

ABSTRACT

Air quality in Indonesia continues to be a major concern due to high levels of pollution, one of which is caused by emissions from coal-fired power plants (CFPP). This issue has triggered public concern, which can be observed through social media and news media platforms. Therefore, this study is essential to understand how public perception is formed and develops regarding CFPP-related issues in Indonesia. The aim of this research is to categorize public discourse and identify the dominant emerging themes. The method used is the Knowledge Discovery in Databases (KDD), which consists of data selection, data processing, data transformation, data mining, and evaluation/interpretation. The K-Means algorithm is applied for clustering, followed by topic modeling to identify the main topic in each cluster. The data for this study were collected from Twitter and national online news sources during the 2023–2024 period. The results show that the K-Means algorithm successfully grouped the data into three main clusters with a silhouette score of 0,6380. The findings indicate that public opinion is divided into three main themes, concern about the environmental impact of CFPP, attention to technical aspects and operational innovations, and discussions on alternative energy sources for electricity generation. These findings demonstrate that machine learning can systematically map public perceptions related to CFPP.

Keywords: CFPP, K-Means, machine learning, social media, topic modeling