

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Salah satu indikator keberhasilan seseorang dalam berkomunikasi adalah merangkai suatu kata dengan menerapkan struktur kalimat yang baik dan benar, baik itu komunikasi secara langsung maupun tidak langsung. Dalam komunikasi secara tidak langsung, struktur kalimat sangat diperlukan untuk membantu komunikan dapat dengan mudah mengerti pesan-pesan yang diberikan oleh komunikator. Suatu kalimat dapat terbentuk karena dipengaruhi oleh beberapa faktor, seperti dimensi ruang, waktu, serta suasana komunikator dalam penyampaian pesan kepada komunikan[1].

Data dari Twitter yang memberikan hasil frekuensi penggunaan kalimat baku oleh pengguna Twitter di Indonesia menunjukkan 67.18% menggunakan kalimat formal, 12.23% menggunakan kalimat non-formal, dan 20.57% tidak dikenali dari 1000 *tweet* yang diperoleh[2]. Frekuensi penggunaan kalimat formal yang cukup banyak tidak menjamin masyarakat Indonesia fasih dalam penggunaan kalimat formal, khususnya dalam menentukan kata-kata baku yang akan digunakan. Banyaknya bahasa asing yang ingin dikuasai, ketidakseimbangan penggunaan kalimat baku dan tidak baku, mengabaikan definisi suatu kalimat, serta penguasaan struktur bahasa yang rendah menjadi faktor sulitnya menentukan kalimat yang baku dalam penulisan[3].

Penelitian di bidang pemrosesan bahasa alami untuk mesin terjemahan telah dilakukan sebelumnya. Penelitian karya Lirong Yao dan Yazhuo Guan pada tahun 2018 yang berjudul “*An Improved LSTM Structure for Natural Language Processing*”. Banyak sistem NLP masih memiliki masalah praktis, seperti kompleksitas pelatihan yang tinggi, kesulitan komputasi dalam skenario konten skala besar, kompleksitas pengambilan yang tinggi dan kurangnya signifikansi probabilitasnya[4]. Penelitian ini mengusulkan metode NLP yang ditingkatkan berdasarkan struktur memori jangka pendek (LSTM), yang parameternya dibuang secara acak ketika dilewatkan ke simpul

sebelumnya dalam lapisan proyeksi rekursif[4]. Penelitian karya Shashi Pal Singh, Ajai Kumar, et al pada tahun 2017 yang berjudul “*Machine Translation using Deep Learning*”. Penelitian ini menyajikan bagaimana melatih jaringan syaraf berulang untuk menata ulang sumber dan menargetkan bahasa dengan menggunakan metode pembelajaran *semi-supervised*[5]. Word2vec diperlukan untuk menghasilkan vektor kata dari bahasa sumber dan Auto encoder membantu penelitian tersebut dalam rekonstruksi vektor untuk bahasa target dalam struktur pohon[5].

Melalui penelitian ini, masalah dalam pengubahan kalimat tidak baku dapat dipelajari dengan penggunaan metode *Natural Language Processing* dan algoritma pembelajaran mesin *Long Short Term Memory*. Penulisan kalimat tidak formal diteliti untuk dapat diterjemahkan kedalam kalimat formal dengan memperhatikan urutan kata pada suatu kalimat. Kalimat yang telah diinput akan diurai menjadi beberapa kata dengan menggunakan metode Tokenisasi. Data hasil pra-pemrosesan berupa vector dari proses *pad_sequences* menjadi suatu masukan untuk algoritma pembelajaran mesin *Long Short Term Memory*. Proses terjemahan mesin juga dilakukan dengan proses encoder-decoder. Encoder digunakan untuk menyandikan suatu susunan kalimat menjadi masukan dalam proses prediksi. Sedangkan decoder digunakan untuk mengembalikan sandi tersebut kedalam kalimat prediksi. Untuk menguji nilai akurasi dari prediksi kalimat formal, peneliti menggunakan evaluasi BLEU untuk mendapatkan nilai evaluasi dari hasil prediksi kalimat. Evaluasi tersebut membandingkan n-grams hasil prediksi dengan n-gram translasi referensi dari dataset dan menghitung jumlah kesamaan.

