

ABSTRAK

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *K – Means* dan *BIRCH* dalam melakukan *clustering* pada data penjualan sepeda berskala besar, serta mengidentifikasi karakteristik *cluster* yang terbentuk dan mengevaluasi efisiensi kedua algoritma. Penelitian menggunakan 50.000 data transaksi penjualan sepeda yang berasal dari dataset *Bike Sales 100K* dari Kaggle. Data diproses melalui tahap *preprocessing*, kemudian dilakukan *clustering* menggunakan lima atribut, yaitu *Price*, *Quantity*, *Bike_Model*, *Store_Location*, dan *Customer_Gender*. Kinerja kedua algoritma dievaluasi menggunakan *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan waktu komputasi. Hasil penelitian menunjukkan bahwa kedua algoritma mampu membentuk empat *cluster* dengan distribusi data yang relatif seimbang. Algoritma *K – Means* memperoleh *Silhouette Score* sebesar 0,1868, DBI sebesar 1,928, dan waktu komputasi 0,708 detik, sedangkan algoritma *BIRCH* memperoleh *Silhouette Score* sebesar 0,1709, DBI sebesar 2,043, dan waktu komputasi 0,653 detik. Berdasarkan hasil evaluasi, algoritma *K – Means* menghasilkan kualitas *cluster* yang lebih baik berdasarkan nilai *Silhouette Score* dan DBI, sedangkan *BIRCH* memiliki waktu komputasi yang sedikit lebih cepat. Hasil *clustering* juga menunjukkan karakteristik yang berbeda pada setiap *cluster* berdasarkan harga, jenis sepeda, lokasi, dan gender pelanggan, sehingga dapat dimanfaatkan sebagai pendukung analisis segmentasi pasar dan pengambilan keputusan berbasis data.

Kata Kunci: Data Mining, Clustering, K – Means, BIRCH, Segmentasi Pasar

ABSTRACT

This research aims to compare the performance of the K-Means and BIRCH algorithms in clustering large-scale bicycle sales data, as well as to identify the characteristics of the resulting clusters and evaluate the efficiency of both algorithms. The study used 50,000 bicycle sales transaction records obtained from the Bike Sales 100K dataset available on Kaggle. The data were processed through a preprocessing stage, followed by clustering using five attributes: Price, Quantity, Bike_Model, Store_Location, and Customer_Gender. The performance of both algorithms was evaluated using the Silhouette Score, Davies–Bouldin Index (DBI), and processing time. The results show that both algorithms successfully formed four clusters with a relatively balanced data distribution. The K-Means algorithm achieved a Silhouette Score of 0.1868, a DBI of 1.928, and a processing time of 0.708 seconds, while the BIRCH algorithm achieved a Silhouette Score of 0.1709, a DBI of 2.043, and a processing time of 0.653 seconds. Based on the evaluation results, the K-Means algorithm produced better cluster quality, as indicated by a higher Silhouette Score and a lower DBI, whereas the BIRCH algorithm required slightly less processing time. The clustering results also revealed distinct characteristics for each cluster based on dominant price, bicycle type, location, and customer gender, providing valuable insights to support market segmentation analysis and data-driven business decision-making.

Keywords: Data Mining, Clustering, K-Means, BIRCH, Market Segmentation