

BAB I

PENDAHULUAN

1.1 Latar Belakang

Objek wisata adalah salah satu daya tarik yang perlu dikembangkan di setiap daerah, karena pengembangan wisata saat ini diharapkan dapat memperluas kesempatan untuk memberikan lapangan pekerjaan dan usaha bagi masyarakat sekitar yang iliranya akan dapat menunjang pembangunan [1]. Kabupaten Bangkalan adalah salah satu kabupaten di Madura yang memiliki sektor wisata yang berpotensi untuk dikembangkan [2]. Pemerintah Kabupaten Bangkalan memiliki kebijakan pada sektor pariwisata yang ditekankan pada penataan objek wisata yang diarahkan untuk menarik wisatawan dengan meningkatkan objek-objek wisata yang ada pada daerah. Dengan kebijakan tersebut maka diperlukan pengembangan pada objek wisata yang ada [3]. Peningkatan jumlah wisatawan menjadi salah satu indikator kesuksesan pariwisata dalam suatu daerah. Namun, pihak pengelola pariwisata perlu memiliki wawasan mengenai faktor yang mempengaruhi kepuasan pengunjung wisata.

Perkembangan dan kemajuan teknologi informasi membawa banyak keuntungan dan manfaat bagi manusia. Perkembangan teknologi informasi juga berbanding lurus dengan perkembangan pariwisata. Saat ini pengunjung wisata dapat memberikan komentar terhadap wisata yang telah dikunjungi. Salah satu contohnya ialah peta daring yang memiliki fitur untuk merencanakan rute serta petunjuk arah untuk mencari sebuah alamat atau yang dikenal dengan *Goole Maps*. Pada tahun 2016 *Google* menambahkan fitur yang memungkinkan pengguna untuk memberikan rating dan ulasan pada tempat yang dikunjungi [4]. Dalam aplikasi *Google Maps* banyak pengunjung objek wisata di Kabupaten Bangkalan yang sudah memberikan ulasan mereka pada objek wisata yang telah dikunjungi. Hal ini dapat membantu untuk mengetahui tanggapan positif, netral atau negatif dari pengunjung yang nantinya akan berguna untuk mengetahui hasil kepuasan pengunjung yang diharapkan mempermudah pengelola wisata di Kabupaten Bangkalan dalam menentukan langkah yang harus dilakukan dalam mengembangkan objek wisata yang ada. Tetapi dengan banyaknya jumlah ulasan yang diberikan oleh pengunjung akan menyulitkan pengelola tempat wisata untuk mengetahui tanggapan yang diberikan. Untuk itu dilakukan analisis sentimen untuk mengetahui tanggapan pengunjung tempat wisata.

Analisis Sentimen merupakan salah satu bidang *Natural Language Processing (NLP)* yang menggunakan *text analytics* untuk mengenali dan mengekstraksi opini dalam bentuk teks. Analisis sentimen digunakan untuk mengelola informasi yang sebelumnya tidak terstruktur

untuk diubah menjadi informasi yang lebih terstruktur, yang digunakan untuk menilai suatu objek apakah cenderung positif, negatif atau netral [5]. Metode untuk melakukan analisis sentimen sebuah topik bisa dengan berbagai macam. Salah satunya adalah *Machine learning* yang dipilih untuk menjadi metode dalam penelitian ini.

Machine learning merupakan teknologi yang memungkinkan mesin untuk mempelajari pola-pola dari data dan melakukan tugas-tugas tertentu secara mandiri. Dalam analisis sentimen *machine learning* digunakan untuk mengenali pola-pola sentimen atau emosi dalam teks, sehingga dapat mengklasifikasikan teks berdasarkan sentimen positif, negatif, atau netral. Salah satu algoritma yang termasuk dalam kategori *machine learning* adalah *Naïve bayes* yang digunakan dalam klasifikasi teks. *Naïve Bayes* merupakan salah satu algoritma *machine learning* yang digunakan untuk klasifikasi teks dengan konsep yang mudah dimengerti dan memberikan hasil akurat dengan efisiensi tinggi [6].

