**Methodology Summary of Wholesale Customers Clustering**

*Related Work*

Riyo et al. (2021) [5] evaluated K-means algorithm by elbow method (3 clusters) and Sihouette score (4 clusters), based on which a decision tree based explainability was introduced. Ilung and Geoff (2015) [4] used Davies-Bouldin (DB) index and returned 3 as the optimal number of clusters. Except K-means, A. S. M. Shahadat Hossain (2017) [2] applied DBSCAN (Density-Based Spatial Clustering of Application with Noise) algorithms with varying radius of neighborhoods and number of minimum points needed to be considered as a cluster, the combo of 0.15 & 4 outperformed others.

*Methodology*

Standardization and checking outliers (15% winsorization)

K-means, with Elbow Method (curve of WSS vs #), Silhouette Analysis, Davies-Bouldin (DB) index.

Visualization of Dimensionality Reduction

**References**

1. R. Bajaj, G. Bathla, A. Gupta, Anurag and L. Pawar, "Optimized Ensemble Model for Wholesale Market Prediction using Machine Learning," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1164-1169, doi: 10.1109/ICESC54411.2022.9885315.
2. A. S. M. S. Hossain, "Customer segmentation using centroid based and density based clustering algorithms," 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 2017, pp. 1-6, doi: 10.1109/EICT.2017.8275249.
3. Rahul Shirole, Laxmiputra Salokhe, and Saraswati Jadhav. Customer Segmentation using RFM Model and K-Means Clustering. International Journal of Scientific Research in Science and Technology. Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com). doi : <https://doi.org/10.32628/IJSRST2183118>.
4. I. Pranata and G. Skinner, "Segmenting and targeting customers through clusters selection & analysis," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 2015, pp. 303-308, doi: 10.1109/ICACSIS.2015.7415187.
5. R. H. Khan, D. F. Dofadar and M. G. Rabiul Alam, "Explainable Customer Segmentation Using K-means Clustering," 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2021, pp. 0639-0643, doi: 10.1109/UEMCON53757.2021.9666609.
6. Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
7. Cardoso, Margarida. (2014). Wholesale customers. UCI Machine Learning Repository. <https://doi.org/10.24432/C5030X>.