

INTISARI

ALTERNATIF METODE WARD *HIERARCHICAL CLUSTERING* MENGGUNAKAN UKURAN JARAK MANHATTAN

Oleh

Qyraturrahmani
14/364186/PA/15938

Analisis kluster merupakan metode statistika multivariat yang bertujuan untuk mengelompokkan objek-objek yang memiliki kemiripan karakteristik ke dalam suatu kluster. Kluster Hirarki merupakan salah satu teknik analisis kluster yang dilakukan tanpa mengetahui jumlah kluster yang diinginkan. Metode Ward dikenal sebagai metode terbaik dalam analisis kluster hirarki. Namun, metode ini hanya dapat digunakan untuk satu ukuran jarak yaitu jarak *squared euclidean*. Sementara ada beberapa ukuran jarak yang memiliki kelebihan, salah satunya seperti jarak Manhattan yang cenderung memiliki sensitivitas yang rendah terhadap data outlier. Oleh karena itu, dilakukan perkembangan metode Ward ini dengan ukuran jarak Manhattan dengan ketentuan tidak melanggar sifat-sifat dari metode Ward *Hierarchical Clustering*.

Melalui rekursif formula *e-distance* dengan jarak Manhattan diperoleh nilai parameter *Lance-Williams* yang sama dengan metode Ward menggunakan jarak *squared euclidean*. Hal ini menunjukkan metode Ward dengan jarak Manhattan tidak melanggar sifat dari metode Ward. Dalam aplikasinya, untuk kasus yang memiliki beberapa data outlier metode Ward dengan jarak Manhattan lebih unggul dibanding dengan jarak *squared euclidean* melalui validitas *dunn index*.

Kata kunci : *Ward Hierarchical Clustering*, *e-distance*, Manhattan, *Dunn Index*.

ABSTRACT

ALTERNATIVE WARD HIERARCHICAL CLUSTERING METHOD USING MANHATTAN DISTANCE MEASURE

By

Qyraturrahmani
14/364186/PA/15938

Cluster analysis is a multivariate statistics method aiming at agglomerating objects which have similar characteristics into one cluster. Hierarchical clustering is one of the cluster analysis performed without knowing the amount of the desired clusters. The Ward method is known as the best method in hierarchical clustering analysis. However, this method can only be used for one distance measure: squared euclidean. Meanwhile, there are several distance measures that have more advantages, one of them is Manhattan distance that has a low sensitivity to outlier data. Therefore, the Ward method development was performed using Manhattan distance measure without violate the properties of Ward Hierarchical Clustering method.

Through recursive formula e-distance using Manhattan distance, it was obtained the same Lance-Williams parameter value using Ward method with squared euclidean distance. This shows that Ward method development using Manhattan distance does not contravene the properties of it. In application, some cases that have several outlier data of the Ward method are more advanced compared to the squared euclidean distance through dunn index validity.

Keywords: *Ward Hierarchical Clustering, e-distance, Manhattan, Dunn Index.*