

Pengkajian metode User-Input-Free Density-Based Clustering (UIFDBC) pada data dengan missing values dan penerapannya pada data real = Assessment of the User-Input-Free Density-Based Clustering (UIFDBC) method on data with missing values and its application to real data

Gabriela Patricia Winny Gracia, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20525167&lokasi=lokal>

Abstrak

Clustering merupakan metode untuk mengidentifikasi kelompok natural pada data berdasarkan ukuran kemiripan seperti jarak Euclidian. Clustering bertujuan untuk mengelompokkan data, dengan kriteria observasi yang berada dalam satu klaster memiliki tingkat kemiripan yang sangat signifikan, sedangkan observasi yang berada dalam cluster yang berbeda, memiliki perbedaan yang sangat signifikan. Pada tahun 2021, Chowdhury, Bhattacharyya, & Kalita mengembangkan metode User-Input-Free Density-Based Clustering (UIFDBC) berdasarkan dari metode density-based clustering yang telah ada sebelumnya. Seperti namanya, metode UIFDBC ini tidak memerlukan input dari pengguna untuk menemukan cluster. Maka dari itu, metode UIFDBC ini berhasil menjawab permasalahan metode clustering sebelumnya yang bergantung pada input dari pengguna. Tujuan dari penelitian ini adalah untuk membahas lebih dalam terkait metode User-Input-Free Density-Based Clustering (UIFDBC), menerapkan metode UIFDBC pada data real, yaitu data konsumen kartu kredit untuk melakukan segmentasi konsumen, serta mengkaji performa metode ini pada data yang mengandung missing values di dalamnya. Dari hasil penelitian, metode UIFDBC berhasil diterapkan pada data konsumen kartu kredit, dan diperoleh sebanyak delapan cluster pengguna, dimana setiap cluster memiliki karakteristik masing-masing. Selain itu, dari hasil pengkajian metode UIFDBC terhadap data dengan missing values diketahui bahwa performa metode UIFDBC dinilai cukup baik untuk proporsi missing values 5%. Namun perlu menjadi catatan bahwa data hasil dari setiap iterasi akan bersifat acak, dikarenakan metode UIFDBC sangat bergantung pada densitas data, sedangkan densitas data bergantung pada missing values yang mana dibangkitkan secara acak sepenuhnya.

.....Clustering is a method to identify natural groups in data based on similarity measures such as Euclidian distance. Clustering aims to group data, with the criteria for observations in one cluster having a very significant level of similarity, while observations in different clusters have very significant differences. In 2021, Chowdhury, Bhattacharyya, & Kalita developed the User-Input-Free Density-Based Clustering (UIFDBC) method based on the previous density-based clustering method. As the name suggests, this UIFDBC method does not require input from the user to find the cluster. Therefore, this UIFDBC method has succeeded in answering the problems of the previous clustering method which depended on input from the user. The purpose of this study is to discuss more deeply the User-Input-Free Density-Based Clustering (UIFDBC) method, to apply the UIFDBC method to real data, namely credit card consumer data to segment consumers, and to examine the performance of this method on data containing missing values in it. From the results of the study, the UIFDBC method was successfully applied to credit card consumer data, and obtained as many as eight user clusters, where each cluster has its own characteristics. In addition, from the results of the study of the UIFDBC method on data with missing values, it is known that the performance of the UIFDBC method is considered quite good for the proportion of missing values 5%. However, it should

be noted that the resulting data from each iteration will be random, because the UIFDBC method is very dependent on data density, while data density depends on missing values which are generated completely randomly.