

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pariwisata merupakan salah satu sektor penting yang memiliki kontribusi besar dalam mendukung pembangunan ekonomi di setiap daerah, yang dapat meningkatkan perekonomian Indonesia dengan adanya peningkatan pendapatan negara[1]. Di Indonesia, Pulau Madura merupakan salah satu wilayah yang memiliki potensi pariwisata yang baik, khususnya di sektor wisata pantai. Kabupaten Bangkalan sebagai salah satu wilayah utama Pulau Madura yang terletak di kawasan pesisir pantai memiliki berbagai destinasi wisata pantai dengan daya tarik cukup besar dari sisi keindahan alam yang berpotensi untuk dikelola dengan baik sehingga, dapat dikenal luas dan mendatangkan wisatawan dari berbagai daerah[1][2]. Selain itu, data Badan Pusat Statistik (BPS) Kabupaten Bangkalan menunjukkan bahwa sektor transportasi dan pergudangan tumbuh paling tinggi pada tahun 2024, yaitu sebesar 11,48% [3]. Pertumbuhan tersebut berkaitan dengan aktivitas pariwisata, karena meningkatnya kunjungan wisatawan akan mendorong kebutuhan transportasi, akomodasi, serta penyediaan makanan dan minuman yang menjadi bagian penting dari ekosistem wisata.

Namun, peran pemerintah daerah Kabupaten Bangkalan dalam pengelolaan pariwisata khususnya pantai, masih belum sepenuhnya optimal. Dari hasil penelitian, dikatakan bahwa pengembangan destinasi wisata pantai di Kabupaten Bangkalan memerlukan strategi pengelolaan yang lebih terarah oleh pemerintah setempat selaku pemangku kebijakan[4]. Di sisi lain, salah satu aspek penting dalam mendukung pengembangan pariwisata adalah pemasaran berbasis media sosial serta pemanfaatan ulasan digital. Pemasaran melalui media sosial terbukti memiliki pengaruh positif terhadap kepuasan wisatawan, meskipun tidak signifikan terhadap minat berkunjung kembali. Sementara itu, daya tarik destinasi terbukti memiliki pengaruh positif dan signifikan terhadap kepuasan dan minat kunjungan ulang wisatawan[5].

Ulasan wisata yang terdapat pada platform digital seperti *Google Maps*, *TripAdvisor*, dan *TikTok* menjadi sumber informasi yang berharga bagi wisatawan saat berkunjung ke suatu tempat wisata dari berbagai aspek. Data ini dapat

dianalisis lebih jauh menggunakan analisis sentimen untuk mengetahui kecenderungan opini positif atau negatif. Hasil analisis sentimen dapat dimanfaatkan pemerintah maupun pengelola wisata dalam menilai kualitas destinasi, sekaligus menyusun strategi pengembangan wisata. Untuk mendukung hal tersebut, maka diperlukan metode klasifikasi yang tepat dalam mengelola ulasan wisata. Salah satu metode yang banyak digunakan dalam penelitian analisis sentimen adalah *Support Vector Machine*.

Pada penelitian sebelumnya menunjukkan bahwa *Support Vector Machine* (SVM) memberikan hasil yang baik dalam analisis sentimen. Penelitian pada ulasan aplikasi *TikTok* di Indonesia, menunjukkan bahwa SVM mencapai akurasi sebesar 88,76% dengan nilai *F1-Score* yang lebih seimbang dalam mendeteksi sentimen positif maupun negatif, sehingga secara konsisten lebih unggul dibandingkan metode *Naïve Bayes*[6]. Temuan tersebut memperlihatkan bahwa SVM merupakan metode yang baik digunakan dalam mengklasifikasikan sentimen pada data media sosial. Selain itu, pada penelitian yang sejalan pada domain wisata juga menunjukkan hasil yang serupa. Penelitian yang dilakukan pada ulasan destinasi wisata di Yogyakarta melalui *Google Maps*, berhasil mengklasifikasikan sentimen positif, negatif, dan netral dengan akurasi rata-rata sebesar 83,8% [7]. Kemudian pada penelitian selanjutnya, yaitu ulasan wisata Pantai Pandawa dengan pendekatan *Aspect-Based Sentiment Analysis* (ABSA) menunjukkan bahwa SVM memiliki performa yang lebih unggul dibandingkan dengan metode lain, dengan nilai *F1-Score* 79,62% sedangkan metode *Naïve Bayes* memperoleh 73,29% [8]. Sehingga, dari hasil penelitian–penelitian tersebut membuktikan bahwa SVM mampu memberikan hasil yang konsisten pada analisis sentimen ulasan wisata.

