ABSTRACT

Information grew increasingly with the rise of internet technology, leading to a significant increase in the amount of news. The manual categorization of news, in which editors read each news article and assign categories based on subjective observation, became inefficient. The large volume of published news and time constraints added challenges in maintaining consistency in categorization, especially for articles with similar content. This research proposed the K-Means method to perform news headline clustering, which addressed the limitations of manual categorization. The K-Means method was chosen because it was more flexible in determining the number of clusters from the outset, making it more efficient at handling large datasets compared to the Agglomerative Hierarchical Clustering (AHC) method. TF-IDF method was used for word vectorization which assigned higher weights to more relevant and unique words in each document. The optimal number of clusters was determined using the Elbow Method, which helped identify the point by minimizing distortion. The results of the research showed that the combination of the K-Means method with TF-IDF and the Elbow Method produced efficient and high-quality news clustering, with clear and structured cluster patterns. The evaluation conducted on new data further reinforced these findings, resulting in clusters of good quality.

Kata Kunci: *clustering, Elbow Method, K-Means, news*