

BAB I

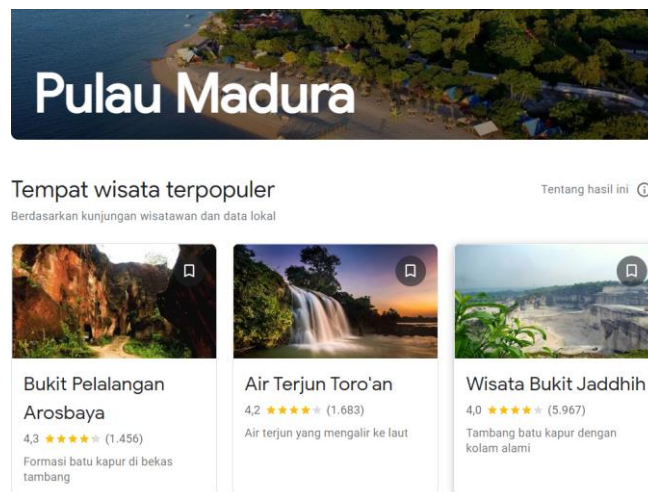
PENDAHULUAN

1.1 Latar Belakang

Teknologi dan informasi pada sektor pariwisata mempunyai peranan yang tidak kalah penting sehingga dapat memberikan efek yang positif dengan menawarkan kemudahan akses informasi dan layanan bagi para wisatawan [1].

Layanan bagi wisatawan dapat ditingkatkan dan dikembangkan secara berkelanjutan oleh petugas atau penyelenggara tempat wisata. Peningkatan ataupun pengembangan tersebut dapat dilakukan dengan melihat dari adanya masukan *feedback* atau umpan balik para wisatawan yang telah mengunjungi objek wisata tersebut. Informasi tersebut dapat dilihat dari *review* para wisatawan melalui *platform* seperti *e-tourism* ataupun melalui *Google Maps*. *Platform* tersebut berguna untuk mencari wilayah yang ingin dikunjungi dan banyak digunakan untuk memposting ulasan dan opini perjalanan, sehingga bisa dijadikan sebagai sumber informasi yang bagus.

Pada penelitian ini, objek wisata yang akan dijadikan objek penelitian analisis sentimen yaitu objek wisata Bukit Pelalangan Arosbaya karena merupakan tempat wisata terpopuler di Pulau Madura masuk ke dalam urutan teratas menurut *Google Maps* hingga tanggal 7 Maret 2023 yang ditunjukkan pada Gambar 1.1.



Gambar 1. 1 Tempat Wisata Populer menurut *Google Maps*

Metode mengenali atau mengategorikan opini dari teks untuk menentukan apakah opini itu positif, negatif, atau netral dikenal sebagai analisis sentimen [2]. Analisis sentimen bisa juga diartikan sebagai proses mengolah data tekstual dan dilakukan komputasi dari sekumpulan opini atau pendapat orang-orang dan emosi melalui entitas atau objek, *event* dan atribut yang dimiliki

[3]. Analisis sentimen dapat digunakan untuk mengumpulkan pendapat dalam media sosial yang berkaitan dengan pelayanan umum, topik terbaru serta dapat juga digunakan untuk mengevaluasi kinerja dan pelayanan [4].

Dengan menganalisis sentimen maka penelitian ini nantinya akan mempermudah petugas atau penyelenggara dalam pengambilan keputusan untuk langkah selanjutnya yang tepat untuk mengembangkan tempat wisata berdasarkan opini atau persepsi wisatawan objek wisata Bukit Pelalangan Arosbaya. Dari *review* yang didapatkan kemudian akan diolah dan dikategorikan sebagai opini dengan sentimen positif atau negatif.

Umumnya, terdapat dua pendekatan dalam melakukan analisis sentimen, yaitu secara *learning-based* (pendekatan menggunakan *machine learning*) dan *lexicon-based* (pendekatan berbasis leksikal) [5]. Pada pendekatan *machine learning* terbagi empat jenis, yaitu pembelajaran terbimbing (*supervised learning*), pembelajaran tidak terbimbing (*unsupervised learning*), pembelajaran semi terbimbing (*semi supervised learning*), pembelajaran dengan penguatan (*reinforcement learning*) [6]. Namun dari empat tersebut yang paling sering digunakan adalah *supervised* dan *unsupervised learning*. Sedangkan pada pendekatan *lexical* yang menjadi acuan adalah kamus dan *corpus*. *Machine learning* lebih favorit digunakan karena tidak diperlukan pemahaman mendalam mengenai struktur bahasa dari kata atau kalimat [6].

