PROJECT REPORT

ANALYSIS OF ONLINE STORE CONSUMER BEHAVIOUR WITH K-MEANS AND AGGLOMERATIVE CLUSTERING ALGORITHMS

VENANSIUS FORTUNATUS WIJAYA
20.K1.0001

Faculty of Computer Science
Soegijapranata Catholic University
2023

ABSTRACT

An online store or e-commerce is a business system that allows buyers and sellers to conduct electronic transactions over the internet. Along with the increasing use of the internet around the world, the e-commerce business is growing rapidly and becoming a very large industry. To find out consumer behaviour, data analysis can be done using clustering algorithms, namely K-Means compared with Agglomerative Clustering and K-Means compared with K-Means++. Through these three algorithms, the data is reduced in dimension by PCA and t-SNE methods. It is then also analysed using the attributes provided by the dataset to further determine its performance measures. The parameters that will be used for clustering are K values 4, 5, and 6 of K-Means algorithm compared to Agglomerative Clustering and the optimal K value of 4 with the initialised centroid center value for K-Means++. Then determining the distance between data using the Euclidean Distance method, while for data grouping using the Average Linkage method in the Agglomerative Clustering algorithm. Through the method, the results are Agglomerative with a K value of 4 and the best t-SNE data type because the K value is good and the data type used is very good so that the results are also good and K-Means++ with an optimal K value of 4 and the best t-SNE data type because the centroid value is well initialised so that this algorithm speeds up the clustering process rather than the usual K-Means which takes longer in the process of clustering online store consumer behaviour data.

Keyword: consumer behaviour, online store, clustering, k-means, agglomerative clustering, k-means++