Raw data (only dropped the ID column)

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| KMeans | n\_clusters= 10, random\_state=0 | 0.35470 |
| KMeans | n\_clusters= 10, random\_state=42 | 0.34885 |
| KMeans | n\_clusters= 10, random\_state=329 | 0.36317 |
| HDBSCAN | n\_clusters = 10 | 0.06723 |
| HDBSCAN | n\_clusters = 5 | 0.06763 |
| SpectralClustering | n\_clusters=10, affinity='nearest\_neighbors', random\_state=42 | 0.38197 |
| GaussianMixture | n\_components=10, random\_state=42 | 0.28269 |
| KMeans | (normalized data)  n\_clusters= 10, random\_state=329 | 0.36317 |

Dimension reduction (PCA) + Normalization using MinMaxScaler()

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| KMeans | (normalized data, reduced with PCA)  n\_clusters= 10, random\_state=329 | 0.37411 |
| SpectralClustering | (normalized data, reduced with PCA)  n\_clusters=10, affinity='nearest\_neighbors', random\_state=42 | 0.40927 |
| AgglomerativeClustering | (normalized data, reduced with PCA)  n\_clusters=2, metric='euclidean' | 0.13182 |
| AgglomerativeClustering | (normalized data, reduced with PCA)  n\_clusters=10, metric='euclidean' | 0.51538 |
| SpectralClustering | (normalized data, reduced with PCA)  n\_clusters=10, affinity='nearest\_neighbors', random\_state=42 | 0.40214 |
| AgglomerativeClustering | (normalized data, reduced with PCA)  n\_clusters=10, metric='euclidean' | 0.46972 |
| KMeans | (normalized data, reduced with PCA)  n\_clusters= 10, init='k-means++', random\_state=42 | 0.33285 |
| Birch | (normalized data, reduced with PCA)  threshold=0.5, branching\_factor=50, n\_clusters=10, compute\_labels=True, copy=True | 0.43041 |

Dimension reduction (t-SNE + assorted unsuccessful attempts)

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| AgglomerativeClustering | (normalized data, reduced with t-SNE = 2)  n\_clusters=10, metric='euclidean' | 0.53368 |
| AgglomerativeClustering | (normalized data, reduced with t-SNE = 3)  n\_clusters=10, metric='euclidean' | 0.46870 |
| AgglomerativeClustering | (normalized data, reduced with truncatedSVD)  n\_clusters=10, metric='euclidean' | 0.12309 |
| AgglomerativeClustering | (normalized data, reduced with Isomap)  n\_clusters=10, metric='euclidean' | 0.25725 |
| SpectralClustering | (normalized data, reduced with t-SNE=2)  n\_clusters=10, affinity='nearest\_neighbors', random\_state=42 | 0.53044 |
| MiniBatchKMeans | (normalized data, reduced with t-SNE=2)  n\_clusters=10, random\_state=0 | 0.46141 |
| Divisive Clustering | (normalized data, reduced with t-SNE)  fcluster(t=10, criterion='maxclust’) | 0.50481 |

Normalization + Dimension reduction (single t-SNE, double t-SNE, triple t-SNE)

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| Birch | (normalized data, reduced with t-SNE = 2 twice)  threshold=1.5, branching\_factor=150, n\_clusters=10' | 0.50472 |
| Birch | (normalized data, reduced with t-SNE = 2 three times)  threshold=1.5, branching\_factor=150, n\_clusters=10' | 0.46999 |
| Birch | (normalized data, reduced with PCA + t-SNE)  threshold=1.5, branching\_factor=150, n\_clusters=10' | 0.37194 |
| Birch | (normalized data, reduced with t-SNE=2)  threshold=0.5, branching\_factor=50, n\_clusters=10 | 0.53448 |

Boolean data + Dimension reduction (single t-SNE, double t-SNE, triple t-SNE)

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| Birch | (boolean data, reduced with t-SNE = 2 twice)  threshold=1.5, branching\_factor=150, n\_clusters=10' | 0.58438 |
| Birch | (boolean data, reduced with t-SNE = 2 three times)  threshold=1.5, branching\_factor=150, n\_clusters=10' | 0.57936 |
| Birch | (boolean data, reduced with t-SNE = 2 twice)  threshold=0.8, branching\_factor=150, n\_clusters=10' | 0.59438 |
| Birch | (boolean data, reduced with t-SNE = 2 three times)  threshold=0.8, branching\_factor=150, n\_clusters=10' | 0.59658 |
| Birch | (boolean data, reduced with t-SNE = 2 twice)  threshold=0.4, branching\_factor=120, n\_clusters=10' | 0.60362 |

Inverse Boolean data + Dimension reduction (single t-SNE, double t-SNE, triple t-SNE)

TSNE(n\_components=2, random\_state=42)

TSNE(n\_components=2, random\_state=222)

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| Birch | (inverse boolean data, reduced with t-SNE = 2, random\_state=42 twice)  threshold=0.4, branching\_factor=120, n\_clusters=10 | 0.62324 |
| Birch | (inverse boolean data, reduced with t-SNE = 2, random\_state=42 twice)  threshold=0.6, branching\_factor=80, n\_clusters=10 | 0.62995 |
| Birch | (inverse boolean data, reduced with t-SNE = 2, random\_state=222 twice)  threshold=0.6, branching\_factor=80, n\_clusters=10 | 0.63257 |
| Birch | (inverse boolean data, reduced with t-SNE = 2, random\_state=222 twice)  threshold=1.3, branching\_factor=200, n\_clusters=10 | 0.63719 |

Inverse Boolean data + Dimension reduction (using HOG) + No other preprocessing

hog(image\_2d, orientations=9, pixels\_per\_cell=(8, 8), cells\_per\_block=(2, 2), block\_norm='L2-Hys')

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| KMeans | n\_clusters=10 | 0.35940 |
| Birch | threshold = 0.6, branching\_factor = 200, n\_clusters=10 | 0.33126 |
| Birch | threshold = 0.7, branching\_factor = 100, n\_clusters=10 | -0.00130 |

Inverse Boolean data + Dimension reduction (using HOG and then t-SNE)

hog(image\_2d, orientations=9, pixels\_per\_cell=(8, 8), cells\_per\_block=(2, 2), block\_norm='L2-Hys')

TSNE(n\_components=2, random\_state=222)

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| KMeans | n\_clusters=10 | 0.46348  0.50212 |
| Birch | threshold = 0.6,  branching\_factor = 200, n\_clusters=10 | 0.44412 |

Inverse Boolean data + Dimension reduction (using UMAP)

umap = UMAP(n\_neighbors=10, min\_dist=0.1, n\_components=2)

| **Model** | **Parameters** | **Kaggle Score** |
| --- | --- | --- |
| KMeans | n\_clusters=10 | 0.35940 |
| AgglomerativeClustering | random\_state=131  Agglo:  n\_clusters=10,  metric='euclidean',  linkage=’ward’ | 0.60699 |
| Birch | random\_state=53  Birch:  threshold = 0.8,  branching\_factor = 80, n\_clusters=10 | 0.63899 |
| Birch | random\_state=56  Birch:  threshold = 0.8,  branching\_factor = 80, n\_clusters=10 | 0.64194 |
| Birch | random\_state=59  Birch:  threshold = 0.8,  branching\_factor = 80, n\_clusters=10 | 0.66129 |
| Birch | random\_state=60  Birch:  threshold = 0.1,  branching\_factor = 10, n\_clusters=10 | 0.66301 |

Inverse Boolean data + CNN:

| 0.38259 |
| --- |

**CODE:**

[github repository](https://github.com/leenelzoubii/Digits-Clustering---ML2)