

Clustering Results

Dejvis Toptani

2023-05-17

Clustering of Ecommerce dataset.

DBSCAN

| min_samples | eps | umap_n_components | NMI | embedding |
|-------------|-------|-------------------|-------|-----------|
| 6 | 0.817 | 8 | 0.647 | BERT |
| 5 | 0.816 | 8 | 0.647 | BERT |
| 6 | 0.818 | 8 | 0.647 | BERT |
| 9 | 0.997 | 16 | 0.785 | fasttext |
| 7 | 0.996 | 16 | 0.785 | fasttext |
| 7 | 0.999 | 16 | 0.785 | fasttext |
| 6 | 0.998 | 8 | 0.643 | tfidf |
| 6 | 0.854 | 2 | 0.632 | tfidf |
| 5 | 0.864 | 2 | 0.632 | tfidf |
| 6 | 0.875 | 16 | 0.723 | word2vec |
| 6 | 0.877 | 16 | 0.723 | word2vec |
| 6 | 0.854 | 16 | 0.723 | word2vec |

KMeans

| n_clusters | umap_n_components | NMI | embedding |
|------------|-------------------|-------|-----------|
| 6 | 16 | 0.726 | BERT |
| 6 | 16 | 0.726 | BERT |
| 6 | 16 | 0.726 | BERT |
| 5 | 2 | 0.847 | fasttext |
| 5 | 2 | 0.847 | fasttext |
| 5 | 2 | 0.847 | fasttext |
| 4 | 16 | 0.803 | tfidf |
| 4 | 16 | 0.803 | tfidf |
| 4 | 16 | 0.803 | tfidf |
| 4 | 16 | 0.762 | word2vec |
| 4 | 16 | 0.762 | word2vec |
| 4 | 16 | 0.762 | word2vec |

Spectral Clustering

| n_clusters | umap_n_components | NMI | embedding |
|------------|-------------------|-------|-----------|
| 8 | 4 | 0.694 | BERT |
| 8 | 4 | 0.669 | BERT |

| n_clusters | umap_n_components | NMI | embedding |
|------------|-------------------|-------|-----------|
| 8 | 4 | 0.667 | BERT |
| 6 | 16 | 0.845 | fasttext |
| 8 | 16 | 0.828 | fasttext |
| 8 | 16 | 0.828 | fasttext |
| 7 | 16 | 0.753 | tfidf |
| 7 | 16 | 0.753 | tfidf |
| 7 | 16 | 0.753 | tfidf |
| 8 | 4 | 0.770 | word2vec |
| 8 | 4 | 0.770 | word2vec |
| 8 | 4 | 0.770 | word2vec |

OPTICS

| min_samples | xi | umap_n_components | NMI | embedding |
|-------------|-------|-------------------|-------|-----------|
| 128 | 0.168 | 4 | 0.703 | BERT |
| 125 | 0.170 | 4 | 0.703 | BERT |
| 124 | 0.167 | 4 | 0.703 | BERT |
| 102 | 0.090 | 8 | 0.814 | fasttext |
| 93 | 0.068 | 8 | 0.814 | fasttext |
| 98 | 0.058 | 8 | 0.814 | fasttext |
| 195 | 0.053 | 16 | 0.756 | tfidf |
| 189 | 0.061 | 16 | 0.756 | tfidf |
| 195 | 0.057 | 16 | 0.756 | tfidf |
| 63 | 0.272 | 4 | 0.746 | word2vec |
| 64 | 0.225 | 4 | 0.746 | word2vec |
| 57 | 0.249 | 4 | 0.746 | word2vec |

DBHD

| min_cluster_size | rho | beta | umap_n_components | NMI | embedding |
|------------------|-------|-------|-------------------|-------|-----------|
| 167 | 1.222 | 0.152 | 8 | 0.643 | BERT |
| 171 | 1.418 | 0.168 | 8 | 0.642 | BERT |
| 168 | 1.368 | 0.142 | 16 | 0.640 | BERT |
| 192 | 0.498 | 0.311 | 4 | 0.763 | fasttext |
| 192 | 0.545 | 0.333 | 4 | 0.763 | fasttext |
| 193 | 0.799 | 0.274 | 4 | 0.763 | fasttext |
| 140 | 0.103 | 0.101 | 16 | 0.670 | tfidf |
| 136 | 0.080 | 0.124 | 16 | 0.670 | tfidf |
| 145 | 0.054 | 0.116 | 16 | 0.670 | tfidf |
| 162 | 0.060 | 0.297 | 16 | 0.703 | word2vec |
| 167 | 0.559 | 0.296 | 16 | 0.703 | word2vec |
| 172 | 0.253 | 0.182 | 16 | 0.703 | word2vec |

Meanshift

| bandwidth | umap_n_components | NMI | embedding |
|-----------|-------------------|-------|-----------|
| 0.909 | 4 | 0.573 | BERT |
| 0.910 | 4 | 0.573 | BERT |

| bandwidth | umap_n_components | NMI | embedding |
|-----------|-------------------|-------|-----------|
| 0.909 | 4 | 0.573 | BERT |
| 0.910 | 16 | 0.588 | fasttext |
| 0.908 | 4 | 0.586 | fasttext |
| 0.908 | 4 | 0.585 | fasttext |
| 0.905 | 8 | 0.530 | tfidf |
| 0.906 | 8 | 0.530 | tfidf |
| 0.905 | 8 | 0.530 | tfidf |
| 0.892 | 16 | 0.598 | word2vec |
| 0.905 | 16 | 0.598 | word2vec |
| 0.908 | 16 | 0.598 | word2vec |

