COMPARISON BETWEEN FIVE CLUSTERING ALGORITHMS APPLIED TO THREE DIFFERENT DATASETS

Elron Bandel
Department of Computer Science
Bar-Ilan University
elronbandel@gmail.com

ABSTRACT

We use 5 different clustering algorithms to learn meaningful clusters in 3 different datasets. The clustering algorithms we use, KMeans, GMM, Louvain, Prim and DBSCAN are all unsupervised learning algorithms that utilize different approaches to find meaningful clusters in the data. In this work we apply every one of them on each of the datasets, search for optimal hyper parameters and discuss the different results. All the code, experiments results and figured can be found in https://github.com/elronbandel/clustering.

1 INTRODUCTION

In the past decades growing amount of information is being stored digitally. The accessibility of large data sets raised the possibility to learn significant insights from the available data. Naturally, most of the data available does not contain human made analysis that can guide learning systems it is required to gather insights without any annotated supervision. Clustering algorithms can indicate on patterns that characterize behavioural different groups in data. We use 5 different such algorithms, then compare their results and discuss each result thoroughly.

2 METHODS

In this work we use few clustering algorithms as well as different evaluation methods and statistical tests to compare between them. The methods section describe each of the methods we use.

2.1 CLUSTERING ALGORITHMS

In this sections we describe the different clustering algorithms we use, their approach to the clustering problem and their unique hyper parameters we later tune.

2.1.1 K-MEANS

K-Means, MacQueen (1967), an algorithm that look for pre-defined number of centeroids points in the data space that minimize the sum of distances from every data point to the closest centeroid. This objective called inertia and can be described by the following formulation: $\sum_{i=0}^{n} \min_{\mu_j \in C}(||x_i - \mu_j||^2)$. K-Means objective can find clusters that are convex and isotropic, which is not always the case. It responds poorly to elongated clusters, or manifolds with irregular shapes and since it is optimizing euclidian distance it perform worse in higher dimension due to the "curse of dimensionality".

2.1.2 GMM

A Gaussian mixture mode, GMM, is a probabilistic model that assumes all the data points are generated from a mixture of a pre-defined number of Gaussian distributions with unknown parameters. The characteristics of those distribution can be approximated with Expectation Maximization Algorithms. After we found the Gaussians the maximize the likelihood of the data we can either softly or hardly assign One Gaussian or more to every data point as its cluster. In higher dimensionality GMM tend to under-perform and to be computationally expensive due to the computation of the large covariance matrices.

2.1.3 Louvain

Louvain Algorithm, Blondel et al. (2008), finds in graphs communities with high modularity that measures the relative density of edges inside communities with respect to edges outside communities. The algorithm is greedy in the sense that it build communities by local decisions of merging existing communities starting from merging separate nodes. Louvain algorithm can be applied to set of vectors by creating fully connected graph from with edges with length of the distance between every two vectors. Louvain algorithm can optimize different definitions of modularity in graphs such as Dugue, Newman and Potts.

2.1.4 Prim

Prim Based Clustering, VanderPlas (2016), Algorithm also operate on graphs but instead of merging densed communities it separates far sub groups in the data. The algorithm first finds Minimal Spanning Tree and then cut off edges above pre-defined threshold by order of length. The intuition behind it is that clusters connected in the MST with short edges and wont separate during the cut off phases.

2.1.5 DBSCAN

DBSCAN - Density-Based Spatial Clustering of Applications with Noise, Ester et al. (1996). Finds core samples of high density and expands clusters from them. Those points defined by how many points are in their pre-defined radios, if they have more then pre-defined amount of points in their surrounding they will be defined as core samples and iteratively connect with other core samples and their satellites points forming a cluster.

2.2 Clustering Evaluation

In order to evaluate the quality of the clusters we received from the clustering algorithm we used few well known clustering evaluation methods:

2.2.1 Silhouette

The Silhouette score, Rousseeuw (1987), is a method to evaluate the clustering using the internal training data after it was clustered by the algorithm. The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is (b - a) / max(a, b). To clarify, b is the distance between a sample and the nearest cluster that the sample is not a part of. This score can measure how densed is every cluster and how seprate are the clusters from each other. The best value is 1 and the worst value is -1. Values near 0 indicate overlapping clusters. Negative values generally indicate that a sample has been assigned to the wrong cluster, as a different cluster is more similar. One of the drawbacks of The Silhouette score is that it is generally higher for convex clusters than other concepts of clusters, such as density based clusters like those obtained through DBSCAN.

