

Pengkajian Metode Ordering Points to Identify the Clustering Structure (OPTICS) pada Data dengan Permasalahan Missing Value dan Penerapan pada Data Real (Data Socio-Economic) = Assessment of the Ordering Points to Identify the Clustering Structure (OPTICS) Method on Data with Missing Value Problems and Applications to Real Data (Socio-Economic Data)

Geraldine Immanuel Tangyong, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920530727&lokasi=lokal>

Abstrak

Clustering merupakan salah satu metode pengolahan data yang bekerja dengan mengelompokkan objek-objek yang memiliki kemiripan ke dalam beberapa kelompok. Salah satu metode dari clustering adalah metode Ordering Points to Identify the Clustering Structure (OPTICS) yang adalah metode turunan dari metode Density Based Spatial Clustering of Application with Noise (DBSCAN). DBSCAN memiliki dua parameter, yaitu epsilon dan MinPts. Parameter ini mengakibatkan DBSCAN kurang bisa beradaptasi terhadap variasi densitas data. Karena itu, OPTICS hadir untuk mengatasi masalah tersebut dengan parameter core distance dan reachability distance. Tujuan dari penelitian ini adalah mengkaji metode OPTICS dan menerapkannya pada data sintetis dengan missing value serta pada data real, dengan data yang digunakan adalah data Socio-Economic dengan 167 observasi. Dari hasil pengkajian diketahui bahwa jika suatu data memiliki proporsi missing value lebih dari 15%, maka nilai akurasi dianggap tidak baik. Selanjutnya, dilakukan penerapan terhadap data Socio-Economic dan diperoleh 3 cluster. Cluster 1 memiliki jumlah anggota terbanyak dan cenderung memiliki skor terendah pada banyak variabel dibanding cluster lainnya dan dapat dilihat bahwa anggota dari cluster 1 banyak terdiri dari negara-negara berkembang. Cluster 2 cenderung memiliki skor tertinggi dan negara-negara di cluster 2 merupakan negara-negara maju. Cluster 3 banyak berisi negara campuran, baik negara berkembang maupun negara maju.

.....Clustering is a data processing technique that works by grouping objects that have similarities into different groups. One method of clustering is the Ordering Points to Identify the Clustering Structure (OPTICS) method, a derivative method of the Density Based Spatial Clustering of Application with Noise (DBSCAN) method. DBSCAN has two kinds of parameters, epsilon and MinPts. These parameters cause DBSCAN to not be able to adapt fully to arbitrary data forms. Therefore, OPTICS is here to solve this problem with the parameters of core distance and reachability distance. The purpose of this study is to examine the OPTICS method and apply it to synthetic with missing value and real data, using the Socio-Economic data with 167 observations. We can conclude that if a data has a missing value proportion of more than 15%, then the accuracy value is considered not good. Furthermore, the Socio-Economic data was applied and obtained 3 clusters. Cluster 1 has the largest number of members and have the lowest scores on many variables, which consists mostly of developing countries. Cluster 2 tends to have the highest score whereas most coming from developed countries. Cluster 3 contains many mixed countries, both developing and developed countries.