Untuk melakukan penelitian analisis sentimen terhadap kepuasan pengunjung wisata di Kabupaten Bangkalan, penelitian ini menggunakan metode *Naïve Bayes* dengan pembobotan kata TF-IDF dan seleksi fitur *Information Gain*. Penggunaan metode *Naïve bayes* terbukti memiliki hasil yang sangat baik dalam banyak penelitian yang dilakukan dengan tingkat akurasi data yang besar. Data yang berupa ulasan pengunjung wisata di Kabupaten Bangkalan akan diolah menggunakan *Naïve Bayes* dengan cara menghitung probabilitasnya. Tingkat akurasi *naïve bayes* yang tinggi adalah salah satu kelebihan *naïve bayes* apabila dibandingkan dengan algoritma *machine learning* lainnya. Hal ini menjadikan algoritma *naïve bayes* cukup baik untuk digunakan pada data yang memiliki jumlah besar [7].

Dalam implementasinya *Naïve bayes* dinilai masih memiliki beberapa kekurangan salah satunya adalah sensitif terhadap pemilihan fitur. Jumlah fitur yang sangat banyak dapat mengurangi kinerja klasifikasi. Oleh karena itu diperlukan sebuah proses pemilihan fitur yang dapat membuat proses klasifikasi lebih efisien dengan cara mengurangi jumlah data yang akan dianalisis. Dalam penelitian ini, digunakan metode seleksi fitur *Information gain*, yang sebelumnya telah digunakan dalam penelitian mengenai penambahan metode Seleksi fitur *Information gain* untuk analisis sentimen terhadap ulasan aplikasi *flip* dan menghasilkan nilai akurasi yang lebih tinggi jika menggunakan seleksi fitur. Hasil dari penelitian ini menunjukkan bahwa *Information gain* memberikan hasil terbaik dengan tingkat akurasi mencapai 96.26% [8].

Pada penelitian ini jumlah kelas pada data ulasan wisata memiliki jumlah yang tidak seimbang. Pada *dataset* yang seimbang, proses klasifikasi dapat langsung dilakukan tanpa melalui tahapan tertentu, namun berbeda halnya ketika kelas *dataset* ini tidak seimbang perlu

dilakukan teknik sampling untuk menangani *dataset* yang tidak seimbang supaya tidak menurunkan kinerja model klasifikasi. Dalam menangani ketidak seimbangan kelas *dataset*, dapat menggunakan teknik *over-sampling* dengan *SMOTE (Synthetic Minority Oversampling Technique)* [9].

Penelitian yang dilakukan untuk melakukan perbandingan hasil akurasi algoritma *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes* dan *Gaussian Naïve Bayes* dalam melakukan analisis sentimen hacker bjorka pada media sosial twitter. Dengan pembobotan kata TF-IDF menggunakan 1000 data. Perbandingan algoritma *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes* dan *Gaussian Naïve Bayes* menghasilkan nilai akurasi 73%, presisi 73%, dan *recall* 100% pada algoritma *multinomial NB*, akurasi 72%, presisi 73%, dan *recall* 98% pada *Bernoulli Naïve Bayes*, dan akurasi 55%, presisi 75%, dan *recall* 63% pada *Gaussian Naïve Bayes* [10].

Menurut penelitian yang menggunakan *Naïve Bayes* dan *Information Gain* untuk melakukan analisis sentimen terhadap penggunaan *E-Wallet* pada saat pandemi dengan sumber data komentar di *twitter*, memperoleh hasil yang cukup tinggi setelah menambahkan seleksi fitur *Information Gain* yaitu 92% hasil akurasi tersebut lebih tinggi jika dibandingkan tanpa menggunakan *Information Gain* yaitu sebesar 84% [11]. Dengan ini dapat disimpulkan bahwa *Naïve bayes* dan seleksi fitur *Information Gain* efektif untuk di implementasi pada analisis sentimen.

