

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi digital telah mengubah lanskap industri pariwisata, di mana citra destinasi tidak lagi hanya ditentukan oleh promosi formal, tetapi juga oleh ulasan wisatawan pada platform digital seperti Google Maps, Tripadvisor, dan media sosial. Fenomena electronic word-of-mouth (eWOM) ini menjadi faktor penting dalam memengaruhi keputusan calon wisatawan. Ulasan tersebut merepresentasikan sumber data yang bernilai karena mampu menyediakan umpan balik yang otentik dan berkelanjutan bagi pemerintah maupun pengelola wisata dalam melakukan evaluasi berbasis data (Husna & Prasetya, 2024). Namun, data ulasan memiliki karakteristik tidak terstruktur dan mengandung noise, seperti penggunaan bahasa informal, singkatan, serta kesalahan ketik (typographical errors), yang menjadi tantangan dalam proses analisis teks (Li & Li, 2022). Selain itu, distribusi data sentimen yang tidak seimbang juga berpotensi menurunkan performa model klasifikasi. Penelitian-penelitian sebelumnya umumnya belum mengintegrasikan penanganan kesalahan ketik, ketidakseimbangan data, dan pengayaan konteks fitur secara simultan dalam satu kerangka analisis.

Oleh karena itu, penelitian ini mengusulkan pendekatan berbasis Long Short-Term Memory (LSTM) yang diintegrasikan dengan teknik Damerau-Levenshtein Distance (DLD) untuk koreksi kesalahan ketik, Text Augmentation untuk penyeimbangan data, serta N-Gram untuk memperkaya representasi fitur, guna meningkatkan akurasi analisis sentimen pada ulasan wisata pantai di Kabupaten Sampang.

Meskipun metode Long Short-Term Memory (LSTM) telah terbukti efektif dalam analisis sentimen dan mampu mengungguli metode machine learning klasik seperti Naïve Bayes dan SVM (Ahmed et al., 2025; Tripathi, 2021), serta relevan dalam domain pariwisata (Sedana et al., 2024), penelitian-penelitian sebelumnya masih memiliki keterbatasan dalam menangani karakteristik data ulasan dunia nyata. Permasalahan utama yang sering muncul meliputi ketidakseimbangan distribusi data sentimen, rendahnya kualitas teks akibat kesalahan ketik, serta keterbatasan dalam menangkap konteks lokal berbasis frasa. Sebagian besar

penelitian cenderung hanya berfokus pada satu aspek permasalahan, seperti penggunaan oversampling untuk mengatasi data tidak seimbang atau pemanfaatan N-gram untuk ekstraksi fitur (Al-Hadhrami et al., 2024) tanpa mengintegrasikan ketiga pendekatan tersebut secara simultan dalam satu kerangka analisis.

Berdasarkan pemetaan literatur dan identifikasi celah penelitian yang mendalam tersebut, penelitian ini mengajukan sebuah kerangka kerja analisis sentimen yang holistik dan tangguh. Penelitian-penelitian sebelumnya cenderung fokus pada satu atau dua aspek permasalahan saja, namun penelitian ini secara unik mengintegrasikan solusi untuk tiga tantangan sentral secara simultan: (1) *Text Augmentation* untuk mengatasi ketidakseimbangan data, (2) *Damerau-Levenshtein Distance* untuk meningkatkan kualitas teks dengan koreksi *typo*, dan (3) fitur *N-gram* untuk memperkaya representasi konteks lokal, yang semuanya diorkestrasi dalam sebuah model klasifikasi LSTM. Dengan mengimplementasikan pendekatan multifaset ini pada studi kasus ulasan wisata pantai di Kabupaten Sampang, penelitian ini tidak hanya memiliki kebaruan (*novelty*) akademis yang tinggi, tetapi juga menawarkan manfaat praktis yang signifikan. Tujuan akhirnya adalah untuk membangun dan mengevaluasi model analisis sentimen yang andal dalam mengklasifikasikan ulasan wisatawan sehingga dapat menghasilkan wawasan yang mendukung pengembangan pariwisata berbasis data di Kabupaten Sampang.

1.2 Rumusan Masalah

Pada bagian perumusan masalah, diuraikan mengenai permasalahan, metode penelitian serta pertanyaan penelitian yang diuraikan sebagai berikut :

1.1.1. Permasalahan

Berdasarkan uraian latar belakang, dapat diambil permasalahan penelitian diantaranya sebagai berikut :

1. Belum tersedianya analisis kuantitatif yang komprehensif mengenai proporsi dan distribusi sentimen (positif, negatif, dan netral) dari wisatawan terhadap destinasi pantai di Kabupaten Sampang. Kondisi ini menunjukkan perlunya dilakukan analisis sentimen yang lebih mendalam dan sistematis terhadap ulasan wisatawan guna memperoleh gambaran yang objektif terkait persepsi pengunjung
2. Perlunya validasi performa model yang diusulkan secara kuantitatif

melalui metrik evaluasi standar (*Accuracy*, *Precision*, *Recall*, dan *F1-Score*) untuk memastikan bahwa model tersebut memiliki tingkat keandalan dan akurasi yang tinggi sehingga layak digunakan sebagai alat bantu evaluasi pariwisata yang dapat dipertanggungjawabkan secara ilmiah.

