

INTISARI

PENINGKATAN PERFORMA ANALISIS KLASIFIKASI MENGGUNAKAN METODE *BALANCED ONLINE K-MEANS++*

Oleh

Fikri Kamaluddin

21/476787/PA/20605

Di era big data saat ini, performa analisis klasifikasi menjadi aspek yang sangat krusial dalam berbagai aplikasi, mulai dari bisnis, medis hingga penelitian ilmiah. Kualitas data yang digunakan dalam analisis sangat mempengaruhi hasil klasifikasi. Data heterogen yang terdiri dari berbagai jenis, format, dan distribusi menimbulkan tantangan tersendiri dalam analisis klasifikasi. Data heterogen dapat menyebabkan algoritma klasifikasi mengalami kesulitan dalam menemukan pola yang konsisten dan akurat, sehingga menurunkan performa analisis. Penelitian ini mengusulkan sebuah pendekatan untuk meningkatkan performa analisis klasifikasi dengan menggunakan metode *Balanced Online K-Means++*. Metode ini bertujuan untuk mengklaster data heterogen terlebih dahulu, sehingga setiap kluster yang dihasilkan memiliki karakteristik yang lebih seragam. Dengan demikian, proses klasifikasi diharapkan menjadi lebih efektif karena setiap kluster dianalisis secara terpisah, memungkinkan algoritma klasifikasi untuk bekerja pada data yang lebih homogen. Pada studi kasus ini dilakukan penerapan *Balanced Online K-Means++* pada analisis klasifikasi jenis-jenis diabetes menggunakan data *Diabetes Dataset*. Hasil eksperimen menunjukkan bahwa metode yang diusulkan mampu meningkatkan performa klasifikasi dibandingkan dengan metode konvensional yang langsung menerapkan algoritma klasifikasi pada data heterogen. Selain itu, metode ini juga menunjukkan peningkatan dalam hal efisiensi waktu komputasi.

Kata kunci: klasifikasi, data heterogen, *Balanced Online K-Means++*

ABSTRACT

IMPROVING CLASSIFICATION ANALYSIS PERFORMANCE USING BALANCED ONLINE K-MEANS++ METHOD

By

Fikri Kamaluddin

21/476787/PA/20605

In today's big data era, the performance of classification analysis has become a crucial aspect in various applications, ranging from business, medical to scientific research. The quality of the data used in the analysis greatly affects the classification results. Heterogeneous data consisting of various types, formats, and distributions poses its own challenges in classification analysis. Heterogeneous data can cause classification algorithms to have difficulty in finding consistent and accurate patterns, thus degrading analysis performance. This research proposes an approach to improve the performance of classification analysis by using the Balanced Online K-Means++ method. This method aims to cluster heterogeneous data first, so that each cluster produced has more uniform characteristics. Thus, the classification process is expected to be more effective as each cluster is analyzed separately, allowing the classification algorithm to work on more homogeneous data. In this case study, Balanced Online K-Means++ is applied to the classification analysis of diabetes types using Diabetes Dataset data. Experimental results show that the proposed method is able to improve classification performance compared to conventional methods that directly apply classification algorithms to heterogeneous data. In addition, the method also shows improvement in terms of computational time efficiency. Translated with DeepL.com (free version).

Keywords: classification, heterogeneous data, Balanced Online K-Means++