

INTISARI

SEGMENTASI PEMAIN SEPAK BOLA PADA DATA CAMPURAN KATEGORIK DAN NUMERIK DENGAN METODE ENSEMBEL *FUZZY C-* *MEANS* DAN *K-MODES* (STUDI KASUS: PENGELOMPOKAN PEMAIN SEPAK BOLA EROPA)

oleh

Muhammad Gufron Arief

18/427702/PA/18662

Analisis kluster merupakan salah satu metode pengelompokan data berdasarkan kemiripan dari suatu data. Pengelompokan pada data campuran numerik dan kategorik menjadi lebih kompleks karena perbedaan tipe data sehingga harus menggunakan ukuran kemiripan yang berbeda. Salah satu metode pengelompokan data campuran adalah metode ensemble. Metode *ensemble* melakukan pengelompokan dengan memisahkan data terlebih dahulu sehingga tidak menghilangkan informasi penting pada data. Pada penelitian ini, digunakan metode *ensemble fuzzy c-means* dan *k-modes*. Metode *fuzzy c-means* mengelompokkan data numerik dengan melihat jarak Euclidean antara objek dengan titik pusat kluster. Metode *k-modes* mengelompokkan data kategorik dengan melihat jarak Hamming antara objek dengan titik pusat kluster. Pada penelitian ini digunakan beberapa validasi kluster untuk menentukan jumlah kluster yang optimum pada tiap metode. Metode *ensemble fuzzy c-means* dan *k-modes* diaplikasikan untuk melihat segmentasi pada data pemain sepak bola di *English Premier League*, *La Liga*, *Bundesliga*, *Seria A*, dan *Ligue 1* musim 2021/2022. Didapatkan 19 segmen pemain sepak bola di lima liga Eropa dengan masing-masing karakteristiknya.

Kata kunci : analisis kluster, data campuran, *ensemble*, *fuzzy c-means*, *k-modes*

ABSTRACT

***FOOTBALL PLAYER SEGMENTATION ON MIXED DATA NUMERIC AND
CATEGORICAL WITH ENSEMBLE FUZZY C-MEANS AND KMODES METHOD
(CASE STUDY: GROUPING EUROPEAN FOOTBALL PLAYER)***

by

Muhammad Gufron Arief

18/427702/PA/18662

Cluster analysis is one of the methods of grouping data based on the similarity of a data. Grouping on numerical and categorical mixed data becomes more complex due to differences in data types so it must use different similarity measures. One of the methods of grouping mixed data is ensemble method. The ensemble method performs grouping by splitting the data, so as not to eliminate important information on the data. In this study, the ensemble fuzzy c-means and k-modes methods were used. The fuzzy c-means method group numerical data by Euclidean distance between the object and the center point of the cluster. The k-modes method group categorical data by Hamming distance between the object and the center point of the cluster. In this study, several cluster validations were used to determine the optimum number of clusters in each method. The ensemble fuzzy c-means and k-modes methods are applied to see the segmentation on football player data in English Premier League, La Liga, Bundesliga, Serie A, and Ligue 1 season 2021/2022. There are 19 segments of football players in five European leagues with each of their characteristics.

Keywords : cluster analysis, mixed data, ensemble, fuzzy c-means, k-modes