

ABSTRACT

Along with the development of information and communication technology, people can easily express their responses and opinions on online transportation services through social networks, especially Instagram. We Are Social 2020 data says there are 79% of the Indonesian population actively using Instagram social media, with the large number of active users producing a lot of data such as community data on online transportation services submitted through the comments column on Instagram, with a variety of positive and negative sentiments. This community sentiment can certainly be information for program implementers for their policies and services.

In this research, sentiment grouping will be carried out on each positive, negative, and neutral sentiment data using the DBSCAN (Density-Based Spatial Clustering of Applications With Noise) algorithm. The main purpose of this clustering is to classify public opinion based on similar characteristics or meaning in writing on between these opinions to determine positive, negative, and neutral based on comments on Instagram social media. By doing preprocessing stages such as tokenize, stopword, and stemming, then word weighting is carried out using TF-IDF to be able to group opinions. From the Clustering results, it is obtained that the positive, negative, and neutral datasets were tested with a minimum sample value range of 10-50 and an eps value of 0.1-1.0 by producing different silhouette values. However, the best value from the third dataset is obtained at the input value of eps = 1.0 and the inputmin sample value = 10, for the positive dataset the silhouette coefficient value is 0.7800973549904059, for the neutral dataset the silhouette coefficient value is 0.7526159947007542, for the negative dataset the silhouette value is the coefficient is 0.8047251594403672. Then the data visualization of the results of the topic grouping will be shown on a web-based software which is also designed in this final project research.

Key Word: *Clustering, TF-IDF, Preprocessing, silhouette coefficient.*