

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu destinasi pariwisata yang memiliki ketenaran di tingkat nasional maupun internasional adalah Provinsi Bali (Surtiningsih dkk, 2018). Sektor pariwisata di Provinsi Bali mengalami pemulihan setelah dampak pandemi COVID-19 pada tahun 2020 (Afriyadi, 2022). Hal tersebut dibuktikan pada bulan bulan Agustus 2022 dengan tingkat hunian hotel yang mencapai 63,17% di kawasan Nusa Dua (Afriyadi, 2022). Selain itu, pemerintah secara aktif menyelenggarakan acara-acara resmi internasional, seperti Konferensi G20 pada bulan November 2022 sebagai bagian dari upaya promosi pariwisata Bali (Dzulfaroh, 2023). Informasi yang diberikan oleh PT Angkasa Pura I menunjukkan bahwa sejak awal tahun 2022 hingga September 2022, jumlah total turis asing yang datang melalui Bandara Internasional I Gusti Ngurah Rai mencapai 1,19 juta orang (Dzulfaroh, 2023). Meskipun demikian, peningkatan kunjungan turis asing ini telah membawa dampak negatif, termasuk pelanggaran hukum seperti pekerjaan ilegal, pelanggaran lalu lintas, dan tindakan pencurian (Dzulfaroh, 2023). Sehingga peristiwa tersebut dapat menjadi salah faktor yang dapat mempengaruhi sentimen dari citra destinasi dan pariwisata Bali kedepannya.

Analisis sentimen penting dilakukan bagi pemerintah kota dalam rangka memahami evaluasi dan persepsi publik terhadap sektor pariwisata di wilayah tersebut (Fanissa dkk, 2018). Informasi mengenai sentimen ini dapat menjadi dasar penting untuk pengambilan keputusan yang berkaitan dengan perbaikan dan pengembangan sektor pariwisata. (Fanissa dkk, 2018). Selain itu evaluasi menggunakan analisis sentimen merupakan cara yang efektif untuk menggali opini dan pandangan masyarakat terhadap suatu subjek yang dapat diperoleh melalui analisis judul berita serta mengidentifikasi aspek yang perlu perbaikan dan evaluasi karena dampaknya yang signifikan terhadap citra terhadap suatu hal di mata publik atau masyarakat (Rohmatul dkk, 2022). Jurnal - jurnal tentang sentimen analisis terhadap berita sudah dibuat oleh Rohmatul (2022) yang meneliti tentang “Analisis Sentimen Berita Universitas Udayana menggunakan metode *Full Text Search (Natural Language) dan Rule Based Method*”. Selanjutnya penelitian dari Hardeva dan Hendratno (2021) membuat “Analisis Pengaruh Berita Tentang *COVID-19* di Media Massa Terhadap Volatilitas Harga Saham di Sub Sektor Hotel, Restoran & Pariwisata Dengan Menggunakan *RavenPack News Analytic*”. Selain itu penelitian tentang analisis sentimen pada twitter sudah diteliti oleh Mandloi dan Patel (2020) melakukan penelitian tentang “*Twitter Sentiment Analysis using Machine Learning*”. Selanjutnya penelitian dari (Yulita dkk, 2021) membuat Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin *Covid-19* Menggunakan Algoritma *Naïve Bayes Classifier*. Penelitian dari Samuel, J (2020) meneliti tentang “*COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification*” begitu juga

dengan penelitian (Rustam dkk, 2021) meneliti tentang “*A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis*”.

Penelitian-penelitian terkait sentimen analisis sudah banyak dilakukan, contohnya Sentimen Analisis tentang *COVID-19* menggunakan twitter yang diteliti oleh (Mandloi & Patel, 2020), (Yulita dkk, 2021) dan (Samuel, J, 2020), namun penelitian terkait sentimen judul berita tentang pariwisata Bali pada media portal online belum pernah dilakukan, dan sejauh peneliti telusuri tentang penelitian terkait analisis sentimen pariwisata adalah penelitian (Fanissa dkk, 2018) dan (Hardeva & Hendratno, 2021) menggunakan data *Tripadvisor* dan data sekunder, yaitu berupa data saham serta masih sedikit pengembangan menggunakan seleksi fitur atau *feature selection* pada penelitian sentimen analisis tersebut. Selain itu, klasifikasi sentimen teks bermasalah dengan jumlah atribut yang digunakan dalam sebuah dataset (Sinanto dkk, 2022). Pemilihan fitur atau seleksi fitur dapat digunakan untuk mengurangi atribut yang tidak relevan dalam dataset (Sinanto dkk, 2022).S

