

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pariwisata saat ini telah berkembang secara signifikan baik dari segi ekonomi maupun sosial pada suatu negara dan khususnya pada destinasi wisata di wilayah tersebut (Rahma, 2020). Pertumbuhan permintaan dalam sektor pariwisata memiliki dampak langsung terhadap peredaran uang, barang, dan jasa, membentuk aliran ekonomi yang luas dan cepat (Isdarmanto, 2017). Dampak ekonomi ini mencakup aspek signifikan seperti pertumbuhan Produk Domestik Bruto (PDB). Aktivitas pariwisata juga memiliki kapasitas untuk menghasilkan devisa melalui kunjungan wisatawan asing (Eddyono, 2021).



Gambar 1. 1. Kontribusi PDB Pariwisata Indonesia ([DataIndonesia.Id](https://dataindonesia.id))

Berdasarkan Data Indonesia ([DataIndonesia.id](https://dataindonesia.id)) pada Gambar 1.1. yang bersumber dari Kementerian Pariwisata dan Ekonomi Kreatif (Kemenparekraf) mencatat kontribusi pariwisata terhadap produk domestik bruto (PDB) Indonesia pada tahun 2022 sebesar 3,6% dan mengalami kenaikan hingga september 2023 sebesar 3,8%. Sementara pada tahun 2020 hingga 2021 kontribusi pariwisata terhadap produk domestik bruto (PDB) Indonesia mengalami penurunan akibat pandemi Covid-19. Peningkatan kontribusi PDB menunjukkan pemulihan sektor pariwisata pasca pandemi semakin membaik.

Meskipun secara kontribusi PDB dan kunjungan wisatawan mancanegara mulai mengalami kenaikan serta banyaknya destinasi pariwisata yang tersebar di Indonesia,

Bali tetap menjadi destinasi pariwisata yang memiliki daya tarik utama bagi para wisatawan, terutama objek wisata pantainya yang terkenal hingga mancanegara (Anzela & Rachmawati, 2023). Berdasarkan hal tersebut, Kementerian Pariwisata dan Ekonomi Kreatif berupaya meningkatkan jumlah kunjungan wisatawan ke berbagai destinasi pariwisata di Indonesia di luar pulau Bali. Upaya tersebut diwujudkan dengan membentuk berbagai strategi, salah satunya pembangunan Destinasi Super Prioritas (DSP).

Strategi pembangunan DSP bertujuan untuk meningkatkan kunjungan wisatawan ke berbagai destinasi di Indonesia selain Bali, baik untuk kunjungan wisatawan domestik maupun mancanegara. Destinasi Super Prioritas tersebut diantaranya Danau Toba di provinsi Sumatera Utara, Borobudur di provinsi Jawa Tengah, Mandalika di provinsi Nusa Tenggara Barat (NTB), Labuan Bajo di provinsi Nusa Tenggara Timur (NTT), dan Likupang di provinsi Sulawesi Utara. Pembentukan DSP tersebut dipilih karena merupakan Kawasan Strategis Pariwisata Nasional yang tersebar di bagian barat, tengah, dan timur provinsi di Indonesia (Pambudi et al., 2020). Dalam penelitian terkait diplomasi komersial lima DSP (Anzela & Rachmawati, 2023), strategi yang dapat dilakukan yaitu dengan mempromosikan DSP tersebut melalui media sosial, pameran, hingga dilakukannya kerja sama sehingga kunjungan wisatawan domestik dan mancanegara dapat meningkat.

