ABSTRACT

This research focuses on the optimization of K-Means algorithm in abalone data clustering using Cuckoo Search Algorithm (CSA). The main objective of this research is to improve clustering accuracy and execution time efficiency. The dataset used consists of 4177 entries with 8 abalone physical features. The preprocessing methods used include data exploration analysis, outlier detection, and feature scaling using Robust Scaler, MinMax Scaler, and Standard Scaler. Three clustering models were evaluated namely K-Means, CSA K-Means, and CSA Euclidean. The results showed that the use of CSA in determining the initial centroid improved the clustering accuracy, with the silhouette score value reaching 0.861 in cluster 2 after Robust Scaler, compared to standard K-Means which only reached 0.444. In addition, CSA K-Means shows a significant performance improvement in terms of execution time, with an execution time of 0.025 seconds for CSA K-Means on cluster 2 after Robust Scaler, compared to standard K-Means which requires 0.093 seconds. Overall, this study shows that the use of CSA in determining the initial centroid in the K-Means algorithm can improve clustering accuracy and execution time efficiency, providing better results compared to the standard K-Means method.

Keywords:

abalone, cuckoo search algorithm, clustering, feature scaling, k-means