

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kata *slang* merupakan istilah nonformal kerap digunakan oleh kalangan generasi muda, untuk menyampaikan ungkapan dan gagasan baik itu secara langsung maupun tidak langsung. *Slang* sering kali mencerminkan identitas generasi atau kelompok kultural tertentu. Berbagai kelompok usia atau komunitas dapat mengembangkan kata-kata atau frasa khusus yang hanya dimengerti oleh mereka. Ini dapat menimbulkan ketidakpahaman atas perbedaan pemahaman antar generasi (Samudro, 2019).

Menurut Samudro (2019) penggunaan kata *slang* sering kali digunakan sebagai sarana komunikasi antar pengguna internet di media sosial terutama *Twitter*. Namun kata *slang* ini menjadi sebuah permasalahan oleh kalangan tertentu. Salah satu *issue* yang muncul adalah bahwa kata *slang* dapat menjadi sulit dipahami jika diucapkan di luar konteks yang tepat. Ini dapat menciptakan situasi di mana komunikasi menjadi kurang efektif, terutama ketika orang yang tidak familiar dengan kata *slang*. Berdasarkan kesenjangan yang terjadi ada beberapa pendekatan yang digunakan pada penelitian ini, salah satunya pendekatan normalisasi.

Pendekatan normalisasi kata saat ini sangat diperlukan, guna mengartikan setiap kata ke dalam bahasa formal untuk memberikan pemahaman terhadap setiap kalangan masyarakat. Meskipun demikian, fenomena penggunaan kata-kata non baku ini dapat menjadi keuntungan bagi akademisi, banyak diantaranya yang memanfaatkannya sebagai bahan penelitian di bidang pemrosesan data teks. Data yang dihasilkan dari media sosial

kemudian dapat diproses menggunakan metode yang dikenal sebagai *Natural Language Processing* (NLP). *Natural Language Processing* (NLP) merupakan suatu pengembangan teknik komputasi bahasa alami dalam menganalisis dan merepresentasikan teks ataupun lisan untuk mencapai pemrosesan bahasa seperti bahasa manusia (Ramadhanti et al., 2019)

Terdapat banyak teknik maupun model yang dapat digunakan dalam NLP. Penelitian ini berfokus pada beberapa teknik ekstraksi fitur seperti *FastText* dan *Word2Vec*, yang mana dilakukan perbandingan pada setiap teknik untuk memetakan kata-kata ke dalam vektor numerik. Adapun perlu dilakukan tahapan-tahapan seperti pra-pemrosesan data, sebelum memulai proses pemrosesan data. Namun, dengan banyaknya variasi kata dalam bentuk tidak baku untuk suatu konteks yang sama, menjadi sulit dilakukan. Contohnya, kata 'semangat pagi' memiliki banyak variasi non baku seperti 'met pagi', 'mangats pagi', 'cemungut pagi' dan sebagainya. Tanpa melakukan pra-pemrosesan data, setiap kata tersebut dapat direpresentasikan sebagai kata yang berbeda.

Sejumlah penelitian sebelumnya telah dilakukan terkait klasifikasi data teks. Salah satunya penelitian (Nurdin et al., 2020) yang membandingkan kinerja *word embedding*, seperti *Word2Vec*, *GloVe*, dan *FastText*, yang diklasifikasikan menggunakan algoritma *Convolutional Neural Network*. Hasil penelitian tersebut menunjukkan nilai *F-measure* berturut-turut sebesar 0.925, 0.958, dan 0.979 (Nurdin et al., 2020). Penelitian lainnya membahas analisis sentimen dari *review* hotel dengan membandingkan akurasi model menggunakan *Word2Vec* dan *FastText*. Dalam penelitian ini, kedua metode *embedding* tersebut digabungkan dengan teknik *ensemble learning*, yakni *Random Forest*, *Extra Tree*,

dan *AdaBoost*. Hasil terbaik diperoleh saat *FastText* digabungkan dengan *Random Forest* dan *Extra Tree*, masing-masing mencapai akurasi sebesar 93% (Khomsah et al., 2021).

Berdasarkan beberapa penelitian banyak literatur tentang analisis teks menggunakan *word embedding* atau pengembangannya. Dari berbagai penelitian, sangat populernya penggunaan *word embedding* digunakan dalam bidang NLP, maka penelitian ini bertujuan untuk melakukan perbandingan kinerja dari model *word embedding* yaitu *FastText* dan *Word2Vec*. Penelitian ini akan membahas “Perbandingan Normalisasi Kata *Slang* Bahasa Indonesia Menggunakan Model *FastText* dan *Word2Vec* Dengan Pendekatan *Natural Language Processing*”.

1.2. Rumusan Masalah

Berdasarkan latar belakang uraian masalah di atas, peneliti dapat merumuskan masalah pada penelitian ini yaitu bagaimana hasil perbandingan Model *FastText* dan *Word2Vec* dalam normalisasi dan memahami kata-kata *slang* Bahasa Indonesia.

1.3. Batasan Masalah

Berikut ini merupakan beberapa batasan masalah terkait dengan penelitian yang akan dilakukan:

1. Penelitian ini membandingkan dua model teknik NLP yaitu *Word2Vec* dan *FastText*.
2. Penelitian ini menggunakan data sumber dari media sosial yang terdiri dari teks dalam Bahasa Indonesia.
3. Penelitian ini nantinya melakukan teknik *scraping* dalam mengumpulkan data teks atau komentar dari media sosial.
4. Luaran dari penelitian ini merupakan dataset dan hanya berfokus pada normalisasi kata *slang* dan mengkonversi menjadi bentuk baku atau formal.

5. Bentuk implementasi pada penelitian ini hanya berupa model.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah untuk mengetahui hasil dari perbandingan Model *Fasttext* dan *Word2Vec* dalam normalisasi dan memahami kata-kata *slang* Bahasa Indonesia.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat bermanfaat sebagai berikut:

1. Penelitian ini dapat membantu dalam memahami lebih baik bahasa *slang* yang digunakan dalam komunikasi sehari-hari di masyarakat Indonesia.
2. Terkait dengan hasil dari perbandingan model NLP, penelitian ini dapat membantu mempelajari karakteristik maupun kinerja dari masing-masing model.
3. Penelitian ini berkontribusi dalam mengumpulkan data set khususnya pada Bahasa Indonesia yang telah dilakukan proses NLP hingga menjadi dataset yang siap digunakan untuk kepentingan pengembangan penelitian.

1.6. Sistematika Penulisan

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan untuk kasus yang akan dipecah.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan tentang teori-teori yang akan digunakan untuk penelitian Perbandingan Normalisasi Kata *Slang* Bahasa Indonesia Menggunakan Model *FastText* & *Word2Vec* Dengan Pendekatan *Natural Language Processing*.

BAB III METODE PENELITIAN

Bab ini menjelaskan langkah-langkah metode penelitian untuk Perbandingan Normalisasi Kata *Slang* Bahasa Indonesia Menggunakan Model *FastText* & *Word2Vec* Dengan Pendekatan *Natural Language Processing*.BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil penelitian yang telah dilakukan yang terdiri dari analisis data, pengujian metode, dan evaluasi metode yang menggunakan *Cosine Similarity*.

BAB V PENUTUP

Bab ini memuat kesimpulan dari hasil penelitian yang dilakukan dan saran untuk penelitian selanjutnya yang dapat dilakukan pengembangan mengenai topik terkait.