Proses pembangunan DSP sudah dibangun semenjak 2021 dan diproyeksikan selesai pada tahun 2024. Akan tetapi, hingga saat ini belum banyak yang memanfaatkan pendapat atau opini masyarakat terkait pengembangan DSP tersebut. Padahal, suatu opini pengguna media sosial dapat memiliki efek yang signifikan terhadap citra destinasi tersebut (Rochmah & Lentera, 2023). Saat ini, media sosial mulai dijadikan sebagai parameter untuk sebuah penilaian karena terdapat informasi, penilaian, curhatan pribadi yang dapat dijadikan orang sebagai referensi seperti *review* film, makanan, tempat wisata (Somantri & Dairoh, 2019), dan sebagainya (Cahyono, 2017). Sehingga, untuk mengeksplorasi nilai dari informasi yang berupa sentimen atau opini terkait DSP dapat dilakukan melalui analisis sentimen.

Analisis Sentimen merupakan bidang studi dalam Pemrosesan Bahasa Alami atau *Natural Language Processing* (NLP) yang berkaitan dengan mengidentifikasi dan menganalisis suasana hati, sentimen, dan opini dari elemen subjektif dalam sebuah

teks (Bhadane et al., 2015). NLP memungkinkan transformasi data yang tidak terstruktur dengan cepat menjadi *insight* yang dapat ditindaklanjuti (Obembe et al., 2021). Twitter salah satu media sosial yang banyak dijadikan tempat penelitian terkait data tidak terstruktur karena berbentuk teks, yang dapat membangun sentimen, pandangan, atau opini (Bagheri & Jahirul, 2018). Karena prosesnya yang mengekstrak opini-opini publik yang tersebar melalui media sosial berbasis teks, Twitter menjadi media sosial yang tepat untuk menganalisis sentimen dari opini-opini pengguna Twitter. Metode analisis sentimen berbasis kamus (*dictionary-based sentiment analysis*) juga menjadi pendekatan yang penting dalam menginterpretasikan sentimen dari teks yang diposting di Twitter. Metode ini menggunakan kamus kata-kata dengan nilai sentimen yang sudah ditentukan untuk menghitung sentimen keseluruhan dari sebuah teks.

Penggunaan Twitter digunakan secara luas dalam berbagai situasi dan ribuan akademisi dan peneliti di semua tingkat pengalaman dan di semua disiplin ilmu sudah menggunakan Twitter setiap hari (Mollett et al., 2015). Karakteristik Twitter yang paling penting ialah kemampuannya bagi pengguna untuk memposting *tweet* tentang peristiwa, situasi, perasaan, opini, atau bahkan sesuatu yang baru, secara *real-time* (Yadranjiaghdam et al., 2017) menyebabkan semakin banyaknya data yang dihasilkan (Pandu et al., 2020). Dengan karakteristik Twitter yang seperti ini, maka seharusnya dapat digunakan secara efektif untuk memposting secara *real-time* mengenai destinasi wisata. Namun, efektivitas Twitter dalam menganalisis data *real-time* untuk keperluan pariwisata, khususnya destinasi super prioritas, belum banyak diteliti.

Mengingat postingan twitter yang terus-menerus muncul dengan kecepatan dan volume yang meningkat, analisis sentimen dengan mengumpulkan *tweet* dilakukan menggunakan big data platform. Pemrosesan big data umumnya membutuhkan pemrosesan data yang kompleks secara terus-menerus dihasilkan dengan kecepatan tinggi untuk memproses secara efisien dalam *real time* atau mendekati *real time* (Chung & Aring, 2018). Menganalisis data secara *real-time* membutuhkan konsumsi data dan *stream processing* (pemrosesan aliran) data (Cha & Wachowicz, 2015).

Penelitian terkait analisis sentimen sebelumnya dilakukan oleh (Hidayat et al., 2021) yang mengklasifikasikan opini publik tentang topik memindahkan Ibu Kota Indonesia dari ribuan tweet yang telah dikumpulkan menggunakan *Support Vector*

