# **CLUSTER VALIDATION:**

**Silhouette Score:**

* Measures the similarity of datapoints within cluster and dissimilarity wrt other clusters.
* (b-a)/max(a,b)

where a-intra-cluster distance

b – inter-cluster distance

|  |  |  |
| --- | --- | --- |
| # of clusters | Hierarchical | Hie + Spectral embedding |
| 3 | 0.56 | 0.6 |
| 4 | 0.5 | 0.57 |
| 5 | 0.53 | 0.54 |

* VAE model – Distributional Semantics in latent space (in progress).

# **CLUSTER PROFILING:**

* With clusters as labels – train supervised model – analyse features learnt / generate rules for clusters (in progress).