

INTISARI

UKURAN JARAK BARU (*NEW DISSIMILARITY*) DALAM ALGORITMA *CLUSTERING K-MODES*

Oleh

Haqiqi Hardandy
13/348064/PA/15435

K-modes merupakan salah satu teknik dalam analisis kluster yang merupakan perluasan dari algoritma *k-means* untuk mengelompokkan data kategorik. Analisis algoritma *k-modes* memodifikasi pengukuran jarak *k-means* standar menggunakan ukuran jarak kecocokan sederhana (*simple matching dissimilarity*), namun pengukuran jarak tersebut menghasilkan kesamaan dalam kluster yang lemah karena tidak mempertimbangkan frekuensi relatif dari nilai atribut masing-masing kluster.

Penelitian ini membahas pengukuran jarak sederhana dengan pengukuran jarak yang telah diperbaharui. Pengukuran jarak baru memperhitungkan frekuensi relatif dari nilai atribut masing-masing kluster. Selain untuk memperbaharui formula menghitung jarak dalam *k-modes*, pengukuran jarak baru bertujuan membuktikan objek masuk dalam keanggotaan kluster memiliki kehomogenan yang kuat. Hasil kluster yang terbentuk lebih merepresentasikan jarak sebenarnya dan menghasilkan kesamaan di dalam kluster yang lebih kuat sehingga hasil kluster lebih akurat.

Kata kunci : algoritma *k-modes*, analisis kluster, data kategorik, *data mining*.

ABSTRACT

NEW DISSIMILARY MEASURE IN K-MODES CLUSTERING ALGORITHM

by

Haqiqi Hardandy
13/348064/PA/15435

K-modes algorithm is one of cluster analysis as an extension of k-means algorithm to classify the categorical objects. The k-modes algorithm modifies the measure of standard k-means distance using simple matching dissimilarity, but it yields such weak similarity in cluster because it does not consider the relative frequency of the attribute value in each cluster.

This research aims to describe the simple matching dissimilarity measure and the new dissimilarity measure. The new dissimilarity measure considers calculating on the relative frequency of the attribute value in each cluster. It is not only to update the formula of calculating the distance in k-modes, but also to prove that the objects entered as the member of each cluster have the strong homogeneous. This method results in a better clustering accuracy because the result represents the distance of objects to central cluster for more real and has stronger similarity in cluster.

Keywords : k-modes algorithm, cluster analysis, categorical data, data mining.