Berdasarkan permasalahan dan penelitian-penelitian tersebut, pada penelitian ini peneliti ingin membuat dan tertarik untuk mengembangkan penelitian sentimen analisis menggunakan *feature selection* dengan judul “Perbandingan Analisis Sentimen menggunakan *Feature Selection Chi Square* dan *Information Gain* pada Judul Berita mengenai Pariwisata Bali pada Portal Media Online” secara bertahap yang didukung oleh penelitian (Septiana dkk, 2021) yang meneliti tentang “Analisis Sentimen Vaksinasi Covid-19 Pada *Twitter* Menggunakan *Naive Bayes Classifier* Dengan *Feature Selection Chi-Squared Statistic* Dan *Particle Swarm Optimization*. Penelitian

tersebut mengimplementasikan *feature selection* pada sentimen analisis untuk meningkatkan kerja prediksi, mengurangi waktu komputasi, dan untuk pemahaman yang lebih baik tentang data dalam *machine learning* atau *pattern recognition applications* (Chandrashekar & Sahin, 2014). Pada penelitian ini ingin membuat sentimen analisis judul berita tentang pariwisata bali pada media portal berita online menggunakan *feature selection Chi-Square* dan *Information Gain* dengan membandingkan kedua *feature selection* tersebut dan melihat model klasifikasi yang dihasilkan setelah seleksi fitur apakah dapat meningkatkan akurasi klasifikasi dan menghasilkan model klasifikasi yang baik atau tidak.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana cara implementasi *feature selection Chi-Square* dan *Information Gain* pada Analisis Sentimen mengenai judul berita Pariwisata Bali?
2. Bagaimana hasil perbandingan akurasi model klasifikasi menggunakan *feature selection Chi-Square* dan *Information Gain*?

1.3 Tujuan

Berdasarkan dari rumusan masalah, didapatkan tujuan dari penelitian ini sebagai berikut:

1. Untuk mengetahui bagaimana cara implementasi *feature selection Chi-Square* dan *Information Gain* pada Analisis Sentimen judul berita pariwisata Bali menggunakan SVM.
2. Untuk mengetahui hasil perbandingan akurasi model klasifikasi SVM dengan *feature selection Chi-Square* dan *Information Gain*.

1.4 Batasan Masalah

Batasan masalah dari penelitian ini mencakup hal-hal berikut:

1. Bahasa pemrograman yang digunakan adalah *python* beserta *library* menggunakan IDE yaitu *Google Colaboratory*.
2. Dataset yang digunakan adalah data judul berita berbahasa Indonesia.
3. Sumber media portal online yang digunakan adalah Detik.com, Kompas.com, dan Balipost.com.
4. Data judul berita yang diambil judul berita dengan *tag* dan *query* yang berkaitan dengan pariwisata bali dari bulan Januari tahun 2022 hingga bulan Maret tahun 2023.
5. Pemberian label pada data dilaksanakan oleh mahasiswa yang memiliki keahlian atau kompetisi dalam bahasa Indonesia sebagai pembantu peneliti.
6. Algoritma perbandingan seleksi fitur yang digunakan adalah *Chi-Square* dan *Information Gain*.

1.5 Manfaat

Manfaat dari penelitian ini dapat dikategorikan ke dalam dua jenis, yakni manfaat teoritis dan manfaat praktis, yang dapat diuraikan sebagai berikut:

1. Manfaat Teoritis

Studi ini bisa menjadi referensi bagi peneliti lain yang akan menjalankan penelitian serupa di masa yang akan datang.

2. Manfaat secara praktis

a. Bagi mahasiswa

Manfaat bagi peneliti dari penelitian skripsi ini adalah memenuhi persyaratan akademik untuk menyelesaikan tugas akhir dan memberikan panduan kepada mahasiswa lain yang melakukan penelitian serupa tentang Perbandingan Analisis Sentimen dengan menggunakan SVM pada Judul Berita Pariwisata Bali.

b. Bagi Pemerintah

Manfaat bagi pemerintah adalah pengklasifikasian berita terkait pariwisata di Bali yang dapat memberikan pandangan yang lebih jelas tentang pandangan publik, baik yang bersifat negatif maupun positif. Informasi ini dapat digunakan sebagai acuan atau panduan bagi *stakeholder* terkait dalam merumuskan fokus pembenahan dan evaluasi dalam sektor pariwisata Bali berdasarkan judul berita yang cenderung memiliki sentimen negatif.