2.2.2 V-measure score

The V-measure, Rosenberg & Hirschberg (2007), is the harmonic mean between homogeneity and completeness. homogeneity means that each cluster contains only members of a single class, completeness means that all members of a given class are assigned to the same cluster. Together the V-measure score gives high scores to clusters that correlate with strongly with given classes and strongly agree on the labeling of every class.

2.3 Dimansion Reduction

High dimensality data can be hard to work with because of the computational cost of optimzation in many dimensions and becuse of the 'curse of dimensionality' that make it hard to gather important insights without being decived by the many different dimensions. Therefore, it can be beneficial to reduce the dimensions of the data samples before using clustering algorithms. we used few methods and compared between them/

2.3.1 PCA

Principal Component Analysis, Pearson (1901), or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. PCA helps us compute the Principal components in data. Principal components are basically vectors that are linearly uncorrelated and have a variance with in data. From the principal components top p is picked which have the most variance.

2.3.2 t-SNE

t-distributed Stochastic Neighbor Embedding (t-SNE), van der Maaten & Hinton (2008), is a echnique to reduce dimensions in high-dimensional data mainly for visualizations. It converts similarities between data points to joint probabilities and tries to minimize the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. t-SNE has a cost function that is not convex, i.e. with different initializations we can get different results.

2.3.3 UMAP

Uniform Manifold Approximation and Projection (UMAP), McInnes et al. (2020), is a dimension reduction technique that can be used for visualisation similarly to t-SNE, but also for general non-linear dimension reduction. The algorithm is founded on three assumptions about the data: (1) The data is uniformly distributed on Riemannian manifold (2) The Riemannian metric is locally constant (or can be approximated as such) (3) The manifold is locally connected. From these assumptions it is possible to model the manifold with a fuzzy topological structure. The embedding is found by searching for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure.

2.4 Statistical Tests

The comparison between the different results can be statistically meaningful only if it the results are not random. To test how likely it is to say that the difference between the results is not random we use statistical tests.

2.4.1 One-Way ANOVA

Analysis of variance (ANOVA) is a form of statistical hypothesis testing heavily used in the analysis of experimental data. A test result (calculated from the null hypothesis and the sample) is called statistically significant if it is deemed unlikely to have occurred by chance, assuming the truth of the null hypothesis. A statistically significant result, when a probability (p-value) is less than a pre-specified threshold (significance level), justifies the rejection of the null hypothesis, but only if the a priori probability of the null hypothesis is not high.

2.4.2 Scheffe Post-Hoc Test

After you have run ANOVA and got a significant results (i.e. you have rejected the null hypothesis that the means are the same), then you run Sheffe's test to find out which pairs of means are significant. The Scheffe test corrects alpha for simple and complex mean comparisons. Complex mean comparisons involve comparing more than one pair of means simultaneously. We use this test in cases that the ANOVA indicate significant means difference and we want to analyze what pairs of results contributed to the difference.

3 Results

3.1 Diabetes Data set

The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. The Information contains only information related to diabetes. The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc. Before using the data we explored it, cleaned it and fix missing values.
For every algorithm we try different hyper paramaters and report the results with their p value.

3.1.1 K-Means

In our experiments we find that, for K-Means Clustering, dimension reduction with UMAP is significantly better than PCA $(p < 0.001)$. It is better to remove outlires $(p < 0.001)$. With the Elbow Method we found that the 5-8 clusters are producing the best interia scores (explained in the K-Means section) with the best marginal contribution. We find that 7 clusters gave the best Silhouette score $(p < 0.001)$. and that the difference between the numbers was significant by the Scheffe test. The v-scores with the labels was very low (< 0.001). indicating that the clusters were not correlated with the gender or the race labels. With the best hyper parameters we found, K-Means final average Silhouette score is 0.64.