Penelitian ini dilakukan dengan pendekatan *machine learning* berdasarkan jenis pemodelan *supervised learning*. Ada beberapa pemodelan *supervised learning* dengan metode klasifikasi yang sering digunakan diantaranya yaitu *Regresi Logistik*, *K-nearest Neighbor*, *Super Vector Machine*, *Naive Bayes*, *Decision Tree* dan *Random Forest* [7].

Pada analisis sentimen tidak terlepas dari penggunaan pembobotan kata. *Term weighting* adalah sebuah metode pembobotan kata (*term*) untuk memberikan sebuah bobot atau nilai untuk kata (*term*) yang terkandung dalam sebuah dokumen [8]. Terdapat beberapa metode pembobotan kata (*term weighting*) diantaranya adalah *Term Frequency* (TF), *Term Frequency Inverse Document Frequency* (TF-IDF), *Weighted Inverse Document Frequency* (*Weighted inverse document frequency* (WIDF)), dan *Term Frequency Relevance Frequency* (TF-RF) [8].

Pembobotan yang paling populer dan menjadi teknik yang banyak digunakan dalam memberi bobot setiap kata saat ini adalah TF-IDF. Lalu pembobotan kata TF-RF merupakan metode yang masih belum dikenal banyak orang pada analisis sentiment. Oleh karena itu peneliti membandingkan teknik pembobotan kata yang populer TF-IDF dengan TF-RF menggunakan *dataset* dari ulasan wisatawan Bukit Pelalangan Arosbaya pada *platform Google Maps* untuk mengetahui performa yang optimal diantara kedua pembobotan kata tersebut dengan dikombinasikan teknik klasifikasi *Naïve Bayes*.

Untuk melakukan analisis sentimen terhadap ulasan *Google Maps* pada objek wisata Bukit Pelalangan Arosbaya, peneliti menggunakan model algoritma *Naive Bayes*. Data yang berupa *review* atau ulasan wisatawan pada *platform Google Maps* terhadap objek wisata Bukit Pelalangan Arosbaya akan diolah menggunakan *Naive Bayes* dengan cara menghitung probabilitasnya. Tingkat akurasi *Naive Bayes Classifier* yang tinggi dan dikenal *simple* adalah kelebihan algoritma *Naive Bayes* [9]. Hal ini menjadikan algoritma *Naive Bayes* cukup baik untuk digunakan dalam perhitungan yang memiliki jumlah data besar.

Penelitian ini bertujuan untuk mengetahui metode yang lebih efektif dengan membandingkan teknik pembobotan kata TF-IDF dan TF-RF jika dikombinasikan dengan teknik klasifikasi *Naive Bayes Classifier* berdasarkan ulasan wisatawan objek wisata Bukit Pelalangan Arosbaya pada *platform Google Maps*.

Penelitian yang dilakukan untuk mengetahui opini masyarakat mengenai kampus Universitas Muhammadiyah Malang (UMM) dengan mengklasifikasikan *tweets* yang berisi sentimen masyarakat mengenai UMM. Metode klasifikasi yang digunakan adalah *Naive Bayes* dan *Support Vector Machine* dengan pembobotan TF-IDF. Hasil Penelitian ini menunjukkan bahwa *Naive Bayes* mendapatkan hasil akurasi yang lebih baik dari *Support Vector Machine* dengan akurasi sebesar 73.65% [10].

