

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Revolusi industri generasi ke empat telah membawa banyak perubahan pada berbagai sektor termasuk pada industri perhotelan yang memiliki urgensi untuk menghadapi dan bertahan dalam persaingan yang terus meningkat. Salah satu faktor keberhasilan dalam menghadapi persaingan tersebut yaitu dengan memperhatikan kebutuhan pelanggan melalui evaluasi produk dan jasa secara berkelanjutan. Evaluasi tersebut dapat dilakukan dengan metode konvensional melalui wawancara, survei dan kuisioner. Hal tersebut berakibat pada hasil analisis menjadi tidak akurat dan tidak konsisten (Guo dkk., 2017).

Dampak yang diakibatkan oleh evaluasi menggunakan metode konvensional dapat dikurangi dengan beralih ke pemanfaatan teknologi *web 2.0*, sebagai contoh situs TripAdvisor yang merupakan tempat untuk berbagi pengalaman melalui ulasan *online*. Ulasan tersebut berupa deskripsi teks yang direpresentasikan dalam bentuk skala peringkat 1 sampai 5 yang bertujuan untuk memberikan informasi dan memudahkan pengguna dalam menilai suatu ulasan. Tetapi skala peringkat tersebut tidak bisa dijadikan sebagai acuan dalam menentukan nilai suatu ulasan karena seringkali tidak akurat dan tidak sesuai dengan isi ulasan (Tama dkk., 2019) sehingga diperlukan metode komputasi yang dapat melakukan proses analisis data ulasan tersebut secara efisien menggunakan teknik *text mining* (Yildiz, 2019). *Text mining* merupakan salah satu bagian dari *data mining*, memiliki definisi proses mengeksplorasi data berbentuk teks untuk dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya sehingga dapat menghasilkan informasi yang berguna, seperti analisis sentimen ulasan pelanggan di media *online* (Kotu dan Deshpande, 2015).

Analisis sentimen adalah proses yang bertujuan untuk memenentukan isi dari dataset yang berbentuk teks (dokumen, kalimat dan paragraf) bersifat positif, negatif atau netral. Analisis sentimen dapat memberikan keuntungan dalam pengambilan keputusan perusahaan untuk berbagai bidang, termasuk sektor

pariwisata (Kontopoulos dkk., 2013). Ulasan pelanggan di media *online* saat ini telah menjadi salah satu sumber informasi yang penting dalam berbagai produk dan layanan (Liu dkk., 2012), karena dapat mengurangi dan membantu konsumen untuk menyimpulkan kualitas suatu produk atau layanan (Koh dkk., 2010).

Analisis sentimen dapat dilakukan dengan menggunakan beberapa algoritma klasifikasi seperti *Support Vector Machine* (SVM), *Artificial Neural Network* (ANN) dan *K-Nearest Neighbor* (KNN). Algoritma ANN memiliki kelemahan harus menggunakan data *training* yang besar. Algoritma SVM mempunyai kelebihan yaitu bisa diterapkan untuk data yang berdimensi tinggi, akan tetapi SVM sulit digunakan untuk data dengan jumlah yang besar (Chandani dkk., 2015). KNN merupakan algoritma yang bekerja dengan menghitung jarak terdekat antar atribut data, memiliki keunggulan dalam hal komputasi berkinerja tinggi (Khamis dkk., 2014), algoritma yang sederhana dan tahan terhadap berbagai karakteristik data yang besar (Imandoust dan Bolandraftar, 2013) dan memiliki akurasi yang baik jika dibandingkan dengan algoritma lainnya (Khamis dkk., 2014) serta memiliki kompleksitas algoritma yang relatif kecil (Yong dkk., 2009).

Penelitian terdahulu yang menggunakan metode KNN seperti klasifikasi teks dalam bahasa Inggris (Bilal dkk., 2015), analisis ulasan pengguna hotel (Khorsand dkk., 2020), klasifikasi sentimen komentar di *twitter* dalam Bahasa Inggris (Kaur dkk., 2018), analisis sentimen terhadap ulasan produk di *online shop* (Suganya dan Vijayan., 2020) dan klasifikasi teks berita pada laman *bbc news* dengan menggunakan pembobotan TF-IDF (Shah dkk., 2020). Namun masalah yang sering muncul dalam proses klasifikasi adalah *imbalanced* dataset, yaitu kondisi salah satu kelasnya mempunyai jumlah data yang jauh lebih banyak pada dataset yang dipakai. Kondisi tersebut membuat hasil klasifikasi akan lebih condong ke bagian data yang memiliki komposisi data lebih besar sehingga menjadi tidak akurat (Ganganwar, 2012).

Tahapan awal yang dilakukan dalam klasifikasi data berbentuk teks adalah *term weighting* (pembobotan), bertujuan untuk mengubah data berbentuk teks menjadi sekumpulan angka (vektor). Teknik pembobotan tersebut dapat dilakukan dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-

IDF) yang mampu menghasilkan pembobotan dengan baik dan waktu komputasi yang cepat (Erra dkk., 2015).

Berdasarkan pada permasalahan dan keunggulan metode pada penelitian sebelumnya, maka penelitian ini menerapkan metode pembobotan TF-IDF dan algoritma klasifikasi KNN untuk analisis sentimen ulasan pada media *online*. Usulan dalam penelitian ini adalah penggunaan metode *random over-sampling* untuk *rebalance* data sehingga diharapkan kinerja dari algoritma klasifikasi lebih optimal, menghasilkan informasi yang efektif dan efisien. Studi kasus yang digunakan dalam penelitian ini adalah ulasan pengunjung hotel kota Mussoorie India di situs TripAdvisor, data yang digunakan sebanyak 1314 data teks dalam bahasa Inggris.

### **1.2. Tujuan Penelitian**

Tujuan dari penelitian ini adalah menerapkan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *K-Nearest Neighbor* yang dirangkum dalam satu sistem yang digunakan untuk analisis ulasan pengguna hotel pada laman TripAdvisor. Hasil keluaran sistem adalah analisis klasifikasi sentimen pengguna hotel dalam tampilan grafik dan tabel.

### **1.3. Manfaat Penelitian**

Manfaat dari penelitian ini adalah implementasi algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *K-Nearest Neighbor* yang dirangkum dalam satu sistem yang digunakan untuk analisis ulasan pengguna hotel pada laman TripAdvisor sehingga dapat menghasilkan informasi berupa klasifikasi sentimen secara efektif dan efisien.