

1 BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa merupakan alat komunikasi untuk mengungkapkan gagasan, pemikiran, maksud serta tujuan kepada orang lain [1]. Selain sebagai alat komunikasi tatap muka, bahasa juga digunakan dalam metode komunikasi lain seperti *online chatting* media sosial melalui Internet. Bertambahnya jumlah pengguna Internet turut meningkatkan jumlah pengguna media sosial, dimana pengguna memiliki kebebasan berinteraksi tanpa melihat usia, status maupun *gender* sehingga tidak adanya penghalang yang memisahkan [2]. Pengguna bebas menggunakan kosakata apapun yang biasanya lebih informal [3]. Namun terkadang penggunaan kata atau kalimat yang dianggap sebagai etika dapat menimbulkan kesalahpahaman [4]. Komunikasi dalam media sosial seringkali tidak mempergunakan bahasa yang baku, hal demikian dapat menyebabkan pengguna acuh pada etika berkomunikasi seperti berkata kasar [4].

Penggunaan bahasa kasar berhubungan dengan makian, dimana bahasa kasar sering digunakan dalam menuturkan hal tersebut [5]. Kata kasar biasanya dipergunakan untuk menyerang pihak tertentu, menyalurkan rasa kesal, atau meluapkan emosi terhadap suatu hal [6]. Bentuk makian berupa kata kasar bagi pendengar akan berdampak pada hilangnya kepercayaan diri, hingga memunculkan perasaan dirinya selalu menjadi korban dan merasa tidak mau berinteraksi dengan orang lain [5]. Dampak negatif juga dirasakan bagi penutur, dimana label abnormal akan diberikan oleh pendengar dan penutur akan merasa dirinya terisolasi dari dunia luar [5]. Dalam *online chatting*, penggunaan kata kasar seperti kata hewan tidak selamanya terlihat kasar, oleh karena itu untuk mengidentifikasi kasar atau tidak suatu kata perlu dilihat dari konteks kalimatnya [6].

Penelitian terkait pendeteksian konteks kalimat telah dilakukan pada beberapa studi dengan menggunakan pemodelan algoritme pada *machine learning*. Studi yang melakukan penelitian terkait pendeteksian berdasarkan konteks kalimat meliputi penelitian dengan model *Naïve Bayes* [7], SVM (*Support Vector Machine*) [8], LSTM (*Long Short Term Memory*) [9]. Umumnya model yang digunakan seperti *Naïve Bayes* dan SVM memiliki performa yang kurang baik dibandingkan model *deep learning* seperti LSTM dan Bi LSTM, karena ketika

data yang banyak diberikan akan cenderung membuat *machine learning* stagnan [10]. Berdasarkan penelitian yang dilakukan oleh [11] bahwa *deep learning* memiliki performa 10-14% lebih baik dibandingkan dengan *machine learning* tradisional dalam melakukan klasifikasi karena memiliki akurasi yang tinggi dan presisi. Untuk LSTM, walaupun merupakan model *deep learning*, pembelajaran yang dilakukan hanya searah sehingga hasil yang didapat kurang akurat. Berbeda dengan Bi LSTM (*Bi directional Long Short Term Memory*) yang merupakan turunan dari LSTM. Bi LSTM dapat melakukan pemrosesan data *sequence* dalam dua arah dengan dua lapisan terpisah yang kemudian diteruskan ke lapisan *output* [12]. Bi LSTM ini dapat mempelajari konteks kalimat lebih baik dibanding LSTM searah dan memiliki performa yang lebih unggul dalam penguraian teks dan prediksi kata dari kalimat *input* [13]. Penelitian yang dilakukan oleh Siami Namini [13] menunjukkan bahwa model Bi LSTM memiliki tingkat akurasi yang lebih unggul dibanding model LSTM dengan pengurangan tingkat kesalahan sebesar 37,78% dimana pada penelitian tersebut tingkat *error RMSE (Root Mean Square Error)* dari 39.09 (LSTM) menjadi 20.17 (Bi LSTM).

Dari penjabaran di atas, maka dikembangkanlah aplikasi *chatting* berbasis *mobile* dan *web* untuk berkomunikasi secara *online* yang menerapkan fitur untuk mencegah pengiriman pesan ketika ditemukan adanya konteks pesan yang dianggap kasar dalam berkomunikasi *online* secara *real-time* atau menyensor kata kasar dari pesan yang dikirimkan, dengan implementasi model Bi LSTM serta menuangkannya ke dalam tugas akhir dengan judul **“PENGEMBANGAN APLIKASI CHATTING PENDETEKSI BAHASA KASAR SECARA REAL-TIME DENGAN IMPLEMENTASI MODEL BI LSTM BERBASIS MOBILE DAN WEB”**.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah pada tugas akhir ini adalah diperlukannya pendeteksian bahasa kasar berdasarkan konteks kalimat pada saat berkomunikasi secara *online*.

1.3 Tujuan

Tujuan tugas akhir ini adalah membuat sebuah aplikasi *chatting* berbasis *mobile* dan *web* dengan mengimplementasikan model Bi LSTM untuk mengidentifikasi pesan bermakna

kasar guna mengurangi penggunaan kata-kata kasar (*Toxic*) dalam berkomunikasi secara *online*.

1.4 Manfaat

Manfaat tugas akhir ini adalah sebagai berikut:

1. Pengguna dapat berkomunikasi *chatting* secara *real-time*
2. Pengguna yang tidak suka menggunakan bahasa kasar dapat berkomunikasi dengan aman tanpa takut lawan bicara berkata kasar.