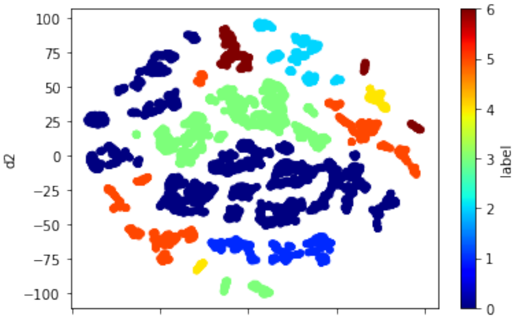

Figure 1: t-SNE Visualization of Clusters with K-Means Algorithm

3.1.2 GMM

In our experiments we find that, for GMM Clustering, dimension reduction with UMAP is significantly better than PCA $(p < 0.001)$. It is better to reduce to 5 dimensions rather than 10, 20 or 30 dimensions $(p < 0.01)$ but the difference between 5 and dimensions is not strong by the Scheffe test. It is better to remove outlires but without strong statistical significance $(p = 0.055)$. With the Elbow Method we found that the 5-8 clusters are producing the best BIC scores (soft version of the interia) with the best marginal contribution. We find that 8 clusters gave the best Silhouette score $(p < 0.001)$ and that the difference between the numbers was significant by the Scheffe test. .We find that the Tied and Diag covariance types are better than the Spherical $(p < 0.001)$ but there is no difference bwetween them by the Scheffe test. The v-scores with the labels was very low (< 0.01) indicating that the clusters were not correlated with the gender or the race labels. With the best hyper parameters we found, GMM final average Silhouette score is 0.628.

3.1.3 Louvain

In our experiments we find that, for Louvain, dimension reduction with UMAP was significantly better than PCA $(p < 0.001)$. There is not major difference in results between the number of dimensions so we reduced it to 5 since it is computationally less costly. It is better to remove outlires $(p = 0.035)$. We find that the Potts modularity definition yield signficantly better results than Dugue

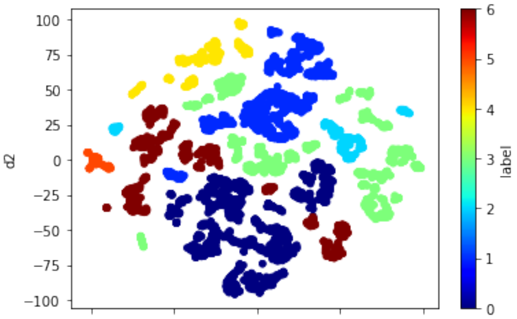

Figure 2: t-SNE Visualization of Clusters with GMM Algorithm

or Newman ($p < 0.001$). The v-scores with the labels is very low (< 0.01) indicating that the clusters were not correlated with the gender or the race labels. With the best hyper parameters we found, Louvain final average Silhouette score is 0.64.

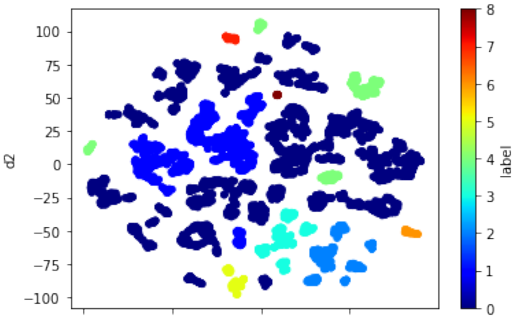

Figure 3: t-SNE Visualization of Clusters with Louvain Algorithm

3.1.4 Prim

In our experiments we find that, for Prim Based Clustering, dimension reduction with UMAP is significantly better than PCA ($p < 0.001$).There is no major difference in results between the number of dimensions except lower dimensions tend to be more noisy in their scores. Outlires removal did not contribute significantly. Even though small cutoff scale yield sometimes much better clustering on avarage we found that 0.8-0.9 is better than 0.7 or 0.6 ($p < 0.001$). The v-scores with the labels is very low (< 0.01) indicating that the clusters were not correlated with the gender or the race labels. With the best hyper parameters we found, Prim final average Silhouette score is 0.71.

