

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen skripsi mengenai analisis sentimen ulasan aplikasi SatuSehat di *Google Play Store* menggunakan *FastText* dan *Long Short Term Memory* (LSTM).

1.1 Latar Belakang

Aplikasi kesehatan untuk perangkat seluler telah menunjukkan dampak signifikan pada sektor kesehatan dan kebugaran individu. Aplikasi seperti PeduliLindungi, misalnya, dapat membantu dalam manajemen kondisi kronis dan meningkatkan kepatuhan terhadap pengobatan. Selain itu, aplikasi tersebut memungkinkan pemantauan jarak jauh, yang berkontribusi pada efisiensi sistem perawatan kesehatan dan pengurangan biaya perawatan (Marcolino dkk., 2018). PeduliLindungi awalnya dikembangkan untuk mengawasi penyebaran Covid-19, namun kemudian berkembang menjadi solusi kesehatan yang lebih luas.

Aplikasi PeduliLindungi resmi diubah pada 1 Maret 2023 menjadi SatuSehat setelah mengalami pembaruan fitur antarmuka. Kementerian Kesehatan Republik Indonesia (Kemkes RI) melakukan transformasi untuk meningkatkan efektivitas aplikasi dalam menangani pandemi dan memperkuat sistem kesehatan nasional. Dengan dukungan infrastruktur *platform-as-a-service* (PaaS), SatuSehat membantu pelaku industri kesehatan terhubung ke data nasional yang dapat diandalkan. Pembaruan terbaru pada 29 Juni 2023 menghadirkan fitur baru guna memperluas peran aplikasi dalam layanan kesehatan masyarakat (Kemenkes RI, 2022).

Sebagai penyedia teknologi kesehatan di Indonesia, SatuSehat harus terus berkembang dan berinovasi sebagai usaha untuk memaksimalkan layanan terhadap pengguna. Oleh karena itu, umpan balik pengguna menjadi sumber yang bermanfaat untuk pengembangan aplikasi. Ulasan di Google Play Store dapat digunakan oleh pengguna untuk memberikan kritik dan saran untuk aplikasi SatuSehat. Sebelum pengguna mengisi ulasan diharuskan

untuk memberikan nilai dalam bentuk bintang dengan nilai mulai dari satu hingga lima. Namun, ada kejadian pengguna memberikan penilaian yang tinggi disertai dengan komentar ketidakpuasan pengguna terhadap suatu hal. Oleh karena itu, analisis sentimen dibutuhkan untuk menentukan apakah kalimat itu sebagai kalimat positif atau negatif.

Analisis sentimen, juga disebut *opinion mining*, adalah subbidang dari linguistik komputasional dan pemrosesan bahasa alami (NLP). Analisis sentimen mencakup proses untuk mengekstrak, mengkategorikan, memahami, dan menilai opini (Aftab dkk., 2023). Dalam konteks komunikasi politik, akademisi menggunakan analisis sentimen untuk mengumpulkan opini publik tentang calon presiden dan memprediksi hasil pemilihan yang akurat. Analisis sentimen sering digunakan dalam bidang ilmu data, pengolahan bahasa natural, dan analisis teks. Analisis sentimen tidak hanya terbatas pada bidang politik, tetapi juga sering digunakan dalam ilmu data, pengolahan bahasa alami, dan analisis teks. Pendekatan tersebut bermanfaat untuk berbagai keperluan, seperti menilai kepuasan pengguna terhadap suatu produk dan memantau reputasi merek.

Dalam pemrosesan bahasa alami atau NLP, terdapat proses penting untuk mengubah kata ke dalam vektor, yang memungkinkan kata-kata dengan makna serupa memiliki representasi yang sama. Proses tersebut dikenal sebagai *embedding* kata, yang memungkinkan model untuk memahami konteks semantik dan hubungan antar kata dalam suatu teks. Representasi kata yang baik sangat penting, karena pemilihan teknik *embedding* yang tepat dapat memengaruhi kinerja dan akurasi pembelajaran model, sehingga berperan besar dalam kesuksesan penerapan model tersebut dalam berbagai tugas NLP seperti klasifikasi teks, penerjemahan, dan analisis sentimen.

