) - is a measure of the similarity between two data clusterings. A form of the Rand index may be defined that is adjusted for the chance grouping of elements, this is the adjusted Rand index. From a mathematical standpoint, Rand index is related to the accuracy.

Silhouette Coefficient (SC): close to 1 means that datum is clustered appropriately.

LSA – Latent sentiment analysis, to include sentiments like synonym and derivation of words etc

1) K-means (n-clusters – 8, generated itself)

Presence of features	N-gram (1,3)		N-gram (1,2)		N-gram (1,1)	
TfIdf + LSA	Evaluation criteria	Value	Evaluation criteria	value	Evaluation criteria	value
	H	0.257	H	0.253	H	0.266
	C	0.383	C	0.371	C	0.385
	V	0.307	V	0.301	V	0.315
	Rand	0.086	Rand	0.079	Rand	0.090
	SC	-0.245	SC	-0.259	SC	-0.240
TfIdf	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.271	H	0.357	H	0.132
	C	0.372	C	0.487	C	0.630
	V	0.314	V	0.412	V	0.218
	Rand	0.115	Rand	0.198	Rand	0.016
	SC	0.016	SC	0.030	SC	0.067
LSA	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.207	H	0.230	H	0.220
	C	0.355	C	0.370	C	0.359
	V	0.262	V	0.284	V	0.273
	Rand	0.048	Rand	0.052	Rand	0.053
	SC	-0.281	SC	-0.254	SC	-0.244

Both absent	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.080	H	0.220	H	0.158
	C	0.245	C	0.505	C	0.612
	V	0.120	V	0.306	V	0.251
	Rand	0.005	Rand	0.062	Rand	0.018
	SC	-0.059	SC	-0.038	SC	-0.001

2) Affinity Propagation

Presence/ absence of features	N-gram (1,3)		N-gram (1,2)		N-gram (1,1)	
TFIDF + LSA (n clusters – 191, generated itself)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	1.000	H	1.000	H	1.000
	C	0.526	C	0.526	C	0.526
	V	0.689	V	0.689	V	0.689
	Rand	0.000	Rand	0.000	Rand	0.000
	SC		SC		SC	
TfIdf (n clusters – 191, generated itself)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	1.000	H	1.000	H	1.000
	C	0.526	C	0.526	C	0.526
	V	0.689	V	0.689	V	0.689
	Rand	0.000	Rand	0.000	Rand	0.000
	SC		SC		SC	
LSA (n clusters – 191, generated itself)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	1.000	H	1.000	H	1.000
	C	0.526	C	0.526	C	0.526
	V	0.689	V	0.689	V	0.689
	Rand	0.000	Rand	0.000	Rand	0.000
	SC		SC		SC	
Both absent (n clusters – 8 (1,3), 41 (1,2), 2 (1,1), generated itself)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.054	H	0.425	H	0.022

<table>
<tr><td></td>
<td><table><tr><td>C</td><td>0.492</td></tr><tr><td>V</td><td>0.097</td></tr><tr><td>Rand</td><td>0.001</td></tr><tr><td>SC</td><td></td></tr></table></td>
<td><table><tr><td>C</td><td>0.398</td></tr><tr><td>V</td><td>0.411</td></tr><tr><td>Rand</td><td>0.040</td></tr><tr><td>SC</td><td></td></tr></table></td>
<td><table><tr><td>C</td><td>0.607</td></tr><tr><td>V</td><td>0.043</td></tr><tr><td>Rand</td><td>0.001</td></tr><tr><td>SC</td><td></td></tr></table></td>
</tr>
</table>

3) Mean Shift (clusters – 16, generated itself)

