

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Sektor pariwisata adalah sektor yang memiliki potensi besar dalam menciptakan lapangan kerja, salah satu contohnya ialah industri perhotelan. Kemudahan akses internet memungkinkan calon wisatawan untuk mencari informasi mendalam tentang hotel, termasuk ulasan dari pengunjung sebelumnya. Ulasan ini berperan krusial dalam membentuk persepsi calon tamu terhadap kualitas fasilitas dan pelayanan hotel. Penelitian Chen et al. (2020) menegaskan bahwa ulasan online sangat memengaruhi keputusan pembelian konsumen, termasuk dalam memilih akomodasi.

Shangri-La Surabaya, bersama hotel-hotel besar lainnya seperti JW Marriott dan Sheraton, berkompetisi untuk mendapatkan ulasan positif dari tamu. Hotel bintang lima ini sangat bergantung pada umpan balik pelanggan untuk terus meningkatkan kualitas layanan dan fasilitasnya. Sebagai salah satu hotel bintang lima termahal di Surabaya, Shangri-La menawarkan tarif menginap yang cukup tinggi, berkisar antara 2 hingga 6 juta rupiah per malam. Karena harga yang premium ini, calon tamu tentu memiliki ekspektasi yang tinggi terhadap layanan dan fasilitas yang akan mereka dapatkan. Dengan harga kamar yang terbilang mewah, Shangri-La Surabaya menjadi pilihan bagi mereka yang mencari pengalaman menginap yang istimewa. Untuk itu, calon tamu seringkali mencari informasi lebih lanjut melalui ulasan di *Google Maps*, guna memastikan bahwa layanan dan fasilitas yang ditawarkan sesuai dengan harga yang mereka bayarkan. *Google Maps* dipilih tidak hanya untuk melihat arah, lokasi, dan rute terbaik menuju tempat yang ingin dikunjungi, tetapi juga karena pengguna dapat melihat rating serta komentar atau ulasan. Ulasan ini sering digunakan sebagai pertimbangan bagi pengguna dalam memutuskan apakah akan mengunjungi tempat tersebut atau tidak. Ulasan di *Google Maps* diberikan dengan rating dari 1 hingga 5. Namun, tidak jarang terjadi ketidaksesuaian antara rating dan ulasan yang diberikan pengguna. Sebagai contoh, Reza memberikan rating 5 tetapi dalam ulasannya reza mengeluhkan banyak sekali kekurangan dari tempat tujuannya. Oleh karena itu dapat diartikan bahwa rating saja tidak cukup untuk menilai suatu

tempat. Ulasan dalam bentuk kalimat dianggap lebih akurat dalam menilai suatu tempat yang sudah dirasakan sendiri oleh pengunjung sebelumnya. Tanpa adanya komentar atau ulasan tersebut, calon pengunjung akan kesulitan mengetahui faktor mana yang sudah baik dan faktor mana yang perlu perbaikan.

Untuk memahami hal apa saja yang mempengaruhi keputusan konsumen dalam memilih hotel, dilakukan penelitian analisis sentimen terhadap ulasan pengguna *Google Maps*. Tujuannya adalah untuk mengidentifikasi aspek-aspek apa saja yang paling menarik minat calon tamu. Analisis sentimen merupakan suatu proses pengolahan bahasa alami yang bertujuan untuk secara otomatis mengidentifikasi, mengekstrak, mengukur, dan mempelajari opini, sentimen, dan emosi yang tersirat dalam sebuah teks. Terdapat beragam algoritma yang dapat dimanfaatkan untuk melakukan klasifikasi sentimen, mulai dari metode yang sederhana seperti *Naïve Bayes* hingga metode yang lebih kompleks seperti SVM, *Logistic Regression* dan juga *XGBoost*.

Analisis sentimen telah menjadi subjek penelitian di berbagai studi sebelumnya. Dalam penelitiannya tahun 2021, Ferian Fauzi Abdullah mengkaji sentimen masyarakat terkait vaksin COVID-19 dengan menggunakan algoritma SVM. Data yang digunakan berupa 2634 komentar *YouTube* yang telah diproses menggunakan teknik TF-IDF untuk mengubah teks menjadi format yang dapat diproses oleh komputer. Hasil penelitian menunjukkan bahwa SVM mampu mengklasifikasikan sentimen dengan akurasi sebesar 77,75% (Ferian Fauzi.,2021). Penelitian yang dilakukan oleh Elza Fitriana dan timnya membuktikan bahwa metode *Recursive Feature Elimination* (RFE) sangat efektif dalam meningkatkan akurasi klasifikasi sentimen. Dalam kasus analisis sentimen *e-wallet* di *Twitter*, penggunaan RFE berhasil meningkatkan akurasi algoritma SVM dari 74% menjadi 81%. (Elza Fitriana et al., 2021). Di sebuah penelitian analisis sentimen PSBB di DKI Jakarta dan dampak terhadap IHSG, metode *Logistic Regression* terbukti paling efektif. Metode ini berhasil mencapai akurasi sebesar 75% dalam mengklasifikasikan sentimen, diikuti oleh random forest (60%), KNN (70%), dan Naive Bayes (70%) (Amelia Syahadati et al., 2021). Dalam penelitian tugas akhir ini, peneliti akan menganalisis 3.719 data yang diambil dari kolom ulasan di *Google Maps* Hotel Shangri-La Surabaya. Sentimen

dalam ulasan-ulasan tersebut akan dikategorikan menjadi dua, yaitu sentimen positif dan negatif. Penelitian ini akan memanfaatkan algoritma SVM dan *Logistic Regression* untuk melakukan klasifikasi. Sebelumnya, data teks akan diubah menjadi bentuk numerik menggunakan teknik *Word Embedding* TF-IDF. *Word Embedding* ini akan merepresentasikan setiap kata dalam bentuk vektor numerik yang dapat dipahami oleh mesin (Muhammad Arief Rahman et al., 2019).

