ABSTRACT

With the start of the data and information age, data has become one of the most useful and desirable things. The data itself will be useful information when data processing is carried out. One example of the results of data processing in business is by creating customer segmentation, customer segmentation is useful for identifying and filtering customers with specified data. Analysis of segmentation is able to provide a more effective target market allocation, more efficient budgeting, more accurate marketing or promotional strategies, and much more. Because segmentation aims to separate customers into several groups or clusters, the clustering algorithm can be used. In this final project, customer segmentation will be carried out based on the value of income and value of expenditure, and will be grouped based on these data. The method of grouping that will be used for this research is to use the K-Means ++ algorithm which is useful for determining the clusters of each given data. In this final project research will be carried out K-Means ++ using the familiar Neo4J query known as Cypher. Neo4J itself claims that Neo4J has high performance because data processing uses graph concepts, is easy to learn, and easy to use. In this study, a comparison of K-Means ++ will also be carried out using python as a comparison of Neo4J. The results obtained in this study are that Neo4J has a good cluster value when calculated using a silhouette score, has a relatively fast execution time, and when compared to python, the cluster results obtained are not much different.

Keywords : Neo4J, K-Means, Graph Database, performance, database