

ABSTRACT

Clickbait Detection for News Article in Bahasa Indonesia using Article Informality

Hilda

17/408354/PA/17661

Due to the flourishing growth of online media, content creators are triggered to create content that is attractive and that they employ several techniques such as the usage of catchy headlines to lure the readers to read their news article. However, these pages might be of low-quality and tend to under-deliver the content promised in the headline. Such that it may utilize some kind of formatting on its headline as well as engaging structures for its content that it appears less formal than professionally written news. These pages are then called clickbait. It is undeniable that clickbait is usually seen through its headline, this does not cease the possibility that the content might show some signals towards its classification.

This research aims to utilize the content of the article in detecting clickbait for news article in Bahasa Indonesia. Features from article along with features from headline will be extracted. In which feature from article represent several informality and readability measures. Feature groupings would be used to measure the performance of different feature combinations across two machine learning algorithm, SVM and Logistic Regression. Feature groupings that are used includes *Headline + Article + Similarity*, *Headline + Article*, and *Headline*.

Best model for this research thus is achieved by *Headline + Article + Similarity* features utilizing SVM which obtained the highest performance across the metrics used. Achieving average accuracy of 92.26%, average precision of 92.73%, average recall of 86.74%, and average F-1 Score of 89.49%.

Keywords: Clickbait Detection, Bahasa Indonesia, Support Vector Machine, Machine Learning