<table>
<tr><td>Presence/ absence of features</td><td>N-gram (1,3)</td><td>N-gram (1,2)</td><td>N-gram (1,1)</td></tr>
<tr><td>TfIdf+LSA</td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td>0.102</td></tr><tr><td>C</td><td>0.532</td></tr><tr><td>V</td><td>0.171</td></tr><tr><td>Rand</td><td>0.010</td></tr><tr><td>SC</td><td>0.055</td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td>0.196</td></tr><tr><td>C</td><td>0.400</td></tr><tr><td>V</td><td>0.263</td></tr><tr><td>Rand</td><td>0.051</td></tr><tr><td>SC</td><td>0.725</td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td>0.225</td></tr><tr><td>C</td><td>0.443</td></tr><tr><td>V</td><td>0.298</td></tr><tr><td>Rand</td><td>0.088</td></tr><tr><td>SC</td><td>0.710</td></tr></table></td>
</tr>
<tr><td>TfIdf
(was not running)</td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td></td></tr><tr><td>C</td><td></td></tr><tr><td>V</td><td></td></tr><tr><td>Rand</td><td></td></tr><tr><td>SC</td><td></td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td></td></tr><tr><td>C</td><td></td></tr><tr><td>V</td><td></td></tr><tr><td>Rand</td><td></td></tr><tr><td>SC</td><td></td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td></td></tr><tr><td>C</td><td></td></tr><tr><td>V</td><td></td></tr><tr><td>Rand</td><td></td></tr><tr><td>SC</td><td></td></tr></table></td>
</tr>
<tr><td>LSA</td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td>0.140</td></tr><tr><td>C</td><td>0.358</td></tr><tr><td>V</td><td>0.202</td></tr><tr><td>Rand</td><td>0.027</td></tr><tr><td>SC</td><td>0.792</td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td>0.142</td></tr><tr><td>C</td><td>0.364</td></tr><tr><td>V</td><td>0.204</td></tr><tr><td>Rand</td><td>0.028</td></tr><tr><td>SC</td><td>0.793</td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td>0.156</td></tr><tr><td>C</td><td>0.583</td></tr><tr><td>V</td><td>0.246</td></tr><tr><td>Rand</td><td>0.029</td></tr><tr><td>SC</td><td>0.749</td></tr></table></td>
</tr>
<tr><td>Both absent
(was not running)</td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td></td></tr><tr><td>C</td><td></td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td></td></tr><tr><td>C</td><td></td></tr></table></td>
<td><table><tr><td>Evaluation criteria</td><td>value</td></tr><tr><td>H</td><td></td></tr><tr><td>C</td><td></td></tr></table></td>
</tr>
</table>

	V		V		V	
	Rand		Rand		Rand	
	SC		SC		SC	

4) Heirarchical

Presence/ absence of features	N-gram (1,3)		N-gram (1,2)		N-gram (1,1)	
TfIdf + LSA	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.076	H	0.076	H	0.072
	C	0.603	C	0.603	C	0.662
	V	0.134	V	0.134	V	0.130
	Rand	0.015	Rand	0.015	Rand	0.013
	SC	0.870	SC	0.867	SC	0.731
TfIdf	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.034	H	0.034	H	0.063
	C	0.540	C	0.540	C	0.786
	V	0.064	V	0.064	V	0.116
	Rand	0.003	Rand	0.003	Rand	0.011
	SC	0.030	SC	0.036	SC	0.049
LSA	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.076	H	0.076	H	0.070
	C	0.603	C	0.603	C	0.616
	V	0.134	V	0.134	V	0.125
	Rand	0.015	Rand	0.015	Rand	0.012
	SC	0.870	SC	0.863	SC	0.766
Both absent	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.034	H	0.034	H	0.078
	C	0.540	C	0.540	C	0.747

	V	0.064	V	0.064	V	0.141
	Rand	0.003	Rand	0.003	Rand	0.015
	SC	0.295	SC	0.318	SC	0.382

5) DBScan

Presence/ absence of features	N-gram (1,3)		N-gram (1,2)		N-gram (1,1)	
TfIdf + LSA (n-cluster – 1, generated itself)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.071	H	0.076	H	0.078
	C	0.585	C	0.603	C	0.569
	V	0.127	V	0.134	V	0.137
	Rand	0.015	Rand	0.015	Rand	0.016
	SC	0.847	SC	0.838	SC	0.780
TfIdf (n-cluster – 0, generated itself)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.000	H	0.000	H	0.000
	C	1.000	C	1.000	C	1.000
	V	0.000	V	0.000	V	0.000
	Rand	-0.000	Rand	-0.000	Rand	-0.000
	SC		SC		SC	
LSA (n-cluster – 1, generated itself)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H	0.076	H	0.078	H	0.092
	C	0.585	C	0.507	C	0.507
	V	0.134	V	0.136	V	0.155
	Rand	0.015	Rand	0.016	Rand	0.025
	SC	0.836	SC	0.766	SC	0.708
Both absent (was giving poor result so did not note down)	Evaluation criteria	value	Evaluation criteria	value	Evaluation criteria	value
	H		H		H	
	C		C		C	
	V		V		V	

	Rand		Rand		Rand	
	SC		SC		SC	