Penelitian yang dilakukan untuk melakukan analisis sentimen dengan membandingkan model algoritma *Naive Bayes Classifier* yaitu *Multinomial Nave Bayes* dan *Bernoulli Naive Bayes* dengan objek sentimen dari Twitter. Penelitian ini juga menggunakan teknik *Term Weighting* yaitu TF-IDF dan TF-RF pada masing-masing model. Jadi pada penelitian ini terdapat 4 skema model yaitu *Multinomial* dan TF-IDF, *Bernoulli* dan TF-IDF, *Multinomial* dan TF-RF serta *Bernoulli* dan TF-RF. Hasil dari 4 skema yaitu TF-IDF dan TF-RF dengan skema *Bernoulli Naive Bayes* dari hasil pengujian *Confusion matrix* menghasilkan akurasi terbaik sebesar 61%, dan rata-rata nilai akurasi dari validasi *5-fold validation* sebesar 60%. Dan yang memiliki nilai akurasi terendah terletak pada model *Multinomial Naive Bayes* dan TF-IDF yaitu 58% dari *Confusion matrix*, dengan nilai rata-rata 59% dari validasi *5-fold validation* [11].

Penelitian yang dilakukan untuk menganalisis dan membandingkan algoritma TF-IDF, TF.RF, dan WIDF. Untuk metode pengujiannya akan digunakan metode klasifikasi *naïve bayes* dan analisis perbandingannya menggunakan *confussion matrix*. Dengan *dataset* yang digunakan sebanyak 130 dokumen yang mana 100 data *traning* dan 30 data uji. Berdasarkan analisa pada hasil klasifikasi yang telah dilakukan, dapat disimpulkan bahwa pembobotan TF.RF dengan

klasifikasi *Naïve bayes* lebih baik dari pembobotan TF.IDF dan WIDF dengan nilai *Accuracy* 98,67%, *Precision* 93,81%, dan *Recall* 96,67% [8].

1.2 Perumusan Masalah

1.2.1 Permasalahan

Ulasan dari Wisata Bukit Pelalangan Arosbaya pada *Google Maps* sudah banyak diunggah berupa opini perjalanan yang dirasakan pengunjung sehingga dapat dijadikan sebagai sumber informasi yang bagus untuk melakukan analisis sentimen. Oleh karena itu penelitian ini dilakukan untuk menganalisis sentimen terhadap ulasan pengunjung dengan data ulasan yang berasal dari *Google Maps* untuk melihat opini pengunjung berupa positif dan negatif agar mempermudah petugas atau penyelenggara dalam pengambilan keputusan untuk langkah selanjutnya untuk mengembangkan tempat wisata berdasarkan opini atau persepsi wisatawan objek wisata Bukit Pelalangan Arosbaya. Dengan penelitian ini juga dapat mengetahui metode yang lebih efektif dengan membandingkan teknik pembobotan kata TF-IDF dan TF-RF jika dikombinasikan dengan teknik klasifikasi *Naive Bayes Classifier*.

1.2.2 Metode Usulan

Analisis sentimen wisatawan terhadap objek wisata Bukit Pelalangan Arosbaya menggunakan metode usulan *Naive Bayes Classifier* dengan membandingkan teknik pembobotan kata *Term Frequency Inverse Document Frequency* (TF-IDF) dan *Term Frequency Relevance Frequency* (TF-RF). Metode *Naive Bayes* digunakan karena metode klasifikasinya tersebut memiliki tingkat akurasi serta kecepatan yang tinggi dengan sekumpulan data yang besar [12].

1.2.3 Pertanyaan Penelitian

1. Bagaimana hasil akurasi yang didapat dengan menggunakan metode klasifikasi *Naive Bayes* dalam pengklasifikasian ulasan wisatawan terhadap objek wisata Bukit Pelalangan Arosbaya dengan pembobotan kata TF-IDF?
2. Bagaimana hasil akurasi yang didapat dengan menggunakan metode klasifikasi *Naive Bayes* dalam pengklasifikasian ulasan wisatawan terhadap objek wisata Bukit Pelalangan Arosbaya dengan pembobotan kata TF-RF?
3. Bagaimana hasil akurasi performa keseluruhan dengan menggunakan metode klasifikasi *Naïve Bayes*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

1. Mengetahui hasil akurasi yang didapat dengan menggunakan metode klasifikasi *Naive Bayes* dalam pengklasifikasian ulasan wisatawan terhadap objek wisata Bukit Pelalangan Arosbaya dengan pembobotan kata TF-IDF.
2. Mengetahui hasil akurasi yang didapat dengan menggunakan metode klasifikasi *Naive Bayes* dalam pengklasifikasian ulasan wisatawan terhadap objek wisata Bukit Pelalangan Arosbaya dengan pembobotan kata TF-RF.
3. Mengetahui hasil akurasi performa keseluruhan menggunakan metode klasifikasi *Naive Bayes*.