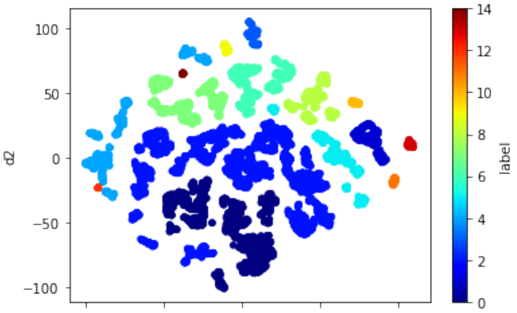

Figure 4: t-SNE Visualization of Clusters with Prim Based Clustering

3.1.5 DBSCAN

We find that, for DBSCAN, dimension reduction with UMAP was significantly better than PCA $(p < 0.001)$. From our experiments it is most benefiticial to reduce to 10 dimensions but it was not statistically significant. Outlires removal did not contribute. We find that higher relative Epsilon values (Neigbourhood Range) are better, specifically, 0.75 and 0.9 are better than 0.25 or 0.5 $(p < 0.001)$. We also find that with good epsilon value the number of neibours defining core nodes did not change much. The v-scores with the labels is very low (< 0.01) indicating that the clusters were not correlated with the gender or the race labels. With the best hyper parameters we found, DBSCAN final average Silhouette score is 0.70.

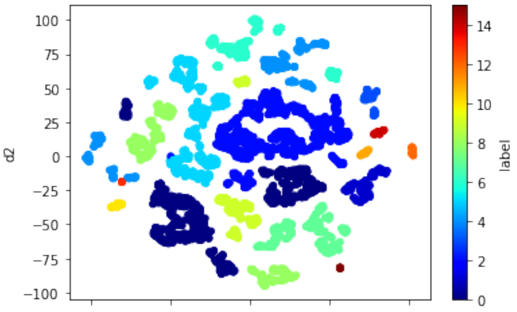

Figure 5: t-SNE Visualization of Clusters with DBSCAN

3.1.6 Diabetes Dataset Clustering Discussion

The dibaties data containes many dimensions of complex data it seems like the patterns that unified the different clusters most sucessfully were not linear since the non linear algorithms did the best and the ones based on Local Density such as Prim and DBSCAN did significantly better proposing that the similar examples formed dense areas that were not necessarily linear.

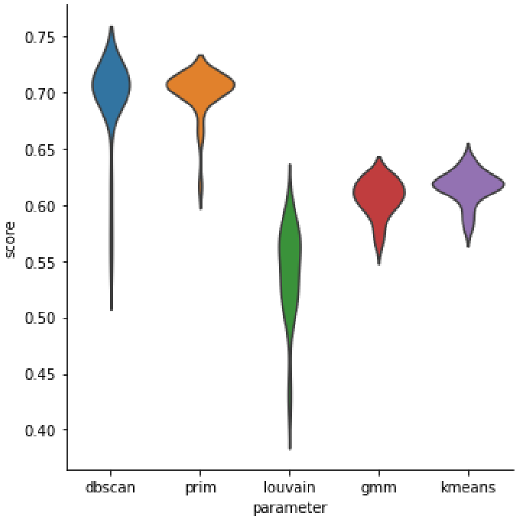

Figure 6: Silhouette Score of The Different Clustering Methods on Diabetes Dataset

3.2 Online Shopping Click Stream Data set

The dataset contains information on clickstream from online store offering clothing for pregnant women. Data are from five months of 2008 and include, among others, product category, location of

the photo on the page, country of origin of the IP address and product price in US dollars. In order to have one data point for every click stream we converted the stream to features like, stream length, items in the stream, categories in the stream in changes between categories to reflect the flow of the stream.
We clustered the data with all the clustering algorithms mentioned before. For every algorithm we try different hyper parameters and report the results along with their p value.

3.2.1 K-Means

In our experiments we find that, for K-Means Clustering, dimension reduction with UMAP is significantly better than PCA ($p < 0.001$). It is better to remove outlires ($p < 0.001$). With the Elbow Method we found that the 2-6 clusters are producing the best interia scores the best marginal contribution. We find that 5 and 6 clusters give the best Silhouette score ($p < 0.001$) with no difference between them which means only k=5 has noticeable marginal contribution. The v-scores with the labels was very low (< 0.001). indicating that the clusters were not correlated with the country label. With the best hyper parameters we found, K-Means final average Silhouette score is 0.31, fairly low.

