ABSTRACT

With the development of social media today, the Bandung City government realizes the importance of the role of social media in channeling public aspirations, especially Instagram. Instagram is a medium for people to express their opinions by commenting. This study aims to cluster comment data from Instagram social media related to Bandung City using the Density-Based Spatial Clustering of Application with Noise (DBSCAN) algorithm. Data was collected through web scraping method using Apify tool from news accounts in Bandung City, resulting in a total of 1,949 comments used as dataset. Next, the text preprocessing process is carried out, such as data cleaning and tokenization, then the comments are extracted features into numerical form using the TF-IDF and LDA methods. After that, the comments are clustered using the DBSCAN clustering algorithm with the final result of the silhoutte score obtained of 0.82 and get 5 clusters. After getting 5 clusters, then the cluster is identified, the topics of each cluster are transportation and infrastructure, weather and daily activities, work and social issues, family and social life, and environmental conditions and health. After clustering, sentiment analysis was conducted to determine whether comments were positive, negative, or neutral. There were 708 negative sentiment comments, 650 neutral sentiment comments, and 429 positive sentiment comments. The research results are visualized through an interactive dashboard that displays the clusters and sentiment analysis of the analyzed comments.

Keywords: DBSCAN, clustering, sentiment analysis, feature extraction, Instagram, web scraping.