*Support Vector Machine* (SVM) merupakan algoritma pembelajaran terawasi yang memanfaatkan pemetaan non-linear untuk mengklasifikasikan data. Algoritma ini bekerja dengan mencari batas optimal yang memisahkan data menjadi beberapa kelas dalam ruang dimensi yang lebih tinggi. Dengan demikian, data dari kedua kategori dapat dipisahkan secara efektif menggunakan *hyperplane*, yang bertindak sebagai batas keputusan antara kedua kelas. (Laia & Setyawan, 2020). Salah satu keunggulan algoritma ini adalah kemampuannya menghasilkan akurasi klasifikasi yang tinggi, meskipun dilatih dengan dataset yang terbatas. Sementara itu, *Logistic Regression* merupakan suatu metode dalam statistika yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dengan satu variabel dependen biner (yang hanya memiliki dua kategori). Dengan menggunakan fungsi sigmoid, *Logistic Regression* menghasilkan nilai probabilitas yang berkisar antara 0 dan 1, yang menunjukkan kemungkinan terjadinya suatu peristiwa. (Pangaribuan, 2021). dan RFE adalah algoritma yang secara iteratif menghilangkan fitur-fitur yang dianggap kurang penting berdasarkan bobot atau koefisien yang diberikan oleh model, sehingga menghasilkan subset fitur yang paling relevan untuk model tersebut.

Setelah melakukan kajian mendalam terhadap literatur terkait, peneliti mengusulkan judul penelitian, “Pengaruh RFE terhadap *Logistic Regression* Dan *Support Vector Machine* Pada Analisis Sentimen Hotel Shangri-La Surabaya”

## **1.2 Rumusan Masalah**

Berdasarkan kajian terhadap latar belakang permasalahan, maka dapat diidentifikasi beberapa rumusan masalah penelitian sebagai berikut:

1. Bagaimana cara mengimplementasikan seleksi fitur *Recursive Feature Elimination* untuk meningkatkan akurasi algoritma *Logistic Regression* dan *Support Vector Machine* untuk menganalisis sentimen pada kolom

ulasan *Google Maps* Hotel Shangri-La Surabaya ?

2. Bagaimana tingkat akurasi pengguna seleksi fitur *Recursive Feature Elimination* pada algoritma *Logistic Regression* dan *Support Vector Machine* untuk menganalisis sentimen pada kolom ulasan *Google Maps* Hotel Shangri-La Surabaya ?
3. Apakah proporsi pembagian data latih dan data uji berpengaruh terhadap nilai akurasi dalam penggunaan seleksi fitur *Recursive Feature Elimination* pada algoritma *Logistic Regression* dan *Support Vector Machine* untuk menganalisis sentimen pada ulasan *Google Maps* untuk Hotel Shangri-La Surabaya?

### **1.3 Tujuan**

Sesuai dengan rumusan masalah yang diajukan, penelitian ini bertujuan untuk :

1. Mengetahui pengaruh dari seleksi fitur *Recursive Feature Elimination* pada algoritma *Logistic Regression* dan *Support Vector Machine* untuk menganalisis sentimen pada kolom ulasan *Google Maps* hotel Shangri-La Surabaya
2. Mengetahui tingkat akurasi penggunaan seleksi fitur *Recursive Feature Elimination* pada algoritma *Logistic Regression* dan *Support Vector Machine* untuk menganalisis sentimen pada kolom ulasan *Google Maps* hotel Shangri-La Surabaya
3. Menganalisis bagaimana proporsi pembagian data latih dan data uji memengaruhi nilai akurasi dalam penerapan seleksi fitur *Recursive Feature Elimination* pada algoritma *Logistic Regression* dan *Support Vector Machine* untuk analisis sentimen pada ulasan *Google Maps* Hotel Shangri-La Surabaya.

### **1.4 Manfaat**

Adapun manfaat yang dapat diambil dari penelitian ini adalah :

1. Penelitian ini ingin menunjukkan bagaimana cara mengungkap penerapan dari seleksi fitur *Recursive Feature Elimination* pada algoritma *Support Vector Machine*.

2. Penelitian ini ingin menunjukkan bagaimana cara mengungkap penerapan dari seleksi fitur *Recursive Feature Elimination* pada algoritma *Logistic Regression*.
3. Sebagai acuan untuk memahami penilaian pelanggan terhadap layanan hotel dengan tujuan untuk meningkatkan kualitas pelayanan hotel.

### **1.5 Batasan Masalah**

Berikut ini adalah batasan masalah yang kami hadapi dalam penelitian ini:

1. Sumber data utama dalam penelitian ini adalah 3719 ulasan pelanggan yang diambil secara langsung dari kolom ulasan *Google Maps* untuk Hotel Shangri-La Surabaya.
2. Algoritma ini diimplementasikan dalam bentuk kode *Python* untuk memudahkan pengembangan dan pengujian.
3. Data ulasan dipilih secara berurutan berdasarkan tingkat relevansinya terhadap tujuan penelitian.
4. Dataset yang digunakan dalam penelitian ini diperoleh melalui proses *scraping* pada tanggal 5 Februari 2024.
5. Data sentimen ulasan akan diklasifikasikan menjadi dua kelas, yaitu kelas positif dan kelas negatif.
6. Pada RFE menggunakan *library*.
7. Penelitian hanya terbatas pada pengujian laboratorium.

*Halaman ini sengaja dikosongkan*