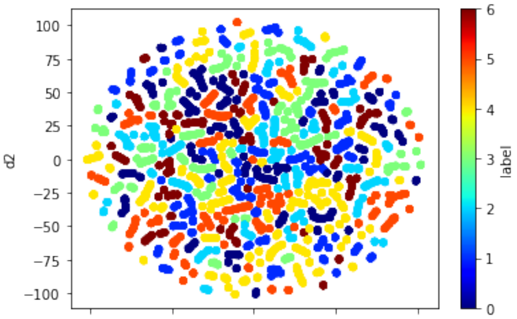

Figure 7: t-SNE Visualization of Clusters with K-Means Algorithm

3.2.2 GMM

In our experiments we find that, for GMM Clustering, dimension reduction with PCA is significantly better than Umap ($p < 0.001$). but had no improvement over no dimension reduction at all ($p < 0.01$). It is significantly better to remove ($p < 0.001$). With the Elbow Method we found that the 5-8 clusters are producing the best BIC scores (soft version of the interia) with the best marginal contribution. We find that 8 clusters gave the best Silhouette score ($p < 0.001$) and that the difference between the numbers was significant by the Scheffe test .We find that the Spherical covariance type is better than the Tied and Diagonal ($p < 0.001$). The v-scores with the labels is very low (< 0.01) indicating that the clusters were not correlated with the gender or the race labels. With the best hyper parameters we found, GMM final average Silhouette score is 0.628.

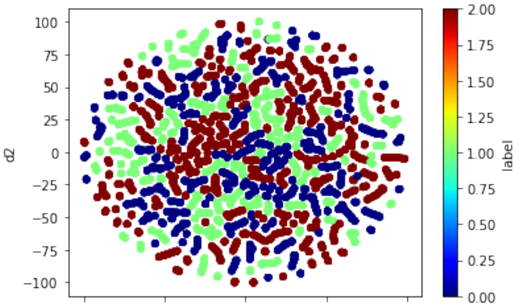

Figure 8: t-SNE Visualization of Clusters with GMM Algorithm

3.2.3 Louvain

In our experiments we find that, for Louvain, dimension reduction with UMAP was significantly better than PCA ($p < 0.001$). There was major difference between the number of dimensions ($p < 0.001$) so 15 or 20 dimension are the best with not difference between them. It is better to remove outlires but with no strong sigincance ($p = 0.13$). We find that the Dugue or Newman modularity definition yield signficantly better results than Potts ($p < 0.001$) but with no difference between them. The v-scores with the labels is very low (< 0.01) indicating that the clusters were not correlated with the country label. With the best hyper parameters we found, Louvain final average Silhouette score is 0.19 which is low.

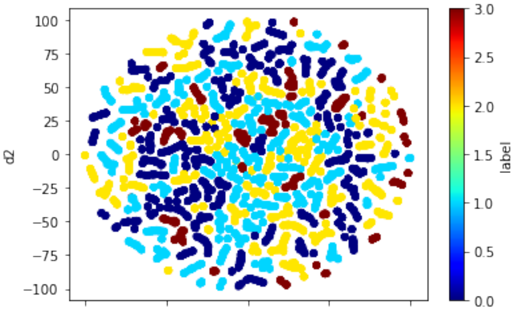

Figure 9: t-SNE Visualization of Clusters with Louvain Algorithm

3.2.4 Prim

In our experiments we find that, for Prim Based Clustering, dimension reduction with UMAP is significantly better by large margin than PCA ($p < 0.001$). It is better to reduce to 10 dimension rather then 5,10,15,20,30 or 40 ($p < 0.001$). Outlires removal made the algorithm to preform less good. There was no difference between the different cutoffs indicating that the distances between clusters were very high to begin with. The v-scores with the labels is very low (0.03) indicating that the clusters were not correlated with the country label. With the best hyper parameters we found, Prim final average Silhouette score is 0.91 which is very high.

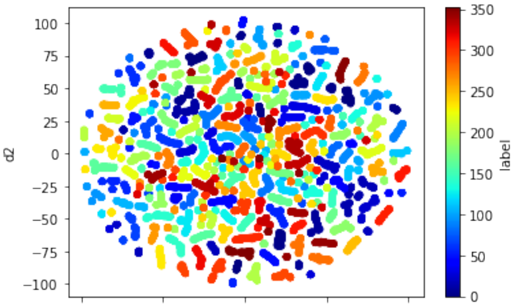

Figure 10: t-SNE Visualization of Clusters with Prim Based Clustering

3.2.5 DBSCAN

We find that, for DBSCAN, dimension reduction with UMAP was significantly better than PCA ($p < 0.001$). From our experiments It is better to reduce to 30 dimension rather then 5,10,20 or 40 ($p < 0.001$). Suprisingly outlires removal contribute over 6 points on avarage to the score ($p < 0.001$). We find that higher relative Epsilon value are better, specifically, 0.9 was signficantly better than 0.75, 0.5 and 0.25 ($p < 0.001$). We also find that with good epsilon value the number of neighbours defining core nodes did not change much. The v-scores with the labels was low (0.15)