Namun dalam pemrosesan teks ulasan, salah satu tantangan yang sering muncul adalah kesalahan pengetikan (*typo*). Hal ini dapat mempengaruhi turunya kualitas analisis karena kata yang salah eja akan diperlakukan sebagai kata baku. Untuk itu, digunakan pendekatan *Damerau-Levenshtein Distance*, salah satu algoritma yang mampu digunakan untuk koreksi ejaan akibat penyisipan, penghapusan, pertukaran, dan penggantian huruf, dengan akurasi mencapai 97% [9].

Setelah tahap koreksi ejaan, proses representasi teks juga menjadi bagian penting dalam meningkatkan kualitas analisis. Optimasi representasi dapat

dilakukan dengan teknik *N-Gram*, dimana pada penelitian sebelumnya menunjukkan bahwa penerapan *Word N-Gram* mampu meningkatkan akurasi klasifikasi hingga 89% dengan nilai *AUC* 0,944, meskipun peningkatan panjang *N-Gram* juga berpengaruh terhadap dimensi data dan waktu pemrosesan[11]. Untuk mencari kombinasi terbaik antara performa dan efisiensi, diperlukan eksplorasi variasi panjang *N-Gram*.

Adapun persoalan lain yang sering terjadi adalah ketidakseimbangan data. Beberapa kelas sentimen biasanya memiliki jumlah data yang lebih banyak dibandingkan dengan kelas lain, sehingga mempengaruhi kinerja model klasifikasi. Berdasarkan hasil penelitian sebelumnya, *Text Augmentation* terbukti efektif dalam menyeimbangkan data dan meningkatkan performa klasifikasi[12]. Penelitian tersebut menunjukkan bahwa teks augmentasi mampu meningkatkan *F1 score* hingga 8%, terutama pada dataset kecil. Selain itu, beberapa studi juga menegaskan bahwa penerapan *Text Augmentation* secara konsisten memberikan dampak positif terhadap peningkatan kualitas data dan hasil klasifikasi[13]. Sehingga, dalam penelitian ini *Text Augmentation* diterapkan untuk memperbaiki kualitas data.

Berdasarkan beberapa uraian tersebut, dapat dilihat bahwa analisis sentimen ulasan wisata memiliki peran dalam membantu memberikan gambaran dari pengalaman pengunjung yang dapat dimanfaatkan oleh pemerintah maupun pengelola destinasi sebagai bahan evaluasi dan strategi pengembangan pariwisata. Namun, untuk memperoleh hasil analisis yang optimal, diperlukan metode yang tepat serta tahapan penelitian yang sistematis. Meskipun telah banyak penelitian terkait analisis sentimen pada sektor pariwisata, masih terdapat keterbatasan dalam pendekatan yang di gunakan, khususnya pada *Aspect-Based Sentiment Analysis* wisata pantai berbasis aspek di Kabupaten Bangkalan. Hingga saat ini, belum ada penelitian menggunakan *Aspect-Based Sentiment Analysis* (ABSA) yang mengkombinasikan *Support Vector Machine* (SVM), koreksi ejaan menggunakan *Damerau-Levenshtein Distance* (DLD), variasi *N-Gram*, serta penyeimbangan data menggunakan *Text Augmentation*. Oleh karena itu, penelitian ini menganalisis sentimen ulasan wisata pantai di Kabupaten Bangkalan yang diperoleh dari platform digital seperti *Google Maps*, *TripAdvisor* dan *TikTok*. Seluruh proses penelitian dilakukan dengan menerapkan kerangka kerja CRISP-DM menggunakan