1.3.2 Manfaat

1. Dapat menambah wawasan serta kemampuan berpikir dalam menerapkan teori yang telah didapat selama perkuliahan kedalam penelitian yang sebenarnya.
2. Mengetahui klasifikasi ulasan wisatawan pada *Google Maps* terhadap objek wisata Bukit Pelalangan Arosbaya.
3. Sebagai acuan referensi untuk dapat dikembangkan dalam penelitian berikutnya.

1.4 Batasan Masalah

Pada penelitian ini, peneliti memberikan batasan masalah agar tidak terjadi kesalahan pada saat penelitian dilakukan. Berikut adalah batasan masalah pada proses penelitian yaitu:

1. Data yang digunakan adalah data yang berhasil diambil dengan teknik *scraping* pada halaman web ulasan *Google Maps* terhadap objek wisata Bukit Pelalangan Arosbaya menggunakan *tool* webharvy.
2. Data yang didapatkan merupakan ulasan yang relevan.
3. Data yang digunakan merupakan data berbahasa Indonesia.
4. Data diambil pada tanggal 6 April 2023 pada situs *Google Maps*.
5. Jumlah data ulasan mentah yang berhasil terambil adalah sebanyak 1470.
6. Pelabelan dilakukan oleh pakar Bahasa Indonesia.
7. Tahap *preprocessing text*, pembobotan kata, dan klasifikasi *Naive Bayes* menggunakan *library* pada Bahasa pemrograman python.
8. Klasifikasi atau sentimen yang dilakukan dikelompokkan menjadi dua kategori yaitu: positif dan negatif.
9. Hasil penelitian berupa angka dari tabel evaluasi *confusion matrix* yaitu akurasi, presisi, *recall*, dan F1-score.

1.5 Sistematika Penulisan

Penulisan skripsi ini terdiri dari lima bab dengan masing-masing bab akan dijelaskan secara singkat sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bab ini berisi tentang teori-teori pendukung skripsi yang digunakan untuk memahami permasalahan yang dibahas pada penelitian ini yaitu tentang analisis sentimen, *preprocessing text*, *Google Maps*, *Scraping data Google Maps*, *Text Preprocessing*, *TF-IDF*, *TF-RF*, *Naive Bayes Classifier*, *confusion matrix*, *k-fold cross validation*, Bukit Pelalangan Arosbaya, dan penelitian terdahulu.

BAB III METODE USULAN

Bab ini dijelaskan mengenai alur penelitian, proses pada analisis sentimen ulasan wisatawan pada *Google Maps* terhadap objek wisata Bukit Pelalangan Arosbaya yaitu tahap pengumpulan data, tahap *labelling*, tahap *preprocessing data*, tahap *TF-IDF*, tahap *TF-RF*, skenario pengujian, dan rincian jadwal pengajuan proposal penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan implementasi sistem yang dikerjakan sesuai dengan rancangan yang dibuat pada bab sebelumnya. Pada bab ini terdapat tahap pengumpulan data, tahap pelabelan data, tahap *preprocessing text* (*data cleansing*, *case folding*, konversi *slangword*, *tokenizing*, *filtering*, dan *stemming*), tahap pembobotan *TF-IDF*, tahap pembobotan *TF-RF*, tahap klasifikasi dengan *naïve bayes classifier*, tahap evaluasi model menggunakan *confussion matrix* dan *k-fold cross validation*, dan analisis hasil dari uji skenario.

BAB V PENUTUP

Pada bab ini dijelaskan mengenai kesimpulan dari hasil penelitian berdasarkan pertanyaan penelitian dan juga terdapat saran bagi penelitian berikutnya yang akan mengangkat topik yang sama yaitu analisis sentimen.

DAFTAR PUSTAKA

Bab ini berisi keterangan mengenai sumber rujukan yang digunakan dalam penyusunan tugas akhir. Pada penelitian ini daftar pustaka disusun secara otomatis dengan menggunakan aplikasi *endnote*.