Analisis sentimen dengan menggunakan *Naïve bayes* tentu telah banyak dilakukan oleh peneliti terdahulu, tetapi belum ada yang berfokus pada wisata khususnya di Kabupaten Bangkalan. Oleh karena itu melalui penelitian ini, akan dilakukan analisis sentimen terhadap ulasan pengunjung wisata di Kabupaten Bangkalan pada *Google Maps* dengan menggunakan algoritma *Naïve bayes*. Dari penelitian ini di harapkan dapat mengetahui pengaruh *SMOTE* dan *Information gain* pada kinerja algoritma *naïve bayes* dalam melakukan analisis sentimen kepuasan pengunjung wisata di Kabupaten Bangkalan melalui ulasan yang diberikan pada *Google Maps*.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Tempat wisata yang dijadikan tujuan destinasi wisatawan akan sangat berpengaruh terhadap banyak tidaknya wisatawan yang datang ke tempat tersebut. Dalam aplikasi *Google Maps* banyak pengunjung objek wisata di Kabupaten Bangkalan yang sudah memberikan ulasan mereka pada objek wisata yang telah dikunjungi. Tetapi dengan banyaknya jumlah

ulasan yang diberikan oleh pengunjung akan menyulitkan untuk mengetahui tanggapan yang diberikan. Dengan adanya permasalahan tersebut, maka dilakukan analisis sentimen untuk mengetahui tanggapan positif, netral, atau negatif yang diberikan oleh pengunjung wisata dengan menggunakan *Naïve Bayes*.

Ketika jumlah fitur yang digunakan memiliki jumlah yang besar dan banyak, *Naïve bayes* perlu menghitung probabilitas kemunculan setiap fitur dalam setiap kelas, yang dapat menyebabkan waktu komputasi lebih lama, untuk itu diperlukan seleksi fitur untuk mengatasi permasalahan tersebut. Metode yang digunakan dalam seleksi fitur adalah seleksi fitur *Information Gain*. Data yang digunakan dalam penelitian ini memiliki ketidakseimbangan jumlah kelas, sehingga diperlukan penyeimbangan untuk mengatasi permasalahan tersebut. Teknik yang digunakan untuk penyeimbangan kelas adalah *SMOTE (Synthetic Minority Over-sampling Technique)*.

1.2.2 Metode Usulan

Penelitian ini mengangkat topik analisis sentimen pada ulasan wisata di Kabupaten Bangkalan dengan menggunakan Algoritma *Naïve Bayes*. Analisis sentimen pada penelitian ini dilakukan menggunakan algoritma *Naïve bayes* dan seleksi fitur *Information Gain*. Penelitian ini akan menggunakan data Bahasa Indonesia yang didapat dari ulasan wisata di Kabupaten Bangkalan pada *Google Maps* dengan teknik *web scraping*. Data tersebut dibagi menjadi tiga kelas yaitu positif, netral, dan negatif. Data tersebut nantinya akan dilakukan pelabelan, preprocessing, pembobotan kata, penanganan kelas yang tidak seimbang dengan teknik *SMOTE*, seleksi fitur *Information Gain*, hingga pemodelan dan menghasilkan nilai akurasi, presisi, *recall* yang didapat untuk membuat sebuah kesimpulan. Selanjutnya dilakukan visualisasi kata menggunakan *wordcloud* untuk mengetahui faktor apa saja yang mempengaruhi kepuasan pengunjung wisata pada setiap sentimen.

1.3 Pertanyaan Penelitian

Berdasarkan hasil uraian permasalahan, maka pertanyaan yang menjadi fokus dari penelitian ini yaitu:

1. Bagaimana tingkat performa (akurasi, presisi, *recall*, dan *f1-score*) algoritma *Naïve Bayes* dalam melakukan analisis sentimen pada *dataset* ulasan pengunjung wisata di Kabupaten Bangkalan?

2. Bagaimana tingkat performa (akurasi, presisi, *recall* dan *f1-score*) algoritma *Naïve Bayes* dengan seleksi fitur *Information gain* dalam melakukan analisis sentimen pada *dataset* ulasan pengunjung wisata di Kabupaten Bangkalan?
3. Bagaimana pengaruh teknik SMOTE (*Synthetic Minority Over-sampling Technique*) dalam kinerja klasifikasi analisis sentimen pada *dataset* ulasan pengunjung wisata di Kabupaten Bangkalan?