1.1.2. Metode penelitian

Penelitian ini mengusulkan sebuah model analisis sentimen berbasis *Long Short-Term Memory* (LSTM) yang diperkuat dengan serangkaian teknik pra-pemrosesan dan ekstraksi fitur canggih untuk mengevaluasi ulasan wisata pantai di Kabupaten Sampang. Arsitektur LSTM dipilih karena kemampuannya yang unggul dalam memahami konteks sekuensial dalam kalimat ulasan. Tidak seperti pendekatan standar yang rentan terhadap data berkualitas rendah, model ini secara proaktif menangani tiga tantangan utama: algoritma *Damerau-Levenshtein Distance* (DLD) untuk mengoreksi kesalahan ketik, teknik *Text Augmentation* untuk menyeimbangkan distribusi kelas sentimen, dan fitur *N-gram* untuk menangkap makna frasa lokal yang krusial. Melalui pendekatan terintegrasi ini, dapat dihasilkan model analisis sentimen yang lebih tangguh dan akurat, sehingga mampu memberikan wawasan yang valid mengenai persepsi wisatawan untuk mendukung strategi pengembangan pariwisata lokal.

1.1.3. Pertanyaan Penelitian

Berdasarkan uraian latar belakang, dapat diambil pernyataan penelitian diantaranya sebagai berikut :

1. Bagaimana proporsi dan distribusi sentimen (positif, negatif, dan netral) dari wisatawan terhadap destinasi pantai di Kabupaten Sampang berdasarkan analisis ulasan wisatawan secara kuantitatif ?
2. Bagaimana efektivitas penerapan algoritma *Damerau-Levenshtein Distance* (DLD), teknik *Text Augmentation*, dan fitur *N-gram* secara terintegrasi untuk mengatasi tantangan spesifik pada data ulasan (kesalahan ketik, data tidak seimbang, dan makna frasa) guna meningkatkan performa model klasifikasi tiga sentimen ?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Berdasarkan rumusan masalah yang ada, dapat ditentukan tujuan dari penelitian ini diantaranya adalah sebagai berikut :

1. Untuk mengetahui proporsi dan distribusi sentimen (positif, negatif, dan netral) dari wisatawan terhadap destinasi pantai di Kabupaten Sampang berdasarkan data ulasan wisatawan. Penelitian ini juga bertujuan untuk menghasilkan informasi kuantitatif yang sistematis mengenai persepsi pengunjung guna mendukung proses evaluasi dan peningkatan kualitas destinasi wisata
2. Membangun dan mengevaluasi sebuah model klasifikasi sentimen berbasis Long Short-Term Memory (LSTM) yang secara terintegrasi menerapkan algoritma Damerau-Levenshtein Distance (DLD), teknik Text Augmentation, dan fitur N-gram untuk mengatasi tantangan data ulasan (seperti kesalahan ketik dan data tidak seimbang) sehingga mampu menghasilkan analisis yang akurat dan andal.
3. Menganalisis dan membandingkan pengaruh penerapan Damerau-Levenshtein Distance (DLD), Text Augmentation, serta variasi fitur N-gram (unigram, bigram, dan trigram) terhadap performa model Long Short-Term Memory (LSTM) berdasarkan metrik evaluasi seperti accuracy, precision, recall, dan F1-score.

1.3.2 Manfaat

Berdasarkan rumusan masalah dan tujuan yang ada, dapat ditentukan manfaat dari penelitian ini diantaranya adalah sebagai berikut:

1. Memberikan kontribusi pada bidang *Natural Language Processing* (NLP) dengan mengusulkan dan menguji sebuah kerangka kerja (*framework*) komprehensif untuk menangani data teks dunia nyata (*real-world Text*). Integrasi spesifik antara algoritma *Damerau-Levenshtein Distance*, teknik *Text Augmentation*, dan fitur *N-gram* dengan arsitektur LSTM menjadi sebuah studi kasus yang valid mengenai cara mengatasi tantangan data ulasan yang informal, mengandung kesalahan ketik, dan tidak seimbang.