algoritma *Support Vector Machine* yang dipadukan dengan koreksi ejaan melalui *Damerau-Levenshtein Distance*, optimasi fitur teks dengan *N-Gram* serta penyeimbangan data melalui *Text Augmentation*. Dengan pendekatan ini, hasil klasifikasi sentimen menjadi lebih akurat dan bermanfaat sebagai dasar pengambilan keputusan dalam pengembangan wisata pantai di Kabupaten Bangkalan.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka diperlukan beberapa rumusan permasalahan yang menjadi fokus pada penelitian. Rumusan masalah mencakup permasalahan, metode penelitian, serta pernyataan penelitian yang dijabarkan sebagai berikut:

1.2.1 Permasalahan

Sektor pariwisata merupakan salah satu bagian penting dalam pembangunan ekonomi daerah, termasuk di Kabupaten Bangkalan yang memiliki berbagai destinasi wisata pantai. Banyak wisatawan yang meninggalkan ulasan mereka melalui platform digital seperti *Google Maps*, *TripAdvisor* dan *TikTok*. Ulasan tersebut berisi informasi mengenai kepuasan, kritik, maupun saran mengenai kualitas layanan dan daya tarik wisata. Namun, dalam proses analisis sentimen, masih mengalami kendala berupa kesalahan penulisan, variasi bahasa, serta ketidakseimbangan jumlah ulasan positif dan negative yang berpotensi mengurangi akurasi hasil klasifikasi sentimen. Oleh karena itu, diperlukan metode yang mampu mengatasi masalah tersebut agar hasil dari analisis sentimen dapat lebih akurat dan bermanfaat sebagai bahan evaluasi pengembangan pariwisata pantai di Kabupaten Bangkalan.

1.2.2 Metode Usulan

Dari permasalahan yang ditemukan, pada penelitian ini menggunakan kerangka kerja CRISP-DM sebagai alur penelitian mulai dari pemahaman data hingga *deployment*. Analisis dilakukan berbasis aspek, sehingga setiap ulasan wisata dipetakan ke kategori tertentu seperti daya tarik, aksesibilitas, amenitas, dan pelayanan tambahan. Dengan demikian, hasil analisis tidak hanya menunjukkan kecenderungan sentimen secara umum, namun memberikan gambaran yang lebih rinci pada tiap aspeknya. Pada tahap pemrosesan data, digunakan metode *Damerau-*

Levenshtein Distance untuk memperbaiki kesalahan penulisan sehingga data lebih bersih dan konsisten. Selanjutnya, digunakan representasi teks dengan *N-Gram* agar hubungan antar kata dapat ditangkap dengan baik. Kemudian untuk mengatasi ketidakseimbangan data, digunakan teknik *Text Augmentation* sehingga distribusi data lebih seimbang. Terakhir, klasifikasi sentimen dilakukan menggunakan algoritma *Support Vector Machine* yang cukup efektif dalam mengklasifikasikan data teks berlabel. Dengan kombinasi metode tersebut, penelitian ini dapat menghasilkan analisis sentimen yang akurat dan relevan.

1.2.3 Pertanyaan Penelitian

Berdasarkan uraian latar belakang dan permasalahan yang telah dijelaskan, maka pertanyaan penelitian yang diajukan adalah sebagai berikut:

1. Apakah penerapan metode *Damerau-Levenshtein Distance* dapat meningkatkan kualitas data dengan memperbaiki kesalahan penulisan pada ulasan wisata pantai?
2. Apakah penggunaan *N-Gram* dapat membantu mempresentasikan teks secara lebih baik sehingga meningkatkan kinerja klasifikasi sentimen?
3. Bagaimana pengaruh penerapan *Text Augmentation* dalam mengatasi ketidakseimbangan data?
4. Apakah kombinasi *Support Vector Machine* dengan *Damerau-Levenshtein Distance*, *N-Gram* dan *Text Augmentation* dapat menghasilkan model analisis sentimen yang akurat terhadap ulasan wisata pantai di Kabupaten Bangkalan?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Tujuan dalam penelitian ini adalah sebagai berikut:

1. Mengukur peningkatan akurasi model setelah penerapan *Damerau-Levenshtein Distance* dalam memperbaiki kesalahan penulisan pada ulasan wisata pantai.
2. Mengukur seberapa baik *N-Gram* dalam mempresentasikan teks untuk membantu *Support Vector Machine* dalam mengklasifikasikan sentimen.

3. Menilai pengaruh *Text Augmentation* dalam menyeimbangkan distribusi data sentimen.
4. Mengetahui akurasi model sentimen yang dibangun dengan kombinasi *Support Vector Machine*, *Damerau-Levenshtein Distance*, *N-Gram* dan *Text Augmentation* pada ulasan wisata pantai di Kabupaten Bangkalan.

1.3.2 Manfaat

Manfaat akademis dari penelitian ini adalah:

1. Memberikan pemahaman tentang *Damerau-Levenshtein Distance* untuk memperbaiki kesalahan penulisan pada ulasan wisata pantai.
2. Menunjukkan bagaimana *N-Gram* dalam membantu model *Support Vector Machine* membaca hubungan antar kata dalam analisis sentimen.
3. Menilai pengaruh penerapan metode *Text Augmentation* dengan teknik *Back Translation* dalam menyeimbangkan distribusi data sentimen.

Manfaat praktis dari penelitian ini adalah:

1. Memberikan informasi yang mudah dipahami oleh pemerintah atau pengelola wisata pantai untuk memperbaiki dan meningkatkan kepuasan pengunjung.
2. Menjadi referensi bagi pengambil keputusan dalam merancang strategi pengelolaan dan promosi destinasi wisata berbasis analisis ulasan digital.

1.4 Batasan Masalah

1. Data berfokus pada ulasan wisata pantai yang berada di Kabupaten Bangkalan yang diambil dari berbagai platform digital, yaitu *Google Maps* diperoleh sebanyak 1.838 ulasan, dari *TripAdvisor* sebanyak 16 ulasan dan dari *TikTok* sebanyak 624 ulasan.
2. Analisis hanya berfokus pada ulasan teks yang memuat opini wisatawan.
3. Fokus penelitian adalah penerapan *Damerau-Levenshtein Distance* untuk perbaikan kesalahan penulisan, *N-Gram* (*unigram*, *bigram*, *trigram*) untuk representasi teks dan *Text Augmentation* untuk menyeimbangkan data.
4. Algoritma yang digunakan adalah *Support Vector Machine*.

1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun dengan mengikuti aturan penulisan skripsi pada program studi S1 Sistem Informasi Universitas Trunojoyo Madura:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, permasalahan, metode usulan, pernyataan penelitian, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bab ini membahas teori-teori dan penelitian terdahulu yang menjadi dasar dalam penelitian ini, termasuk teori analisis sentimen, *Support Vector Machine*, *Damerau-Levenshtein Distance*, *N-Gram* dan *Text Augmentation*.

BAB III METODE PENELITIAN

Bab ini menjelaskan arsitektur sistem analisis sentimen, sumber dan pengumpulan data, pelabelan, penggunaan *Text Augmentation*, *Preprocessing* data, perbaikan kata menggunakan *Damerau-Levenshtein Distance*, ekstraksi fitur *N-Gram* dan TF-IDF, klasifikasi dengan *Support Vector Machine*, evaluasi model, serta skenario uji coba.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi metode yang telah diusulkan, meliputi hasil pengolahan data, hasil pengujian model analisis sentimen multi-aspek, serta pembahasan terhadap performa model berdasarkan nilai evaluasi yang diperoleh.

BAB V PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran yang dapat digunakan sebagai bahan pertimbangan untuk pengembangan penelitian selanjutnya.