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan dilakukan penelitian ini adalah:

1. Mengetahui tingkat performa (akurasi, presisi, dan *recall* dan *f1-score*) algoritma *Naïve Bayes* dalam melakukan klasifikasi teks ulasan ulasan pengunjung wisata di Kabupaten Bangkalan
2. Mengetahui tingkat performa (akurasi, presisi, dan *recall* dan *f1-score*) algoritma *Naïve Bayes* dengan seleksi fitur *Information gain* dalam melakukan klasifikasi teks ulasan pengunjung wisata di Kabupaten Bangkalan
3. Mengetahui pengaruh teknik SMOTE (*Synthetic Minority Over-sampling Technique*) dalam kinerja klasifikasi analisis sentimen pada *dataset* ulasan pengunjung wisata di Kabupaten Bangkalan

1.4.2 Manfaat

Adapun manfaat dari penelitian ini adalah untuk melakukan analisis sentimen terhadap ulasan pengunjung wisata di Kabupaten Bangkalan. Analisis sentimen ini digunakan untuk mengetahui performa algoritma *Naïve Bayes*, seleksi fitur *Information gain* dan Teknik SMOTE dalam melakukan klasifikasi data ulasan pengunjung wisata Kabupaten Bangkalan pada *Google Maps*, sehingga dapat mengetahui tanggapan positif, netral, atau negatif serta topik yang paling sering dibahas untuk mengetahui kepuasan pengunjung wisata.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

- a. Ulasan yang digunakan berasal dari *Google Maps*.
- b. Ulasan yang digunakan merupakan ulasan dalam kurun waktu 3 tahun terakhir.
- c. Ulasan yang digunakan adalah ulasan pengunjung wisata pada Bukit Pelalangan Arosbaya, Bukit Geger, Bukit Jaddih, Pantai Siring Kemuning, Labuhan Mangrove Education Park, Pantai Rongkang, Mercusuar Sambilangan, Makam Syaikhona Kholil, Bukit Lampion Beramah, dan Sumber Pocong di Kabupaten Bangkalan.

d. Sentimen yang digunakan adalah sentimen positif, negatif, dan netral.

1.6 Sistematika Penulisan

Sistematika penulisan dibuat agar pembaca dapat lebih mudah memahami isi laporan penelitian. Secara umum, sistematika penulisan laporan ini terdiri dari:

BAB I

PENDAHULUAN

Pada bab ini berisi tentang latar belakang penelitian analisis sentimen ulasan wisata di Kabupaten Bangkalan, rumusan masalah, dan pertanyaan penelitian dari obyek yang diteliti, tujuan dan manfaat penelitian analisis sentimen menggunakan algoritma *Naïve Bayes*, batasan masalah penelitian, dan sistematika penulisan

BAB II

TINJAUAN PUSTAKA

Bab ini berisi tinjauan literatur yang relevan dengan penelitian yang dilakukan sebelumnya, dan penelitian terdahulu dijadikan peneliti sebagai acuan pada penelitian ini. Pada bab ini membahas tentang teori-teori yang digunakan dalam penelitian mengenai *Natural Language Processing*, analisis sentimen, algoritma *naïve bayes*, dan teori yang berhubungan dengan analisis sentimen.

BAB III

METODE USULAN

Bab ini berisi berisi tahapan proses pengerjaan penelitian berupa pengumpulan data yang didapat melalui *web scraping* ulasan pada *Google Maps*, labelisasi data, *pre-processing*, pembobotan kata, penyeimbangan data, pembagian data, seleksi fitur, pemodelan klasifikasi teks dengan menggunakan algoritma *naïve bayes*, evaluasi, dan visualisasi kata menggunakan *word cloud*.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini berisi hasil pengolahan data, pengujian dari hasil proses klasifikasi menggunakan metode *Naïve Bayes*

menggunakan seleksi fitur *Information Gain* pada *dataset* ulasan wisata di Kabupaten Bangkalan.

BAB V

PENUTUP

Bab ini berisi rangkuman kesimpulan yang diperoleh dari hasil penelitian serta proses yang telah dilakukan. Selain itu, bab ini juga akan mencakup saran-saran yang akan disampaikan untuk keperluan pengembangan penelitian di masa mendatang.

DAFTAR PUSTAKA

Daftar pustaka berisi daftar penelitian terdahulu yang dijadikan rujukan dalam penelitian ini.