but higher than everything we saw so far, indicating that the clusters are different in behaviours for different countries. With the best hyper parameters we found, DBSCAN final average Silhouette score is 0.945 a very high score.

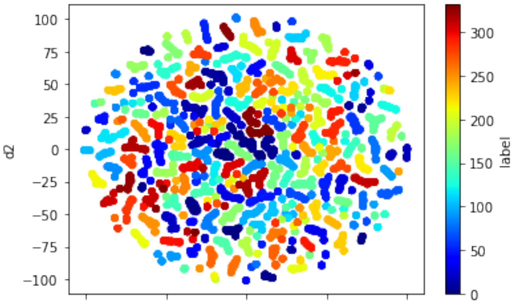

Figure 11: t-SNE Visualization of Clusters with DBSCAN

3.2.6 Online Shopping Clickstream Dataset Clustering Discussion

The Online Shopping Clickstream data containes non linear patterns that can seperated only with highly non linear algorithms such as DBSCAN and Prim. Not like the prior dataset in this one the linear GMM and K-Means preformed extremyl poorly indicating that the clustes where much more complex then before.

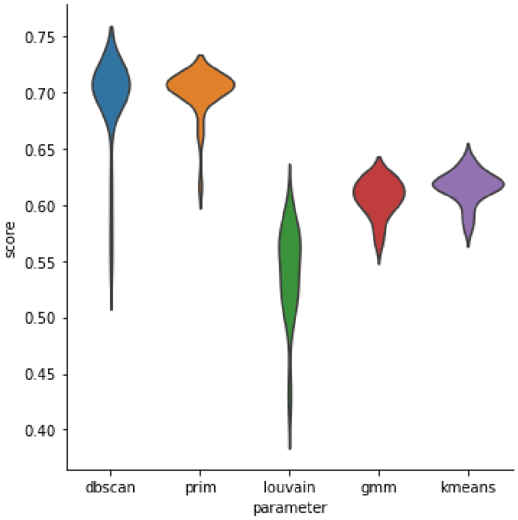

Figure 12: Silhouette Score of The Different Clustering Methods on Online Shopping Clickstream Dataset

3.3 Online Shoppers Purchasing Intention Data set

The Online Shoppers Purchasing Intention dataset contains information of 12,330 online shoppers encouners and weather they ended up with a sale. Of the 12,330 sessions in the dataset, 84.5% (10,422) were negative class samples that did not end with shopping, and the rest (1908) were positive class samples ending with shopping. The dataset consists of feature vectors belonging to the differen sessions. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period. We clustered the data with all the clustering algorithms mentioned before. For every algorithm we try different hyper parameters and report the results along with their p value.

3.3.1 K-Means

In our experiments we find that, for K-Means Clustering, dimension reduction with UMAP is significantly better than PCA ($p < 0.001$). It is better to remove outlires but without strong significance ($p = 0.3$) indicating that most of the sessions were similar to one of the clusters. With the Elbow Method we found that the 2-6 clusters are producing the best interia scores the best marginal contribution. We find that 7 clusters give the best Silhouette score ($p < 0.001$) with noticeable marginal contribution. The v-scores with the each of the labels was very low (< 0.001). indicating that the clusters were not correlated with the visitor, revenue or weekend label. With the best hyper parameters we found, K-Means final average Silhouette score is 0.31, fairly low.