Salah satu algoritma yang digunakan dalam pemrosesan bahasa alami adalah *FastText*, yang memetakan kata ke dalam vektor. *FastText*, yang merupakan pengembangan dari *Word2Vec*, menggunakan informasi subword (n-gram) untuk mempelajari representasi kata, berbeda dengan *Word2Vec* yang mengabaikan struktur kata. Bojanowski dkk. (2017), pendekatan tersebut membuat *FastText* lebih efektif dalam menangani kata-kata yang jarang muncul atau tidak ada dalam korpus, karena dapat memanfaatkan pola subword untuk menghasilkan representasi yang lebih akurat.

Penelitian yang dilakukan oleh Nurdin dkk. (2020) membandingkan kinerja tiga metode *word embedding* *Word2Vec*, *GloVe*, dan *FastText* dalam klasifikasi. Hasilnya

menunjukkan bahwa *FastText* memiliki kinerja terbaik dengan *f1-score* sebesar 0,979 untuk *dataset 20 Newsgroup* dan 0,715 untuk *Reuters News*, dibandingkan dua metode lainnya. Keunggulan *FastText* terletak pada kemampuannya menangani masalah *out of vocabulary* (OOV), yakni dapat merepresentasikan kata yang tidak ada dalam data latih. Meskipun *Word2Vec* dan *GloVe* memiliki pendekatan berbeda dalam mempelajari semantik kata. *FastText* dibangun dari *Word2Vec* dengan memetakan *subword* menunjukkan kinerja yang lebih baik dalam menghasilkan representasi kata yang lebih fleksibel dan cepat.

Berbagai penelitian telah dilakukan dalam menggunakan algoritma dan teknik yang berbeda untuk analisis sentimen untuk memahami pengalaman pengguna. Pulungan dkk. (2022) menggunakan TF-IDF dan *Random Forest* pada ulasan aplikasi PeduliLindungi, mencapai akurasi 72% dengan *precision*, *recall*, dan *f1-score* 71%. Penelitian oleh Abror (2023) meningkatkan akurasi SVM dari 74,33% menjadi 84% dengan seleksi fitur *information gain*. Sementara itu Lustiansyah dkk. (2022) menggunakan *Long Short Term Memory* (LSTM) pada 3000 ulasan aplikasi PeduliLindungi, dengan hasil akurasi 82,44%, menunjukkan potensi LSTM dalam memahami opini publik. Studi lain oleh Islamy dkk. (2022) mengombinasikan *FastText* dan LSTM untuk mengklasifikasikan 5.000 data IMDB *movie reviews*, mencapai akurasi terbaik 86,3%. Hasil tersebut menunjukkan bahwa berbagai pendekatan, mulai dari teknik *word embedding* hingga model *deep learning* seperti LSTM, menawarkan kelebihan unik dalam menganalisis data opini dari beragam platform.

Model LSTM merupakan salah satu jenis variasi dari *Recurrent Neural Network* (RNN) yang dirancang untuk memproses data dengan urutan panjang. LSTM bisa digunakan dalam pemrosesan data untuk membuat klasifikasi dan prediksi. Kemampuan LSTM untuk mengingat informasi penting melalui *memory cell*, yang mengatasi masalah *vanishing gradient* yang sering terjadi pada RNN saat menangani dependensi jangka panjang (Isnain dkk., 2022). LSTM dilengkapi dengan tiga gerbang, yaitu *input gate*, *forget gate*, dan *output gate*. Gerbang-gerbang LSTM mengatur informasi yang diperbarui dan disimpan, dan *forget gate* menentukan apakah informasi akan dilupakan atau dipertahankan. yang memungkinkan LSTM untuk menangani masalah *vanishing gradient* (Muhammad dkk., 2021).