Hasil dari penelitian ini menguji akurasi pembelajaran mesin dalam menentukan kata-kata baku menjadi kalimat formal serta menentukan faktor-faktor dari nilai akurasi yang dihasilkan dalam penelitian. Evaluasi BLEU menentukan diterima atau tidaknya kalimat hasil prediksi dengan mengukur kedekatan konteks kalimat formal dari dataset. Selain itu, dari penelitian ini akan dihasilkan aplikasi berbasis website untuk penyebaran hasil model prediksi.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya. Maka rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana menerapkan algoritma *long short term memory* dalam pemrosesan bahasa alami untuk alih bentuk kalimat non-formal menjadi kalimat yang formal dengan pendekatan *machine translation*?
2. Apa saja faktor-faktor yang menjadi penunjang serta penghambat dari nilai akurasi yang dihasilkan dalam proses pembelajaran mesin?
3. Bagaimana membuat aplikasi berbasis website untuk memprediksi kalimat non-formal menjadi kalimat formal menggunakan algoritma *long short term memory*?

1.3. Batasan Masalah

Peneliti akan membatasi masalah penelitian sebagai berikut:

1. Dataset dikumpulkan dari portal berita Kompas.id dengan topik berita bisnis dan ekonomi. Keterbatasan waktu penelitian membatasi banyaknya terjemahan kata ke dalam kalimat non-formal.
2. Dataset berupa kata serta kalimat dalam bentuk formal dan non-formal yang berjumlah 3000 untuk *training* dan 100 untuk *testing*.
3. Pengembangan model dilakukan dengan membandingkan hasil arsitektur model dari proses pembuatan Matriks Embedding dengan panjang matriks 100, 200, dan 300.
4. Proses pembelajaran mesin menggunakan masukkan kalimat per kalimat.
5. Penggunaan kata tidak baku tidak meliputi singkatan dan kata *slang*.
6. Menggunakan *framework* komputasional dengan Keras dan Tensorflow dalam pelatihan model pelatihan.
7. Menggunakan *framework* Flask dalam pengembangan aplikasi penerjemah kalimat non-formal menjadi kalimat formal.

1.4. Tujuan Penelitian

Tujuan dari penelitian adalah sebagai berikut:

1. Mengembangkan model dengan menerapkan algoritma *Long Short Term Memory* dalam pemrosesan bahasa alami untuk alih bentuk kalimat tidak formal menjadi kalimat formal berbasis website.
2. Menentukan faktor penunjang dan penghambat dari nilai akurasi yang dihasilkan dalam proses pembelajaran mesin.

1.5. Manfaat Penelitian

Manfaat dari penelitian sebagai berikut:

1.5.1. Manfaat Akademis

Manfaat akademik dari penelitian ini yaitu untuk mengetahui bahwa penerapan algoritma *long short term memory* dalam pemrosesan bahasa alami untuk alih bentuk kalimat non-formal menjadi kalimat yang formal dengan pendekatan *machine translation* dapat dilakukan. Selain itu, faktor penunjang dan penghambat dari nilai akurasi yang dihasilkan dalam proses pembelajaran mesin dapat dijadikan acuan untuk pengembangan pemrograman bahasa alami dengan Bahasa Indonesia. Khususnya dalam topik kalimat formal dan non-formal.

1.5.2. Manfaat Praktis

Manfaat praktis dari penelitian ini yaitu dapat digunakan untuk penerjemah kalimat tidak formal menjadi suatu susunan kalimat.

1.6. Struktur Penulisan

Untuk memahami lebih jelas laporan ini, maka materi-materi yang tertera pada Laporan Skripsi ini dikelompokkan menjadi beberapa subbab dengan sistematika penyampaian sebagai berikut:

BAB I PENDAHULUAN

Berisi tentang latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian yang mencakup manfaat akademik serta manfaat praktis, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisikan teori yang berupa pengertian, definisi, dan cara pandang yang diambil dari kutipan buku yang berkaitan dengan penyusunan laporan skripsi serta beberapa tinjauan literatur yang berhubungan dengan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini mencakup pembahasan jenis penelitian, metode penelitian, instrumen penelitian, objek penelitian, serta kerangka pemikiran dalam melakukan proses perancangan sistem penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan uraian hasil akurasi dalam menentukan kata-kata baku menjadi susunan kalimat formal yang memperhatikan struktur kalimat itu sendiri. Pembahasan metode penelitian secara lengkap dalam mendapatkan hasil penelitian.

BAB V SIMPULAN DAN SARAN

Bab ini berisikan kesimpulan yang merupakan intisari hasil penelitian yang dilihat keterkaitannya dengan perumusan masalah dan tujuan penelitian. Saran yang berisikan rekomendasi untuk penelitian lanjutan diperlukan karena adanya keterbatasan penelitian.

DAFTAR PUSTAKA

Berisikan sumber-sumber pustaka yang digunakan sebagai acuan dalam menyelesaikan penelitian baik dari buku, jurnal ilmiah, media massa, internet, atau data sekunder lainnya.



KALBIS Institute

Transforming • Hearts and Minds