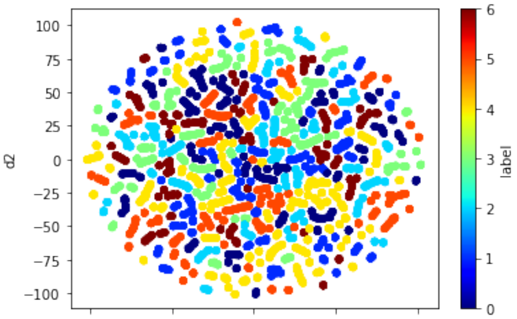

Figure 13: t-SNE Visualization of Clusters with K-Means Algorithm

3.3.2 GMM

In our experiments we find that, for GMM Clustering, dimension reduction with UMAP is significantly better than PCA or no dimension reduction at all ($p < 0.001$) and it is better to reduce to 5 or 10 dimensions rather then 20 or 30 ($p < 0.001$). It is signifcantly better to remove outlires ($p = 0.001$). With the Elbow Method we found that the 2-8 clusters are producing the best BIC scores (soft version of the interia) with the best marginal contribution. We find that 8 clusters gave the best Silhouette score ($p < 0.001$) and that the difference between the numbers was significant by the Scheffe test. .We find that the Tied and Diag covariance types are better than the Spherical ($p < 0.001$) but there is no difference bwetween them by the Scheffe test. The v-scores with the labels was very low (< 0.01) indicating that the clusters were not correlated with visitor revenue or the weekend labels. With the best hyper parameters we found, GMM final average Silhouette score is 0.29 which is quite low.

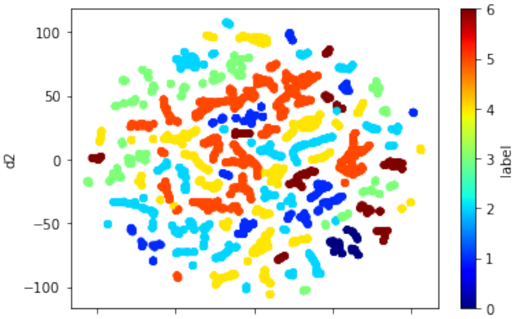

Figure 14: t-SNE Visualization of Clusters with GMM Algorithm

3.3.3 Louvain

In our experiments we find that, for Louvain, dimension reduction with UMAP was slightly better than PCA with no statistical significance. There was major difference between the number of dimensions $(p < 0.001)$ so 20 dimension is better then 5 10 or 15 dimensions with strong significance by the Scheffe test $(p < 0.001)$. It is slightly better to remove outlires but with statistical significance. We find that the Dugue or Newman modularity definition yield signficantly better results than Potts $(p < 0.001)$ but with no difference between them. The v-scores with the labels is very low (< 0.01) indicating that the clusters were not correlated with the visitor revenue or the weekend labels. With the best hyper parameters we found, Louvain final average Silhouette score is 0.35.

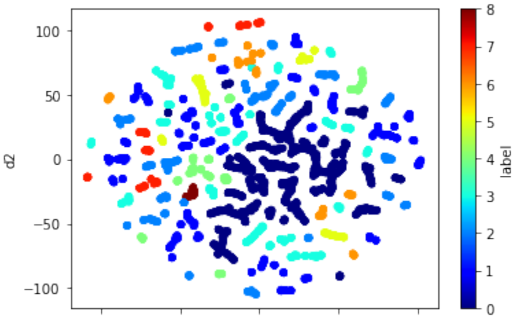

Figure 15: t-SNE Visualization of Clusters with Louvain Algorithm

3.3.4 Prim

In our experiments we find that, for Prim Based Clustering, dimension reduction with UMAP is significantly better by large margin than PCA $(p < 0.001)$. It is better to reduce to 5 dimension rather then 10,15,20,30 or 40 but without strong statistical significance. Outlires removal made the algorithm preform much better $(p < 0.001)$. Higher cutoffs preformed better $(p < 0.001)$ indicating that the clusters where separated by large gaps. The v-scores with the labels is low $< (0.04)$ indicating that the clusters were not correlated with the visitor revenue or the weekend labels. With the best hyper parameters we found, Prim final average Silhouette score is 0.67.

