

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam komunikasi sehari-hari, ada dua macam model bahasa yang banyak digunakan masyarakat di Indonesia, yaitu bahasa formal dan bahasa tidak formal. Bahasa formal merupakan bahasa yang mengikuti kaidah bahasa, sedangkan bahasa yang tidak formal adalah bahasa yang keluar dari kaidah baku bahasa. Namun demikian, masyarakat lebih sering menggunakan bahasa nonformal untuk komunikasi sehari-hari. Selain dalam percakapan secara langsung, penggunaan bahasa nonformal banyak pula digunakan di media sosial, *messenger*, dan sebagainya. Penggunaan bahasa tersebut digunakan di antaranya dengan alasan untuk memperlancar komunikasi dan menyederhanakan tulisan pesan.

Adapun contoh bentuk bahasa nonformal di antaranya kata *slang* dan kata yang disingkat. Kata *slang* yang biasa digunakan misalnya “ayok”, “asik”, “aqu”, “lobang”, “nagih” dan lain lain. Sedangkan kata yang disingkat misalnya “mngrti”, “yg”, “sdng”, “mmpu”, dan “bkn”. Selain itu sering juga terjadi kesalahan penulisan seperti posisi huruf yang tertukar, kurang huruf, kelebihan huruf, dan tertukarnya huruf. Adapun contoh untuk masing-masing kesalahan penulisan tersebut dapat dilihat pada Tabel 1.1.

Tabel 1.1 Jenis dan contoh kesalahan penulisan

No.	Jenis kesalahan penulisan	Contoh (kata sedang)	Penjelasan
1	Posisi huruf tertukar	Sedagn	Posisi Huruf ‘g’ dan ‘n’ tertukar
2	Kurang huruf	Sedng	Kurang huruf ‘a’
3	Kelebihan huruf	Sedanng	Kelebihan huruf ‘n’
4	Tertukarnya huruf	Sedqng	Kesalahan penulisan huruf ‘a’ dengan ‘q’

Pengolahan bahasa alami adalah salah satu bidang ilmu yang menangani bentuk bahasa tersebut untuk berinteraksi dengan komputer. Dalam pengolahan bahasa alami, kemunculan kata tidak standar seperti di atas sangat mempengaruhi dalam proses pengolahan kata. Oleh karena itu, ada sebuah tahapan yang disebut dengan *preprocessing*. *Preprocessing* bertujuan untuk menghilangkan data dari *noise*. Salah satu tahap yang sering dilakukan dalam

preprocessing adalah normalisasi kata. Normalisasi merupakan salah satu tahap yang mengubah menjadi kata yang sesuai dengan standar kaidah bahasa Indonesia.

Berdasarkan permasalahan yang dipaparkan di atas, maka diperlukan adanya proses secara otomatis untuk menangani perubahan kata yang tidak standar menjadi standar. Oleh karena itu, maka pada penelitian ini diusulkan sebuah solusi untuk menanggapi permasalahan tersebut dengan algoritma *Levenshtein Distance*.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini yaitu

- a. Bagaimana mengimplementasikan algoritma *Levenshtein Distance* dalam memperbaiki kalimat yang mengandung *Out of Vocabulary* menjadi kata bahasa Indonesia yang baik dan benar?
- b. Bagaimana akurasi penggunaan algoritma *Levenshtein Distance* untuk melakukan normalisasi *Out of Vocabulary*?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini di antaranya:

- a. Pemeriksaan kata pada kalimat hanya terbatas di bahasa Indonesia saja.
- b. Seluruh kata dalam bahasa Inggris akan dianggap sebagai istilah dan diabaikan.
- c. Data uji yang digunakan adalah data Twitter dengan pencarian acak pada tanggal 24 April 2019 di *timeline* akun @Lambe_Turah.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah membuat sistem *preprocessing* teks yang dapat mengenali *Out of Vocabulary* dan memperbaikinya dengan kata bahasa Indonesia yang tepat.

1.5 Manfaat Penelitian

- a. Bagi peneliti

Menambah wawasan, pengetahuan serta kemampuan teknis peneliti dalam pembuatan tahap *preprocessing* di lingkup pemrosesan bahasa alami. Penelitian ini juga menambah wawasan, pengetahuan serta kemampuan teknis peneliti dalam melakukan normalisasi terhadap kata yang tidak standar dengan menggunakan algoritma *Levenshtein Distance*.

