

INTISARI

PENERAPAN METODE ISOLATION FOREST PADA DETEKSI DAN PENANGANAN OUTLIER DALAM ANALISIS KLASTER MENGGUNAKAN ALGORITMA K-MEANS, K-MEANS++, DAN K-MEDOIDS

oleh

Ilham Rizky Fauzi

21/477229/PA/20640

Algoritma *K-Means*, *K-Means++*, dan *K-Medoids* merupakan algoritma yang paling banyak digunakan dalam analisis kluster karena kemudahan dalam penggunaan dan efisiensi dalam komputasi. Akan tetapi, ketiga algoritma tersebut masih dapat dioptimumkan kekompakan dan pemisahan hasil klasternya dengan menggunakan deteksi dan penanganan *outlier* yang tepat. Metode *Isolation Forest* merupakan metode yang dapat mendeteksi *outlier* atau anomali dengan efektif dan efisien. Metode tersebut mengidentifikasi *outlier* sebagai data yang lebih mudah untuk diisolasi dari mayoritas data yang normal dengan berbasis pembentukan hutan isolasi yang terdiri dari beberapa pohon isolasi yang kemudian menghitung skor anomali yang didapatkan setiap titik data. Pada penelitian ini, data yang digunakan untuk studi kasus yaitu data jumlah tenaga kesehatan berdasarkan kategori menurut kabupaten/kota di Indonesia tahun 2021. Berdasarkan hasil analisis yang diperoleh, algoritma *K-Means++* dengan deteksi *outlier* menggunakan *Isolation Forest* dan pemisahan antara data normal dengan data *outlier* menghasilkan 4 kluster dengan pengukuran nilai *Dunn Index* yaitu 0,05373902. Kemudian, dalam profilisasi kluster final, data *outlier* didefinisikan sebagai kluster ke-5.

Kata kunci: *Isolation Forest*, Deteksi *Outlier*, Analisis Kluster, *K-Means*, *K-Means++*, *K-Medoids*

ABSTRACT

***APPLICATION OF THE ISOLATION FOREST METHOD FOR OUTLIER
DETECTION AND HANDLING IN CLUSTER ANALYSIS USING
K-MEANS, K-MEANS++, AND K-MEDOIDS ALGORITHMS***

by

Ilham Rizky Fauzi

21/477229/PA/20640

The K-Means, K-Means++, and K-Medoids algorithms are the most widely used algorithms in cluster analysis because of their ease of use and efficiency in computing. However, the three algorithms can still optimize the compactness and separation of cluster results by using proper outlier detection and handling. The Isolation Forest method is a method that can detect outliers or anomalies effectively and efficiently. This method identifies outliers as data that is easier to isolate from the majority of normal data based on the formation of an isolation forest consisting of several isolation trees which then calculate the anomaly score obtained by each data point. In this study, the data used for the case study was data on the number of health workers based on categories according to districts/cities in Indonesia in 2021. Based on the analysis results obtained, the K-Means++ algorithm with outlier detection using Isolation Forest and separation between normal data and outlier data produced 4 clusters with a Dunn Index value measurement of 0.05373902. Then, in the final cluster profiling, outlier data was defined as the 5th cluster.

Keywords: Isolation Forest, Outlier Detection, Cluster Analysis, K-Means, K-Means++, K-Medoids