

Abstrack

Currently the spread of Hoax information is very common in various media. One of the media is news. Much research has been done with the method of grouping the news using K-Means, but there is more research using original news data. In this study, there is the K-Means algorithm used. This study uses a 1,231 record dataset containing Hoaxes and used 16 groups consisting of advice fees, bogus warnings, death, facebook scams, fake news, faux images, health, general, legend, malware, misleading, lottery, scams, special features. Then the dataset will be pre-processed first. For the preprocessing stage is divided into several parts, namely clear url, clear symbols, tokenisasi, stop words, normalization, garbage. Then the results of preprocessing are processed in the word weighting process. After passing the word weighting stage, it can be grouped using the method of k-means clustering. In this study conducted by entering the number of clusters that vary. From the results of validation by entering a different cluster value, the best cluster value has been obtained, namely $K = 16$ with the value of SSE (Sum Square Error) generated 1058.0332293029496.

Keywords : *E-mail, Hoax, Clustering, Algoritma K-Means, SSE(Sum Square Errorr)*.