b. Bagi pengguna

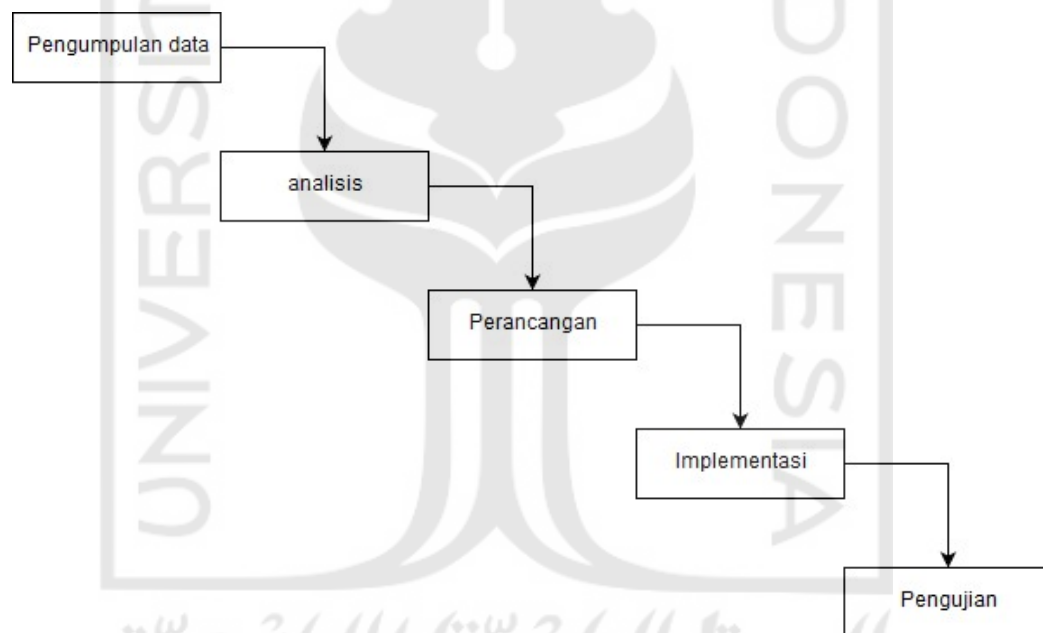
1. Membantu tahap *preprocessing* saat *text mining*
2. Mempermudah mendeteksi kata yang tidak sesuai dengan bahasa Indonesia.

c. Bagi Institusi Pendidikan

Mendapatkan informasi dalam pengembangan ilmu pengetahuan yang berhubungan dengan normalisasi teks, sehingga dapat digunakan sebagai referensi dalam penelitian selanjutnya.

1.6 Metode Penelitian

Langkah-langkah yang digunakan untuk mengembangkan dapat dilihat pada Gambar 1.1.



Gambar 1.1 Tahap metodologi penelitian

a. Pengumpulan Data

Tahap ini merupakan tahap awal dari penelitian. Pada tahap ini dilakukan pengumpulan data dengan menggunakan metode studi literatur. Studi literatur dilakukan melalui penelusuran literatur dari buku, dan penelitian-penelitian sebelumnya baik skripsi, tesis, jurnal dan artikel yang berhubungan dengan normalisasi teks bahasa Indonesia.

b. Analisis

Analisis kebutuhan sistem merupakan tahapan untuk mendefinisikan kebutuhan dari sistem yang dibangun. Analisis dilakukan pada algoritma *Levenshtein Distance* dengan cara memahami cara kerja algoritma tersebut.

c. Perancangan sistem

Tahap ini dipaparkan deskripsi kebutuhan dan perancangan bentuk sistem yang akan dibuat berdasarkan analisis kebutuhan sistem. Perancangan sistem menggunakan gambaran alur utama sistem, *Flowchart*, dan perancangan antarmuka (*interface*) untuk mempermudah dalam pembuatan sistem sesuai dengan kebutuhan, serta perancangan pengujian untuk mengetahui performa dari algoritma yang digunakan.

d. Implementasi

Tahap sistem akan dibuat berdasarkan analisis kebutuhan sistem dan perancangan yang sudah dilakukan dengan menuliskan kode program menggunakan bahasa pemrograman Python beserta contoh implementasi kode program tersebut dalam sistem dan implementasi *framework Django* sebagai *interface*.

e. Pengujian

Pengujian sistem dilakukan setelah proses implementasi selesai dan kemudian dilanjutkan dengan proses evaluasi dengan cara menjalankan aplikasi dan mengisi masukan dan menilai hasil keluaran. Tahap ini dilakukan agar dapat mengetahui apakah aplikasi tersebut sudah sesuai dengan tujuan atau masih terdapat kekurangan.

1.7 Sistematika Penulisan

Dalam penulisan skripsi ini, secara keseluruhan terdiri dari lima bab yang masing-masing bab disusun dengan sistematika sebagai berikut:

BAB I PENDAHULUAN

Bab ini merupakan bagian awal, Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penelitian laporan.

BAB II LANDASAN TEORI

Bab ini berisi tentang teori-teori yang berhubungan dengan permasalahan yang diangkat dari penelitian ini. Teori-teori tersebut antara lain mengenai penjelasan *Natural Language Processing*, *Text Mining*, *Preprocessing*, *Twitter*, *framework Django*, *spelling correction*, dan metode yang digunakan yaitu *Levenshtein Distance*.

BAB III METODOLOGI PENELITIAN

Berisi analisis dan sistem yang akan dibangun, uraian pada bab ini mencakup tahapan penelitian, analisis kebutuhan, tahap pengumpulan data, *preprocessing* yang dipakai, perancangan *flowchart*, perancangan antarmuka (*interface*), dan rancangan pengujian sistem.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Berisi pembahasan tentang hasil dari setiap proses, hasil pengujian, dan antarmuka yang ada dalam sistem sesuai dengan penyelesaian masalah yang diangkat, serta pengujian validasi dan implementasi sistem beserta kelebihan dan kekurangan sistem.

BAB V KESIMPULAN DAN SARAN

Bab terakhir berisi kesimpulan yang didasarkan pada hasil yang telah dicapai dari pembahasan, selain itu juga berisi saran yang diharapkan dapat digunakan sebagai bahan pertimbangan bagi penelitian selanjutnya yang akan mengembangkan penelitian serupa.

