Abstract

Short message service (SMS) is one of the communication services to send and receive short text message on your mobile phone. SMS is still used in daily life because of the ease of use, its simplicity, fast, and cheap. The increasing use of SMS is used by many people to gain some benefits, one of which is to send spam via SMS. SMS *spam that circulating widely in community tend to have a certain pattern. It's just common people do not know much about it, so that they got misguided by those SMS. Of this condition, then in this final project, SMS spam corpus based on Indonesian was made to learn about SMS spam filtering, which also in the same time will be tested it's performance for SMS spam filtering system.*

SMS Data collected by crowdsourcing models involving mobile phone users. SMS Data collected linguistic approach is validated by Indonesian experts to determine the class spam and ham, resulting in a corpus SMS spam. Corpus SMS spam will be tested its performance by using web-based Decision Tree classification program and the tool RapidMiner using Naïve Bayes method. The test resulted in the conclusion that the SMS data in the corpus has a strong character suit each class, with an average of 92.91% accuracy of both testing.

Keywords: SMS *spam*, SMS *ham, Indonesian, crowdsourcing, corpus, unique term, RapidMiner*