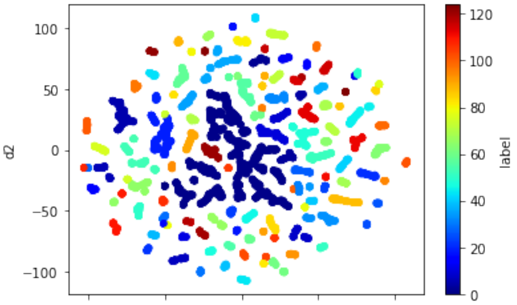

Figure 16: t-SNE Visualization of Clusters with Prim Based Clustering

3.3.5 DBSCAN

We find that, for DBSCAN, dimension reduction with UMAP was significantly better than PCA $(p < 0.001)$. From our experiments It is better to reduce to 5 dimensions rather then 10,20, 30 or 40 $(p < 0.001)$. Suprisingly, again, outlires removal contribute over 10 points on avarage to the score $(p < 0.001)$. We find that 0.25 as the relative Epsilon value was better with 4 neighbours $(p < 0.01)$ indicating that the clusters were not seprertad by large gaps but were dense . We also find that with high epsilon value the number of neighbours defining core nodes did not change much. The v-scores with the visitor revenue or the weekend labels is low $< (0.04)$. With the best hyper parameters we found, DBSCAN final average Silhouette score is 0.665 .

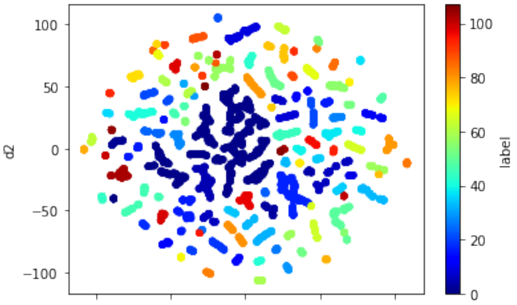

Figure 17: t-SNE Visualization of Clusters with DBSCAN

3.3.6 Online Shoppers Purchasing Intention Data set Clustering Discussion

The Online Shoppers Purchasing Intention dataset clusters, like before, better with non-linear clustering methods, indicating as before that the dataset is quite complex and non contain non-linear patterns that unite sub groups in the the data. The relatively better success of Prim algorithm $(p < 0.001)$ might be due to the MST part of the algorithms since it is a global operation over the data rather then a greedy one.

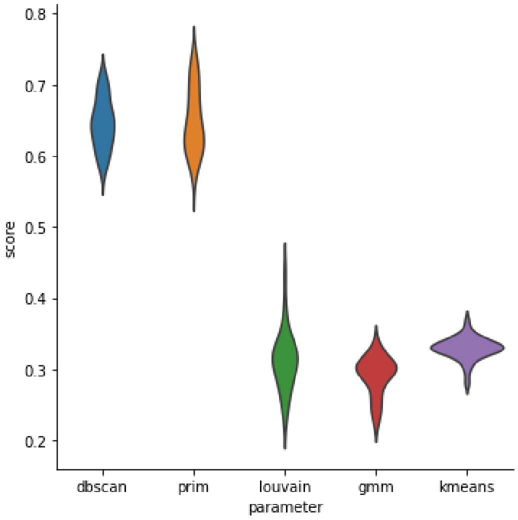

Figure 18: Silhouette Score of The Different Clustering Methods on Online Shoppers Purchasing Intention Dataset

4 Summary

In this work we compared between different clustering algorithms in different dataset, even though the datasets were very different than each other we can recognize some mutual trends. The first is that DBSCAN and Prim based clustering preformed better in all the datasets. The second is that in vast majority of the cases the clusters did not correlate with the external labels indicating that they might capture different, harder to interpret, patterns in the data.

References

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, Oct 2008. ISSN 1742-5468. doi: 10.1088/1742-5468/2008/10/p10008. URL http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

M. Ester, H. Kriegel, J. Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, 1996.

James B. MacQueen. Some methods for classification and analysis of multivariate observations. 1967.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.

Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Series 6*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720. URL https://www.tandfonline.com/doi/abs/10.1080/14786440109462720.

Andrew Rosenberg and Julia Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 410–420, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/D07-1043.

Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53 – 65, 1987. ISSN 0377-0427. doi: https://doi.org/10.1016/0377-0427(87)90125-7. URL http://www.sciencedirect.com/science/article/pii/0377042787901257.

L.J.P. van der Maaten and G.E. Hinton. Visualizing high-dimensional data using t-sne. 2008.

Jake VanderPlas. mst$_clustering : Clusteringviaeuclideanminimumspanningtrees.$*Journal of Open Source Software*, 1 $(1) : 12, 2016. doi : .$ URL https://doi.org/10.21105/joss.00012.

A Appendix

You may include other additional sections here.