2. Kerangka kerja dan hasil evaluasi model dari penelitian ini dapat dijadikan sebagai *baseline* atau titik acuan bagi para peneliti selanjutnya. Ini memberikan landasan yang kuat bagi pengembangan metode analisis sentimen yang lebih canggih di masa depan, baik pada domain pariwisata maupun domain lainnya yang memiliki karakteristik data serupa.
3. Hasil analisis sentimen dapat menjadi dasar pengambilan keputusan yang berbasis data (*data-driven decision making*). Pemerintah daerah dapat memperoleh wawasan kuantitatif mengenai persepsi publik, mengidentifikasi destinasi pantai mana yang memerlukan perhatian khusus, serta merumuskan kebijakan dan program promosi yang lebih tepat sasaran untuk meningkatkan citra pariwisata daerah.

1.4 Batasan Masalah

Berikut batasan-batasan yang akan diberikan agar permasalahan tidak terlalu luas dan tidak menyimpang sehingga sesuai dengan tujuan penulis berikut ini :

1. Fokus penelitian ini adalah data ulasan yang berasal dari platform Google Maps. Penelitian ini tidak akan mengambil atau menganalisis data dari platform lain seperti Tripadvisor, Twitter, Instagram, atau blog perjalanan.
2. Objek yang dianalisis secara spesifik adalah destinasi wisata pantai yang berada di wilayah administratif Kabupaten Sampang, Madura.
3. Analisis sentimen hanya dilakukan pada ulasan yang ditulis dalam Bahasa Indonesia.
4. Model *deep learning* yang dibangun dan dievaluasi secara mendalam dalam penelitian ini adalah arsitektur *Long Short-Term Memory* (LSTM).
5. Klasifikasi sentimen terbatas pada tiga kategori polaritas utama, yaitu Positif, Negatif, dan Netral pada level ulasan (dokumen).
6. Implementasi algoritma *Damerau-Levenshtein Distance* (DLD) difokuskan untuk mengoreksi kesalahan ketik (*typo*) dan tidak dirancang untuk menangani normalisasi kata gaul atau singkatan yang tidak baku.

1.5 Sistematika Penulisan Skripsi

Sistematika penulisan skripsi ini penulis susun dengan mengikuti aturan penulisan proposal pada program studi S1 Sistem Informasi Universitas Trunodjoyo Madura. Berikut merupakan metodologi penulisan proposal pada penelitian ini.

BAB I PENDAHULUAN

Menjelaskan mengenai latar belakang, perumusan masalah, permasalahan, metode penelitian, pertanyaan penelitian, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan terkait penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Berisi teori-teori penunjang yang berkaitan dengan Analisis Sentimen melalui pendekatan LSTM dengan *Text Augmentation*, DLD, dan *N-gram*. Selain itu, bab ini mencakup penelitian-penelitian terdahulu yang relevan dengan dengan proses NLP.

BAB III METODE PENELITIAN

Membahas metode yang diusulkan, termasuk metode penelitian, tahapan penelitian, proses pengolahan dataset, penerapan model LSTM, serta skenario pengujian dan evaluasi model.

BAB IV HASIL DAN PEMBAHASAN

Menjelaskan hasil dari penerapan metode *Long Short-Term Memory* (LSTM) dalam menganalisis sentimen ulasan wisata pantai di Kabupaten Sampang. Pembahasan mencakup tahapan penelitian yang meliputi proses pengumpulan data, *preprocessing* teks, penerapan teknik *text augmentation* dan *data labeling* (DLD), ekstraksi fitur menggunakan pendekatan *N-gram*, pembagian data latih dan data uji, proses pelatihan model LSTM, hingga evaluasi performa model menggunakan metrik yang relevan. Selain itu, bab ini juga membahas hasil perbandingan performa model dengan dan tanpa penerapan teknik *text augmentation*, DLD, serta variasi fitur *N-gram* untuk mengetahui konfigurasi terbaik dalam analisis sentimen.

BAB V PENUTUP

Menjelaskan kesimpulan dari hasil penelitian yang telah dilakukan terkait analisis sentimen ulasan wisata pantai di Kabupaten Sampang menggunakan metode LSTM dengan teknik *text augmentation*, DLD, dan fitur *N-gram*. Kesimpulan disusun berdasarkan hasil analisis dan evaluasi performa model yang telah dilakukan. Selain itu, bab ini juga memuat saran untuk pengembangan

penelitian selanjutnya, baik dari segi metode, pengolahan data, maupun penerapan teknik lain yang dapat meningkatkan akurasi dan kualitas hasil analisis sentimen di masa mendatang.

DAFTAR PUSTAKA

Memuat semua sumber rujukan yang digunakan dalam penyusunan proposal penelitian. Penulisan daftar pustaka dilakukan secara otomatis menggunakan aplikasi pengelola referensi seperti Mendeley, untuk memastikan konsistensi dan akurasi dalam pencantuman referensi