

ABSTRAK

Abstrak— Faktor utama peningkatan jumlah spam adalah kemampuan untuk mengirimkan pesan kepada banyak penerima sekaligus (Belém & Duarte-Figueiredo, 2011). Selain didukung oleh perkembangan teknologi, jumlah pengguna telepon seluler menjadi faktor utama penyebaran SMS spam. Berdasarkan data dari *International Telecommunication United (ITU)* aktivitas pengguna telepon seluler di Indonesia meningkat setiap tahun (ITU, 2016). Dan menurut data dari *Cloudmark* pada tahun 2010, 30% dari SMS yang beredar di Asia adalah SMS spam (Agarwal, Sanmeet, & Sunita, 2015). Penelitian ini bertujuan untuk membangun aplikasi SMS Spam *Filtering* pada android menggunakan algoritma *Naïve Bayes*. *Naïve Bayes* adalah algoritma klasifikasi yang sederhana namun memiliki akurasi yang tinggi untuk kasus yang sederhana seperti SMS. Konsep dasar algoritma *Naïve Bayes* adalah peluang kemunculan. Klasifikasi dilakukan berdasarkan teks dari SMS. Data yang digunakan sebanyak 1140 SMS. Data *training* akan dilatih untuk mendapatkan model. Sebelum melakukan klasifikasi, data perlu dipraproses untuk meningkatkan kualitas model, dengan menghilangkan *noise* pada data terlebih dahulu. Klasifikasi dilakukan dua kali. Klasifikasi pertama mengelompokkan SMS spam dan bukan spam. Klasifikasi kedua SMS spam akan dikelompokkan menjadi spam penipuan dan promo. Adapun hasil evaluasi dan validasi menggunakan metode *10-fold cross validation*. Rata-rata akurasi yang didapatkan pada klasifikasi pertama adalah 94,12% dan klasifikasi kedua adalah 94,56%.

Kata kunci— SMS, Spam, Text Classification, Machine Learning, Naïve Bayes, Android.

ABSTRACT

Abstract— The main factor of the increase in the number of spam is the ability to send messages to many recipients at once (Belém & Duarte-Figueiredo, 2011). Besides being supported by the development of technology, the number of mobile phone users is a major factor in the spread of SMS spam. Based on data from the International Telecommunication United (ITU), the number of activities of cellular phone users in Indonesia increases every year (ITU, 2016). And according to data from Cloudmark 2010, in Asia up to 30% of the SMS were represented by spam (Agarwal, Sanmeet, and Sunita, 2015). This research aims to develop a SMS spam filtering application on Android using Naïve Bayes Algorithm. Naïve Bayes Algorithm is a simple classification method but has high accuracy for simple case like SMS case. The base concept of Naïve Bayes is probabilistic of appearance. Classification based on the text of the SMS. The dataset used is 1140 SMS. Data will be trained to get the model. Before classifying, data needs to be praprocessed in order to improve quality of model, by removing the noises in data. The classification will be applied twice, first classification groups SMS spam and not spam. The second classification, SMS spam will be grouped into spam contains fraud and spam contains ads. The results of the evaluation and validation used 10-fold cross validation. The average accuracy for the first classification is 94.12% and for second classification is 94.56%.

Keyword— SMS, Spam, Text Classification, Machine Learning, Naïve Bayes, Android.