1.5 Batasan Masalah

Batasan Masalah dari tugas akhir ini adalah sebagai berikut:

1. Minimal sistem operasi *smartphone* yang digunakan adalah Android 5.0+ Lollipop untuk aplikasi berbasis *mobile*.
2. Aplikasi hanya dapat mengirimkan kalimat (tidak dapat mengirimkan gambar, video, *link*, *sticker*, dan dokumen).
3. Kalimat bermakna kasar yang tidak dipelajari oleh model dari *dataset* tidak dapat terdeteksi.
4. Jenis kata kasar yang digunakan dalam kalimat pada dataset adalah sebagai berikut:
 - a. Pengkondisian, contohnya seperti to***, bodoh
 - b. Mengandung unsur hewan, contohnya seperti anj***, ba**
 - c. Bagian dari tubuh, contohnya seperti k****1, kepala otak
 - d. Anggota keluarga, contohnya seperti bapakmu
 - e. Profesi, contohnya seperti l**te
 - f. Bahasa asing dan daerah, contohnya seperti *su, *mother****er*

1.6 Metodologi Penelitian

Sebelum mengembangkan sistem, tahap pertama yang dilakukan adalah melakukan kajian terhadap penelitian sebelumnya terkait permasalahan dan melakukan pemahaman secara teoritis dari algoritme *bi directional long short term memory* (Bi LSTM). Penelitian ini menggunakan *dataset* yang telah tersedia pada *repository* jurnal ilmiah penelitian yang berkaitan dengan penggunaan bahasa kasar [14], [15]. *Dataset* tersebut merupakan data yang diambil dari *tweets* pengguna aplikasi Twitter dan komentar-komentar postingan Instagram.

Setelah itu juga dilakukan analisis dari proses kerja aplikasi dimulai dari penginputan pesan oleh pengguna lalu akan dilakukan prediksi oleh model Bi LSTM melalui API untuk mendapatkan hasil apakah pesan berkonteks kasar atau tidak yang kemudian akan dikirim ke aplikasi chat.

Untuk proses pengembangan sistem pada penelitian ini menerapkan model *waterfall*. Model *waterfall* merupakan salah satu model pengembangan sistem yang sistematis [16]. Proses yang dilakukan adalah sebagai berikut:

a. Analisis

Pada tahap ini dilakukan analisis seperti analisis proses, kebutuhan fungsional maupun kebutuhan non-fungsional. Analisis proses menjelaskan tentang proses kerja aplikasi dalam melakukan prediksi kalimat. Untuk analisis kebutuhan fungsional digambarkan menggunakan *use case diagram*, sedangkan kebutuhan non-fungsional menggunakan metode PIECES (*Performance, Information, Economy, Control, Efficiency, Service*).

b. Perancangan

Perancangan dan pemodelan antarmuka aplikasi menggunakan *website* Figma sedangkan alur proses sistem dirancang dalam *diagram flowchart* menggunakan Microsoft Visio.

c. Implementasi

Tahap berikutnya adalah melakukan pembersihan teks (*text cleaning*) dari *dataset* yang dipakai, mengubah semua teks menjadi huruf kecil (*lower case*), melakukan integrasi data, melakukan transformasi data dan membuang *noise* pada teks. Lalu melakukan pengkodean program untuk aplikasi *mobile* dan *web* menggunakan bahasa pemrograman Dart dan menggunakan *framework* Flutter 2.0. Melakukan pengembangan model Bi LSTM untuk pengembangan fitur pencegahan pengiriman dan pengembangan pendeteksian kata kasar pada aplikasi menggunakan bahasa pemrograman Python dan menggunakan *framework* Tensorflow untuk membantu dalam melatih model *machine learning*, mengukur akurasi model, menyimpan bobot pada model, merancang *layer* pada model, melakukan *pre-processing* pada data terhadap model, melakukan kompilasi pada model, merancang sistem *filtering*, untuk melakukan prediksi terhadap data.

d. Pengujian

Pada tahap ini dilakukan pengujian terhadap aplikasi, apakah berjalan sesuai dengan yang diharapkan meliputi:

- i. Pengujian aplikasi *chatting mobile* dan *web* dilakukan menggunakan metode *black box testing* terhadap spesifikasi untuk melihat apakah implementasi sesuai yang diharapkan.
- ii. Pengujian penerapan Bi LSTM pada aplikasi dengan memberikan serangkaian kalimat sebagai *test case* yang didapat dari hasil observasi terhadap komentar pada akun Instagram @lambe_turah sebanyak 50 kalimat yang dianggap kasar berdasarkan penggunaan pengelompokan kata kasar dan 50 kalimat yang tidak dianggap kasar selama 1 minggu dimulai dari tanggal 28 Januari 2022 hingga 4 Februari 2022 [17]. Kalimat yang memiliki makna kasar tidak bisa dikirim dan akan memberikan *feedback* kepada pengguna bahwa kalimat tersebut mengandung makna kasar, sedangkan kalimat yang tidak mengandung makna kasar akan dikirim dan diteruskan kepada lawan bicara. Pengujian tingkat akurasi prediksi dari model Bi LSTM akan menggunakan tabel *Confusion Matrix* sebagai alat ukurnya.

Setelah pengujian selesai dilakukan, maka dilakukan proses menarik kesimpulan apakah sesuai dengan hasil yang diinginkan.

UNIVERSITAS
MIKROSKIL