

ABSTRACT

The continuous advancement of technology has significantly increased the ease and speed with which people can access information. At the same time, social media has emerged as a powerful platform for efficient information dissemination and expanded its role in public expression. Twitter, in particular, is a social media platform widely used by Indonesians to voice their aspirations and criticize the local and central government. This research aims to analyze public opinion on issues or problems related to the Bandung City government on a daily basis to be grouped based on the similarity of patterns that appear to provide new information through clustering methods. The dataset consists of comments from social media accounts that use Sundanese language and are related to Bandung. The total number of comments successfully taken to be used as a dataset is 3,605 data, then the data will be processed by applying clustering algorithms, with the cleaning and feature extraction stages first. K-Means clustering is the algorithm used in this research with the final silhouette score obtained as much as 0.792, followed by five clusters produced. The five cluster topics were successfully identified for each cluster, cluster 0 for convenience and traffic congestion, cluster 1 for environmental issues and city infrastructure, cluster 2 for social issues and public perception, cluster 3 for transportation and city noise, and cluster 4 showing welfare and social activities. This method is implemented using Python as the language for building machine learning models.

Keywords: social media, data scraping, clustering, feature extraction, k-means clustering