SNN-DPC

| k | nc | umap_n_components | NMI | embedding |
|----|----|-------------------|-------|-----------|
| 20 | 6 | 4 | 0.701 | BERT |
| 20 | 6 | 4 | 0.701 | BERT |
| 20 | 6 | 4 | 0.701 | BERT |
| 15 | 6 | 2 | 0.833 | fasttext |
| 15 | 6 | 2 | 0.833 | fasttext |
| 15 | 6 | 2 | 0.833 | fasttext |
| 11 | 6 | 16 | 0.768 | tfidf |
| 13 | 6 | 16 | 0.768 | tfidf |
| 11 | 6 | 16 | 0.768 | tfidf |
| 21 | 6 | 4 | 0.770 | word2vec |
| 21 | 6 | 4 | 0.770 | word2vec |
| 21 | 6 | 4 | 0.770 | word2vec |

HDBSCAN

| min_cluster_size | min_samples | umap_n_components | NMI | embedding |
|------------------|-------------|-------------------|-------|-----------|
| 160 | 72 | 4 | 0.683 | BERT |
| 142 | 73 | 4 | 0.683 | BERT |
| 175 | 71 | 4 | 0.683 | BERT |
| 98 | 108 | 2 | 0.822 | fasttext |
| 148 | 108 | 2 | 0.822 | fasttext |
| 195 | 11 | 4 | 0.813 | fasttext |
| 130 | 15 | 16 | 0.751 | tfidf |
| 161 | 15 | 16 | 0.751 | tfidf |
| 176 | 15 | 16 | 0.751 | tfidf |
| 95 | 28 | 2 | 0.783 | word2vec |
| 90 | 32 | 2 | 0.783 | word2vec |
| 86 | 32 | 2 | 0.783 | word2vec |

SpectralACL

| n_clusters | epsilon | umap_n_components | NMI | embedding |
|------------|---------|-------------------|-----|-----------|
| 4 | 0.450 | 2 | 0 | BERT |
| 4 | 0.443 | 16 | 0 | BERT |

| n_clusters | epsilon | umap_n_components | NMI | embedding |
|------------|---------|-------------------|-----|-----------|
| 4 | 0.637 | 16 | 0 | BERT |
| 4 | 0.761 | 4 | 0 | fasttext |
| 4 | 0.071 | 2 | 0 | fasttext |
| 4 | 0.253 | 4 | 0 | fasttext |
| 4 | 0.521 | 8 | 0 | tfidf |
| 4 | 0.931 | 16 | 0 | tfidf |
| 4 | 0.296 | 8 | 0 | tfidf |
| 4 | 0.281 | 4 | 0 | word2vec |
| 4 | 0.756 | 2 | 0 | word2vec |
| 4 | 0.797 | 2 | 0 | word2vec |

DBADV

| perplexity | MinPts | probability | umap_n_components | NMI | embedding |
|------------|--------|-------------|-------------------|-------|-----------|
| 23 | 8 | 0.997 | 8 | 0.650 | BERT |
| 23 | 8 | 0.997 | 8 | 0.650 | BERT |
| 23 | 5 | 0.997 | 8 | 0.650 | BERT |
| 30 | 22 | 0.997 | 16 | 0.806 | fasttext |
| 29 | 25 | 0.997 | 16 | 0.806 | fasttext |
| 29 | 25 | 0.997 | 16 | 0.806 | fasttext |
| 28 | 7 | 0.997 | 8 | 0.679 | tfidf |
| 28 | 5 | 0.997 | 8 | 0.679 | tfidf |
| 28 | 7 | 0.997 | 8 | 0.679 | tfidf |
| 27 | 8 | 0.997 | 16 | 0.725 | word2vec |
| 27 | 8 | 0.997 | 16 | 0.725 | word2vec |
| 26 | 1 | 0.997 | 16 | 0.725 | word2vec |

DPC

| density_threshold | distance_threshold | umap_n_components | NMI | embedding |
|-------------------|--------------------|-------------------|-------|-----------|
| 0.666 | 0.949 | 8 | 0.636 | BERT |
| 0.659 | 0.969 | 8 | 0.636 | BERT |
| 0.654 | 0.940 | 8 | 0.636 | BERT |
| 0.848 | 0.996 | 16 | 0.659 | fasttext |
| 0.533 | 0.999 | 16 | 0.659 | fasttext |
| 0.591 | 0.998 | 16 | 0.659 | fasttext |
| 0.706 | 1.000 | 8 | 0.547 | tfidf |
| 0.773 | 0.997 | 8 | 0.547 | tfidf |
| 0.651 | 0.999 | 8 | 0.547 | tfidf |
| 0.457 | 0.995 | 8 | 0.652 | word2vec |
| 0.502 | 0.995 | 8 | 0.652 | word2vec |
| 0.503 | 0.999 | 8 | 0.652 | word2vec |