

INTISARI

METODE DUA LANGKAH DALAM PENGELOMPOKKAN DATA UNTUK PENGELOMPOKAN DATA CAMPURAN BERJENIS KATEGORIK DAN NUMERIK

Oleh

Naufal Ibnu Amzani
12/336539/PA/15089

Data dengan jenis atribut campuran numerik dan kategorik semakin sering kita temukan dalam kehidupan sehari-hari. Sementara itu, algoritma analisis pengelompokan data yang sering dipakai dalam kehidupan sehari-hari kebanyakan hanya mampu menganalisis data yang berjenis kategorik saja atau berjenis numerik saja. Metode dua langkah dalam pengelompokan data untuk pengelompokan data campuran berjenis kategorik dan numerik ini merupakan salah satu metode pengelompokan data berjenis campuran tersebut dengan menggunakan *co-occurrence* sebagai dasar untuk memberikan nilai numerik pada atribut kategorik. Penggunaan *co-occurrence* dimaksudkan untuk menemukan kesamaan hubungan antar masing-masing atribut kategorik. Sehingga, semua data berjenis kategorik akan memiliki nilai numerik. Kemudian, metode ini menggunakan dua langkah pengelompokan data. Pertama metode ini mengaplikasikan analisis *cluster* hirarki aglomeratif yang kemudian hasilnya diintegrasikan dengan jumlahan item kategorik sebagai item tambahan. Metode ini menggunakan analisis *cluster* k-means sebagai algoritma pengelompokan data terakhir. Kemudian, cluster yang terbentuk kemudian dianalisis kualitasnya menggunakan entropi. Metode ini menghasilkan hasil yang baik sebagai analisis *cluster* dengan data berjenis campuran.

Kata kunci: analisis *cluster*, analisis *cluster* k-means, analisis *cluster* hirarki aglomeratif, data campuran, *co-occurrence*, entropi

ABSTRACT

CLUSTERING MIXED CATEGORICAL AND NUMERIC DATA WITH TWO-STEP METHOD OF CLUSTERING

by

Naufal Ibnu Amzani
12/336539/PA/15089

Data with mixed attributes of numerical and categorical types are increasingly common in real life. Meanwhile, clustering analysis algorithm that is often used is mostly only able to analyze data with pure categorical and numeric only. Clustering mixed categorical and numeric data with two-step method of clustering is one of the methods of clustering mixed data by using co-occurrence as a basis for assigning numerical values into categorical attributes. The use of co-occurrence is intended to find similarity of relationships between each categorical attributes. Thus, all categorical data will have numeric values. Then, this method uses two-step of clustering. First, this method applied hierarchial agglomerative clustering which is then the results integrated with the sum of categorical items as additional items. This method then uses k-means clustering analysis as the final clustering algorithm. Then, the formed clusters will be tested for its quality with entropy. This method produced good results as cluster analysis for mixed data.

Key word : *clustering analysis, k-means clustering, hierarchial agglomerative clustering, mixed attributes, co-occurrence, entropy*