Machine dan TF-IDF, penelitian oleh (Rochmah & Lentera, 2023) yang menerapkan *text mining* dan *software* Orange dalam menganalisis eWOM (*electronic word of mouth*) dari media sosial Twitter terhadap Destinasi Wisata Super Prioritas. Penelitian oleh (Pandu et al., 2020) menganalisis sentimen pasca pengumuman hasil pilpres 2019 yang diambil dari data tweet dengan metode *lexicon analysis* dan menggunakan R. Penelitian terkait analisis sentiment secara *real-time* sebelumnya dilakukan oleh (Tumbel et al., 2017) dengan melakukan analisis sentimen data layanan Uber berbentuk CSV menggunakan Apache Spark untuk pengolahan data secara real-time dengan metode *windowed real-time stream processing*. Penelitian oleh (Elzayady et al., 2019) tentang analisis sentimen dengan bantuan Apache Spark memanfaatkan library MLIB untuk menangani jumlah big data pada data twitter, kemudian melakukan klasifikasi sentimen menggunakan algoritma klasifikasi *Naïve Bayes*, *Logistic Regression*, dan *Decision Tree* dan membandingkan kinerja atau akurasi ketiga algoritma tersebut. Penelitian lain yang masih berkaitan dengan analisis real-time dan streaming oleh (Anupama et al., 2020; Sotiris et al., 2018) yang menganalisis sentimen terhadap ulasan film dan Brexit dari twitter.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen menggunakan big data platform, dengan pendekatan *Dictionary-Based Sentiment Analysis*, untuk memantau dan menganalisis sentimen publik tentang destinasi super prioritas di Twitter. Penelitian ini memanfaatkan big data platform Apache Kafka untuk proses konsumsi dan streaming data dan Apache Spark untuk melakukan pemrosesan data dalam mengolah dan menganalisis data yang masuk dari Apache Kafka. Apache Kafka efektif untuk membangun rute data streaming real-time sedangkan Apache Spark memfasilitasi pemrosesan data stream yang efektif untuk menganalisis sentimen tentang pariwisata khususnya destinasi super prioritas di Twitter. Hasil analisis dalam penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan sistem analisis sentimen yang lebih canggih dan dapat diaplikasikan dalam berbagai industri termasuk pariwisata, dapat dijadikan sebagai pendukung keputusan baik itu untuk wisatawan maupun pengelola tempat wisata, rekomendasi pimpinan daerah dalam hal peningkatan fasilitas dan layanan, analisis keluhan wisatawan secara real time, promosi, dan lainnya.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka secara umum permasalahan yang dibahas dalam penelitian adalah sebagai berikut:

1. Bagaimana perancangan model komputasi analisis sentimen Destinasi Super Prioritas pada media sosial Twitter berbasis *stream processing* menggunakan big data platform Apache Kafka dan Apache Spark?
2. Bagaimana implementasi model komputasi analisis sentimen Destinasi Super Prioritas pada media sosial Twitter berbasis *stream processing* dengan menerapkan Apache Kafka dan Apache Spark?
3. Bagaimana hasil eksperimen analisis sentimen Destinasi Super Prioritas pada media sosial Twitter berbasis *stream processing* dengan Apache Kafka dan Apache Spark?
4. Bagaimana kinerja komputasi sebagai analisis sentimen Destinasi Super Prioritas pada Twitter secara streaming?

1.3. Tujuan Penelitian

Dengan mengacu pada rumusan masalah, maka tujuan penelitian ini adalah sebagai berikut:

1. Merancang model komputasi analisis sentimen Destinasi Super Prioritas pada media sosial Twitter secara *streaming* dan *real-time* menggunakan big data platform Apache Kafka dan Apache Spark.
2. Mengimplementasikan model komputasi analisis sentimen Destinasi Super Prioritas pada media sosial Twitter secara *streaming* dan *real-time* menggunakan Apache Kafka dan Apache Spark.
3. Menganalisis hasil eksperimen sentimen pengguna twitter terhadap Destinasi Super Prioritas secara *streaming* dan *real-time* menggunakan Apache Kafka dan Apache Spark.
4. Menganalisis hasil kinerja model komputasi dalam menganalisis sentimen Destinasi Super Prioritas pada Twitter secara streaming.

