

# **The Accuracy Improvement of Text Mining Classification on Hospital Review through Alteration in The Preprocessing Stages**

**Triyas Hevianto Saputro, Arief Hermawan**  
*Department of Master of Information Technology*  
*University of Technology Yogyakarta*  
*North Ringroad St., Jombor Sleman Yogyakarta*  
*E-mail: triyasheviantosaputro@student.uty.ac.id*

## **ABSTRACT**

Sentiment analysis is part of text mining which is used to extract information from a sentence or document. This study focuses on the problem of text classification for sentiment analysis of hospital reviews through google maps review. The collection of text data obtained still contains many words that use non-standard language. This non-standard word becomes a problem at the pre-processing stage. The purpose of this study is to increase the accuracy of the hospital review text classification model for sentiment analysis modeling. In this study, using pre-processing technique scenarios: (1) The first scenario uses tokenization, punctuation and number removal, lower case (case folding) techniques, stemming, special character removal, and stop words removal. (2) The second scenario uses tokenization, punctuation and number removal, lower case (case folding) pre-processing techniques, spelling correction, stemming, special character removal, and stop words removal. (3) The third scenario uses tokenization, punctuation and number removal, lower case (case folding) pre-processing techniques, slang words, stemming, special character removal, and stop words removal. Then this study will also compare each of these scenarios without involving stop words removal. The results of this study obtained that the accuracy scores for the first, second and third preprocessing scenarios were 80.0%, 81.7%, and 76.7%, respectively. From this scenario, it can be concluded that spelling correction can help in increasing accuracy. But for slang words tend to give a decrease in accuracy. Then this study tested each scenario of the pre-processing stage without involving the stop words removal technique. The accuracy scores for the first, second and third scenarios without involving stopwords removal are 83.3%, 88.3% and 83.3%, respectively. For this case study without using stop words removal can provide increased accuracy.

**Keywords:** increased accuracy, preprocessing, text classification