

## ABSTRAK

Pemanfaatan teknologi berkembang pesat di masa sekarang ini. Kepercayaan konsumen dalam melakukan transaksi bisnis online adalah sesuatu hal yang sangat sulit diprediksi dan dikendalikan, karena perilaku dan sifat tiap konsumen yang mencari informasi di situs yang menyediakan bisnis online berbeda-beda. Tujuan dari penelitian ini adalah mengelompokkan data para pengunjung pembeli toko online berdasarkan keputusan mereka untuk membeli di toko tersebut atau tidak membeli dengan menggunakan metode *Hierarchical K-Means Clustering* dan mengevaluasi *cluster – cluster* yang terbentuk dari proses *clustering*. Data yang diperoleh dari *UCI Machine Learning Repository*. Penelitian ini akan menggunakan metode pengelompokan dengan metode *Hierarchical K-Means Clustering*. Metode ini akan mengelompokkan data berdasarkan pengunjung yang memutuskan membeli dan pengunjung yang tidak memutuskan untuk membeli. Penelitian ini menggunakan metode penyeimbangan data yang dilakukan sebanyak sepuluh kali, normalisasi *z-score* dan *min-max*, variasi PCA, variasi atribut yang terbaik dengan *Information Gain*, dan uji validasi eksternal dan internal. Hasil dari penelitian menyatakan jika penggunaan metode penyeimbangan data, penggunaan PCA, dan penggunaan *Information Gain* berpengaruh. Hasil uji validasi eksternal menyatakan data hasil *output* dikatakan memiliki akurasi yang belum baik. Hasil uji validasi internal menyatakan data hasil *output* di setiap *cluster* dinyatakan *reasonable structure* dan *strong structure*.

Kata Kunci : Pengelompokkan, penyeimbangan data, metode Hierarchical K-Means Clustering

## ABSTRACT

The use of technology is growing rapidly in the present. Consumer confidence in conducting online business transactions is something that is very difficult to predict and control, because the behavior and nature of each consumer looking for information on sites that provide online business is different. The purpose of this research is to group the data of visitors to online shop buyers based on their decision to buy at the store or not to buy using the Hierarchical K-Means Clustering method and evaluate the clusters formed by the clustering process. Data obtained from UCI Machine Learning Repository. This study will use a clustering method with the Hierarchical K-Means Clustering method. This method will group data based on visitors who decide to buy and visitors who don't decide to buy. This study uses a data balancing method performed ten times, z-score and min-max normalization, PCA variations, best attribute variations with Information Gain, and external and internal validation tests. The results of the study stated that the use of data balancing methods, PCA usage, and the use of Information Gain had an effect. The results of the external validation test state the output data is said to have accuracy that is not good. Internal validation test results state that the output data in each cluster is declared to be a reasonable structure and strong structure.

Keywords : Grouping, data balancing, Hierarchical K-Means Clustering method