

ABSTRACT

We present our work in the area of sentiment analysis for Indonesian language. We focus on building automatic semantic orientation using available resources in Indonesian. In this research we used Indonesian corpus that contains 9 million words from kompas.txt and tempo.txt that manually tagged and annotated with of part-of-speech tagset. And then we construct a dataset by taking all the adjectives from the corpus, removing the adjective with no orientation. The set contained 923 adjective words.

This systems will include several steps such as text pre-processing and clustering. The text pre-processing aims to increase the accuracy. And finally clustering method will classify each word to related sentiment which is positive or negative. With improvements to the text preprocessing, can be achieved 72% of accuracy.

Keywords: sentiment analysis, Indonesia language, automatic semantic orientation, adjective words.