Abstract

Product review is one of the most important sources for consumers in finding the most suitable products for their needs. However, there is a chance a reviewer has other intentions other than providing an honest review, such as advertising the brand or other brands. A review that does not contain any information related to the product aspects/features could be considered spam. This paper presents our work on spam review detection, specifically on the beauty product domain. We used SVM and Logistic Regression classifier and the following features: the review sentiment, product-related features, and review-centric features extracted from the reviews. We classify the beauty product review text as spam and non-spam reviews. The experimental result shows that the best accuracy, 81% was obtained when we used the sentiment and review-centric features with the SVM algorithm.

Keywords: Spam Review Detection, Support Vector Machine, Logistic Regression, Indonesian