

PERBANDINGAN ALGORITMA *NAÏVE BAYES, RANDOM FOREST, DAN C4.5 (DECISION TREE)* DALAM KLASIFIKASI SMS SPAM

Nama Mahasiswa : Shafira Nadila Paramartha
NIM : 3011710056
Pembimbing : Lailatul Hidayah, S.Kom., M.S.

ABSTRAK

Short Message Service (SMS) atau fitur layanan pesan singkat merupakan salah satu teknik komunikasi pada media perangkat selular yang masih digunakan hingga saat ini. Meningkatnya penggunaan layanan SMS telah banyak dimanfaatkan pihak tertentu yang bertujuan selain untuk media komunikasi sehari-hari, diantaranya pesan yang mengandung konteks tidak bermanfaat, penipuan hadiah, iklan/promosi, dana tunai, pelecehan seksual, modus, dan lainnya. SMS yang masuk terkadang berisi konten ilegal yang membuat ketidaknyamanan oleh pihak penerima pesan. Sehingga diperlukan filtering terhadap SMS yang masuk untuk membantu para pengguna ponsel atau penerima SMS mengetahui apakah SMS yang diterimanya spam atau tidak. Maka dari itu, perlu dilakukannya analisis spam. Algoritma yang digunakan dalam penelitian ini adalah *Naïve Bayes, Random Forest, dan C4.5 (Decision Tree)*. Penelitian ini menghasilkan hasil kinerja tertinggi ada pada algoritma Random Forest dengan nilai akurasi sebesar 99%, nilai rata-rata presisi sebesar 99%, nilai rata-rata recall sebesar 99% dan nilai rata-rata F1 sebesar 99,5%.

Kata kunci: SMS, spam, naïve bayes, random forest, decision tree

COMPARISON BETWEEN NAÏVE BAYES, RANDOM FOREST, AND C4.5 (DECISION TREE) ALGORITHM FOR SPAM OF SMS CLASSIFICATION

Student Name : Shafira Nadila Paramartha
Identity Number : 3011710056
Supervisor : Lailatul Hidayah, S.Kom., M.S.

ABSTRACT

Short Message Service (SMS) or the short message service is one of the communication feature on mobile devices that are exist still now. The increased use of SMS services has been widely used by certain parties for purposes other than daily communication media, including messages that contain useless contexts, fraudulent gifts, advertisements / promotions, cash funds, sexual harassment, modes, and many others. The incoming SMS sometimes contains illegal content which causes inconvenience to the recipient of the message. So it requires filtering of incoming SMS to help cellphone users or SMS recipients know whether the SMS they receive is correct or not. One of the filtering implementation techniques uses the accuracy of the data mining classification method for analysis and gets the accuracy value from the comparison between the naïve bayes, random forest, and C4.5 (decision tree) algorithms. This research have the highest performance results are in the Random Forest algorithm with an accuracy value of 99%, an average precision value of 99%, an average recall value of 99% and an average F1 value of 99.5%.

Keywords: SMS, spam, naïve bayes, random forest, decision tree