BAB I
PENDAHULUAN

1.1 Latar Belakang

SMS (*Short Message Service*) merupakan salah satu media komunikasi tanpa kabel yang masih digunakan masyarakat dengan biaya murah, yang memungkinkan pengirim SMS dan penerima SMS dapat dilakukan dengan cepat dan mudah. Pesatnya perkembangan teknologi SMS dipengaruhi oleh banyaknya *provider* penyedia jasa telekomunikasi yang menawarkan jasanya dengan harga yang cukup terjangkau oleh masyarakat luas. Semakin luas pengguna SMS pada masyarakat disalahgunakan oleh pihak yang tidak bertanggung jawab untuk melakukan tindak kejahatan dengan menyebarkan SMS *spam* yang tidak diminta dan tidak diinginkan, seperti promosi, penipuan, pesan porno, dan lain sebagainya (Mahardika, 2018).

Berdasarkan data dari ITU [1], pada tahun 2013 jumlah nomor selular yang dipakai di Indonesia sudah melebihi jumlah penduduk Indonesia atau sekitar 313 juta. Perangkat selular yang dipergunakan saat ini semuanya memiliki fasilitas untuk mengirimkan pesan singkat atau SMS. Sekalipun sudah ada sarana lain yang dapat dipergunakan untuk berkirim pesan selain menggunakan sms seperti *email*, *instant messaging*, dan sebagainya tetapi dibutuhkan perangkat genggam khusus untuk memanfaatkannya, sehingga SMS masih tetap bertahan sampai dengan sekarang (Indiarto, 2016).

Dengan banyaknya jumlah pengguna perangkat genggam yang memiliki fasilitas SMS tersebut maka tidak mengherankan jika SMS masih menjadi salah satu media yang dipergunakan untuk berkomunikasi antar pengguna. Namun, pesan yang berisikan hal-hal yang tidak diinginkan oleh pengguna karena SMS spam terkadang merugikan dan mengarah pada hal penipuan. Dikarenakan hal tersebut para pengguna ponsel diharuskan sangat berhati-hati terhadap SMS yang diterimanya. Sehingga diperlukan filtering terhadap SMS yang masuk untuk membantu para pengguna ponsel atau penerima SMS mengetahui apakah SMS yang diterimanya benar atau tidak. Filtering spam adalah teknik klasifikasi teks

untuk mengatasi spam. Salah satu teknik implementasi filtering dengan menggunakan akurasi dari metode klasifikasi data mining untuk analisis dan mendapatkan nilai akurasi dari komparasi algoritma *Naïve Bayes, Random Forest,* dan *C4.5 (Decision Tree)* dalam klasifikasi SMS.

1.2 Perumusan Masalah

Berdasarkan uraian masalah diatas, maka dapat dirumuskan permasalahan sebagai berikut:

a. Bagaimana tahap pengolahan data dalam penelitian?
b. Bagaimana cara membandingkan algoritma-algoritma yang dipakai?
c. Algoritma mana yang menghasilkan akurasi tertinggi?

1.3 Batasan Masalah

Agar penelitian yang dilakukan tidak menyimpang dari pokok pembahasan dan mencapai sasaran, maka diperlukan batasan sebagai berikut:

a. Data diambil dalam rentan waktu antara bulan Januari 2020 – Agustus 2021
b. Data yang digunakan dalam penelitian ini adalah data dari beberapa penerima teks SMS
c. Data teks SMS diklasifikasikan ke dalam dua kelas yaitu negatif dan positif
d. Data teks SMS yang di analisis berbahasa Indonesia

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka ditentukan tujuan penelitian:

a. Dapat mengetahui akurasi Algoritma Naïve Bayes, Random Forest, dan C4.5 *(Decision Tree)* dalam mengklasifikasikan SMS spam
b. Dapat menjadi referensi untuk masyarakat sebelum merespon SMS spam

1.5 Manfaat Penelitian

Adapun beberapa manfaat yang diharapkan dari penelitian ini adalah:

a. Mengetahui akurasi dari masing-masing algoritma
b. Mengetahui algoritma mana yang cocok untuk permasalahan ini