

## BAB 1. PENDAHULUAN

### 1.1 Latar Belakang

*Short Message Service* atau SMS adalah salah satu media komunikasi berupa pesan singkat yang sering digunakan masyarakat saat ini. Meskipun sekarang telah banyak sosial media sebagai sarana komunikasi yang fungsinya sama dengan SMS seperti, *Whatsapp Messenger*, *Blackberry Messenger*, *Line Messenger*, *Wechat Messenger* dan banyak lagi lainnya. Namun bukan berarti masyarakat meninggalkan media komunikasi SMS yang telah ada lebih dahulu. SMS masih menjadi suatu media komunikasi yang digemari masyarakat, selain dari faktor penggunaan pulsa yang tidak begitu banyak, untuk dapat menikmati media SMS tidak perlu menggunakan *handphone* canggih dengan harga yang mahal. Karena media SMS terdapat pada semua jenis *handphone*.

Dengan banyaknya gaya penulisan dan gaya bahasa yang digunakan dalam mengirim pesan sehingga kalimat SMS yang digunakan menjadi kalimat tidak baku. Misalnya pada kasus mahasiswa terhadap dosen. SMS merupakan media yang dianggap sopan untuk berkomunikasi dengan orang yang dihormati namun bukan berarti mahasiswa dapat mengirimkan SMS dengan bahasa dan gaya tulisan yang berlebihan. Hal tersebut akan berubah menjadi komunikasi yang tidak sopan, karena tidak mengandung etika berkomunikasi terhadap orang yang dihormati.

Maka dengan melihat kasus tersebut, perlu dibuat sebuah sistem untuk mendeteksi kalimat tidak baku. Untuk membangun sistem pendeteksi kemiripan pada dokumen teks diperlukan suatu ilmu yang mengacu pada pengolahan data teks. *Text mining* adalah proses menganalisis teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu (Ian H. Witten 2001:1). Tahap-tahap pada text mining secara umum terdiri dari *case folding*, *tokenizing*, *filtering*, *stemming*, *tagging* dan *analyzing*.

Penelitian ini akan menguji algoritma *cosine similarity* untuk mendeteksi kalimat SMS berdasarkan kata-kata yang digunakan baik dari gaya tulisan atau gaya bahasa. *Cosine similarity* adalah salah satu metode dalam menentukan nilai kemiripan antar dua objek yang dapat diterapkan dalam menentukan nilai kemiripan pada dua berkas dokumen teks. Parameter yang digunakan adalah jumlah kata-kata pada setiap dokumen teks. Komposisi kata pada dokumen teks merupakan penentuan nilai yang dihasilkan dari metode *cosine similarity*.

Dengan adanya deteksi kalimat tidak baku berdasarkan Kamus Besar Bahasa Indonesia dengan metode *cosine similarity* diharapkan dapat membantu untuk mendeteksi kebakuan pada teks SMS. Untuk itu dilakukan sebuah penelitian tentang “DETEKSI KALIMAT TIDAK BAKU BERDASARKAN KAMUS BESAR BAHASA INDONESIA DENGAN METODE *COSINE SIMILARITY*”.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang diatas, dapat dirumuskan masalah sebagai berikut :

1. Bagaimana membangun suatu sistem deteksi kalimat tidak baku berdasarkan Kamus Besar Bahasa Indonesia untuk mendeteksi kalimat pada teks SMS berdasarkan kata yang digunakan dengan metode *cosine similarity* ?
2. Bagaimana menerapkan metode *cosine similarity* pada sistem deteksi kalimat tidak baku berdasarkan Kamus Besar Bahasa Indonesia ?
3. Bagaimana menganalisa tingkat akurasi metode *cosine similarity* ?

## **1.3 Batasan Masalah**

Batasan masalah yang diambil adalah sebagai berikut :

1. Data yang digunakan adalah 100 teks *Short Message Service*.
2. Penelitian ini mendeteksi kalimat tidak baku berdasarkan kata yang digunakan dengan pedoman Kamus Besar Bahasa Indonesia.

3. Kata pada kalimat teks *Short Message Service* tidak berupa kata yang bersifat teknis.
4. Dokumen Teks yang dibandingkan adalah dokumen teks yang seluruhnya atau sebagian besar menggunakan bahasa Indonesia.
5. Penerapan sistem hanya diterapkan pada dekstop.
6. *Wordlist* yang digunakan telah ditentukan oleh peneliti.
7. Dokumen SMS yang memiliki karakter maksimal 160 karakter.
8. Proses *text mining* yang dilakukan adalah proses *tokenizing* dan *stemming*.
9. Data *testing* dapat dilakukan dengan mengirimkan teks SMS langsung.

#### **1.4 Tujuan**

Tujuan penelitian ini adalah sebagai berikut :

1. Membangun suatu sistem deteksi kalimat tidak baku berdasarkan Kamus Besar Bahasa Indonesia untuk mendeteksi kalimat pada teks SMS berdasarkan kata yang digunakan dengan metode *cosine similarity*.
2. Menganalisa hasil perhitungan *cosine similarity* pada studi kasus teks SMS dan mengukur tingkat akurasi.

#### **1.5 Manfaat Penelitian**

Manfaat dalam penelitian ini adalah sebagai berikut :

1. Dengan sistem deteksi kalimat tidak baku yang akan dibangun dengan pedoman Kamus Besar Bahasa Indonesia maka akan lebih mudah mendeteksi tingkat kebakuan kalimat pada teks SMS.
2. Dengan menerapkan metode *cosine similarity*, maka akan dapat mengetahui tingkat akurasi metode tersebut pada suatu sistem untuk mendeteksi SMS.