1.4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memonitor sentimen publik terkait destinasi super prioritas yang sedang populer di media sosial Twitter.
2. Memberikan *insight* mendalam mengenai destinasi super prioritas yang dapat mendukung pengambilan keputusan yang responsif untuk membantu pemerintah atau pengelola destinasi dalam menyesuaikan strategi promosi, fasilitas, atau layanan.
3. Memberikan penelitian dan kontribusi pada pengembangan teknologi big data, khususnya dalam penggunaan *stream processing* untuk analisis sentimen

1.5. Batasan Masalah

Adapun batasan pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data *tweet* dari media sosial twitter yang berbahasa Indonesia selama tahun 2022.
2. Penelitian ini mengambil studi kasus hanya tempat wisata yang termasuk ke dalam lima Destinasi Super Prioritas.
3. Penelitian ini hanya melakukan simulasi analisis sentimen dari media sosial twitter terhadap lima Destinasi Super Prioritas secara streaming.

1.6. Sistematika Penulisan

Pada dokumen skripsi ini terdapat sistematika yang digunakan sebagai pedoman penulisan. Sistematika penulisan dalam dokumen skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi pendahuluan yang melatarbelakangi penelitian ini, permasalahan yang akan dijawab, tujuan penelitian yang ingin dicapai, dan manfaat yang didapatkan dari penelitian ini, serta batasan masalah yang akan diterapkan dalam penelitian. Bab ini akan menjelaskan pentingnya sektor pariwisata bagi perekonomian Indonesia dan mengapa Destinasi Super Prioritas (DSP) ditetapkan oleh pemerintah. Bab ini juga membahas peran media sosial, terutama Twitter, dalam mencerminkan sentimen publik serta pentingnya analisis sentimen berbasis big data untuk pengelolaan Destinasi Super Prioritas.

BAB II KAJIAN PUSTAKA

Bab ini menguraikan teori-teori dan literatur yang relevan dengan penelitian. Tinjauan pustaka mencakup konsep dasar tentang Pariwisata, Twitter, *Natural Language Processing*, Analisis Sentimen, Big Data, Data Stream, Apache Kafka, dan Apache Spark. Selain itu, bab ini juga membahas studi-studi terdahulu yang berkaitan dengan penggunaan big data dan stream processing dalam sektor pariwisata, serta bagaimana analisis sentimen dapat digunakan untuk memahami opini publik.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian ini, mulai dari desain penelitian, alat dan bahan penelitian yang digunakan. Adapun tahapan desain penelitian terdiri dari tahapan persiapan, studi literatur, data penelitian, tahapan pembentukan model komputasi, tahapan pengembangan sistem, tahapan desain eksperimen, dan yang terakhir adalah melakukan analisis dan hasil.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi penjelasan mengenai proses dari setiap tahap penelitian. Proses tersebut meliputi pengumpulan data dari Twitter, sementara Apache Kafka dan Apache Spark digunakan untuk stream processing dan analisis data secara real-time. Menyajikan hasil penelitian dan analisis data yang telah dilakukan. Hasil penelitian meliputi temuan-temuan utama terkait sentimen pengguna Twitter terhadap Destinasi Super Prioritas, efisiensi dan efektivitas stream processing dengan big data platform, serta pola dan tren yang diidentifikasi dari data. Pembahasan dalam bab ini mengevaluasi hasil-hasil tersebut dalam konteks tujuan penelitian, serta keterlibatannya bagi strategi pemasaran dan peningkatan layanan di DSP.

BAB V KESIMPULAN DAN SARAN

Bab ini menyimpulkan hasil penelitian berdasarkan temuan yang ada selama penelitian dan analisis. Selain itu, bab ini juga memberikan rekomendasi berdasarkan hasil analisis sentimen, termasuk saran untuk potensi penggunaan teknologi big data dan *stream processing* dalam konteks pariwisata lainnya dan saran yang dapat digunakan untuk penelitian selanjutnya.