Pelabelan dalam analisis sentimen adalah proses penting untuk mengkategorikan teks ke dalam kelas seperti positif, negatif, atau netral, yang berfungsi sebagai *ground truth* bagi algoritma pembelajaran mesin. Proses tersebut membantu model *FastText* dan LSTM dalam

memahami pola data dan mengklasifikasikan ulasan dengan akurat. Penelitian oleh Fathoni dkk. (2024) membandingkan metode *Indonesian Sentiment Lexicon (InSet Lexicon)* dan *Valance Aware Dictionary and Sentiment Reasoner (VADER)*. Hasil menunjukkan bahwa *InSet Lexicon* memiliki akurasi rata-rata 87,83%, sedikit lebih tinggi daripada VADER yang mencapai 87,66%. Hasil mengindikasikan bahwa *InSet Lexicon* lebih efektif dalam meningkatkan akurasi klasifikasi, memberikan wawasan penting dalam pemilihan metode pelabelan untuk analisis sentimen ulasan aplikasi.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk menganalisis sentimen ulasan aplikasi SatuSehat yang diperoleh dari Google Play Store dengan menggunakan pelabelan *InSet Lexicon* dan representasi kata melalui *word embedding FastText*, serta model LSTM. Untuk memastikan hasil analisis yang optimal, *tuning hyperparameter* dilakukan guna meningkatkan kinerja model LSTM dalam menghasilkan prediksi sentimen yang akurat.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang dijelaskan sebelumnya, maka dirumuskan permasalahan yaitu bagaimana mengembangkan analisis sentimen ulasan aplikasi SatuSehat di Google Play Store dengan menerapkan *FastText word embedding* dan model LSTM.

1.3 Tujuan dan Manfaat

Tujuan penelitian ini secara umum adalah mendapatkan model analisis sentimen ulasan aplikasi SatuSehat menggunakan *FastText word embedding* dan model LSTM. Tujuan khusus dari penelitian ini adalah untuk mengidentifikasi pengalaman subjektif pengguna aplikasi SatuSehat dan mengelompokkannya ke dalam kelas positif dan kelas negatif berdasarkan pendapat pengguna.

Adapun manfaat dari penelitian skripsi ini adalah sebagai berikut:

1. Memberikan rekomendasi evaluasi yang dapat digunakan pengelola layanan Kementerian Kesehatan Republik Indonesia dalam menilai kekurangan dan kelebihan fitur pada aplikasi SatuSehat berdasarkan ulasan pengguna di Google Play Store.
2. Menjadi sumber referensi untuk penelitian selanjutnya terkait analisis sentimen menggunakan *FastText word embedding* dan model LSTM.

1.4 Ruang Lingkup

Ruang lingkup dari penelitian skripsi ini adalah sebagai berikut:

1. Data yang digunakan berasal dari *scraping* ulasan berbahasa Indonesia pada aplikasi SatuSehat di Google Play Store.
2. Data ulasan yang diambil mencakup periode ulasan dari Januari 2023 hingga Desember 2023.
3. Sebanyak 16.749 ulasan, dengan 4.840 ulasan berlabel positif dan 11.909 ulasan berlabel negatif, dilabeli berdasarkan bobot kata menggunakan *InSet* Lexicon.

1.5 Sistematika Penulisan

Sistem penulisan yang digunakan dalam skripsi ini terbagi dalam beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab ini memaparkan latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan penelitian mengenai analisis sentimen ulasan aplikasi SatuSehat di Google Play Store menggunakan *FastText* dan LSTM.

BAB II LANDASAN TEORI

Bab ini menyajikan studi pustaka mengenai dasar teori yang mendukung penelitian mengenai analisis sentimen ulasan aplikasi SatuSehat di Google Play Store menggunakan *FastText* dan LSTM.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan pelaksanaan penelitian yang meliputi pengumpulan data, *preprocessing data*, pembobotan kata dengan *FastText*, pembentukan model LSTM, dan evaluasi kinerja model dengan menggunakan *confusion matrix*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas mengenai hasil pengujian dan analisis dari hasil analisis sentimen ulasan aplikasi SatuSehat di Google Play Store menggunakan *FastText word embedding* dan LSTM.

BAB V PENUTUP

Bab ini berisi kesimpulan yang didapatkan dari penelitian tentang analisis sentimen ulasan aplikasi SatuSehat di Google Play Store menggunakan *FastText* dan LSTM, serta saran yang dapat diterapkan untuk pengembangan lebih lanjut.