Sentiment analysis based on probabilistic classifier techniques in various Indonesian review data

ABSTRACT

Sentiment analysis is the field in data science to achieve a broader holistic view of users' needs and expectations. Indonesian user opinions have the potential to manage to be valuable information using sentiment-analysis tasks. One of the most supervised-learning techniques used in Indonesian sentiment analysis is the Naïve Bayes classifier. The classifier can be optimized and tuned in various models to increase the sentiment analysis model performance. This research aims to examine the performance of various Naïve Bayes models in sentiment analysis, especially when implemented in small datasets to handle overfitting problems. Four different Naïve Bayes models used are Gaussian, Multinomial, Complement and Bernoulli. We also analyze the effect of various pre-processing techniques on the models' performance. Moreover, we build the first fashion dataset from the Indonesian marketplace which has a unique character compared to the datasets from other domains. Finally, we also use various datasets in the experiment to test the Naïve Bayes models' performance. From the experimental results, Complement Naïve Bayes is superior to other models, especially in handling overfitting with an F1-score of approximately 0.82.