

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Semakin pesatnya pertumbuhan internet dan meningkatnya popularitas media sosial memungkinkan orang untuk secara terbuka mengungkapkan pendapat dan pengalaman akan suatu hal. Informasi yang terdapat di Internet sangat beragam, salah satunya adalah informasi mengenai perspektif pelanggan terhadap aspek pariwisata pantai Malang Selatan. Perspektif pelanggan bisa didapatkan dari opini pelanggan dalam bentuk ulasan yang ada pada platform Web. Namun, begitu banyaknya opini akan mempersulit pengguna lain untuk memperoleh informasi tersebut. Analisis sentimen merupakan sebuah proses untuk menganalisis atau mengidentifikasi sebuah opini seseorang yang menunjukkan sikap terhadap suatu topik atau produk tertentu masuk ke dalam kategori positif, negatif, atau netral [1]. Secara umum analisis sentimen dibagi menjadi tiga level, yaitu level dokumen, level kalimat, dan level aspek. Pada penelitian ini menggunakan analisis sentimen level aspek. Analisis sentimen level aspek menunjukkan performa yang lebih baik dibandingkan level dokumen dan level kalimat. Hal ini dikarenakan dalam suatu opini yang dikemukakan seseorang cenderung membahas tiap aspek suatu entitas, bukan keseluruhan [2]. Entitas dalam analisis sentimen berbasis aspek adalah objek yang dapat berupa produk, layanan, topik, isu, orang, organisasi atau *event* sedangkan aspek adalah bagian atau atribut dari entitas [3].

Analisis sentimen memerlukan beberapa hal yang harus dipersiapkan terlebih dahulu, salah satunya dengan memilih *classifier* yang akan digunakan. *Classifier* metode yang dapat mengklasifikasikan data menjadi beberapa kelas [4]. Dalam penelitian ini, *classifier* yang dipilih adalah *Support Vector Machine* (SVM), merupakan salah satu metode yang sangat populer paling kuat dan akurat perihal klasifikasi maupun regresi [5]. Tetapi proses klasifikasi pada *Machine Learning* sering kali ditemukan kelemahan terhadap pemilihan fitur atau parameter yang dapat mempengaruhi akurasi. Seleksi fitur adalah alternatif cara

untuk mengurangi fitur yang tidak relevan sehingga dimensi data berkurang dan diperoleh performa klasifikasi yang tinggi. Fitur individu terbaik dapat dilakukan menggunakan beberapa metode, salah satunya *Information Gain* (IG) [6]. Kemudian, dalam penulisan tanggapan maupun saran sering ditemukan beberapa kesalahan yang dapat mempersulit proses klasifikasi. Kesalahan tersebut dapat berupa kesalahan pada ejaan, aksen bahasa, dan gramatikal dalam bahasa tulis yang menyulitkan pemrosesan bahasa alami [7]. Perbaikan kata tidak baku dalam penerapannya memiliki banyak algoritma. *Peter Norvig* merupakan algoritma yang menggabungkan operasi *delete*, *insert*, *replace*, dan *transpose* dari kata yang terdeteksi *typo* dan mencarinya dalam kamus [8]. Sebelum dilakukan klasifikasi diperlukan pembobotan fitur untuk membantu meningkatkan akurasi klasifikasi, pembobotan yang umum digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) [9].

Berdasarkan penelitian sebelumnya terkait analisis sentimen berbasis aspek di kasus pariwisata pantai Malang Selatan menyatakan bahwa perlu adanya penggunaan metode untuk memperbaiki penggunaan kata yang tidak baku dengan tujuan untuk mengoptimalkan jumlah fitur yang akan digunakan pada tahapan klasifikasi [10]. Berikutnya penelitian terkait fungsi penggunaan seleksi fitur pada proses klasifikasi menyatakan bahwa seleksi fitur berguna untuk mereduksi fitur-fitur sehingga proses klasifikasi bisa lebih efektif dan efisien [11]. Kemudian penelitian terkait normalisasi kata pada tahap *pre-processing* dan menggunakan seleksi fitur *Information Gain* menyatakan bahwa seleksi fitur *Information Gain* dengan normalisasi terbukti dapat meningkatkan hasil akurasi yang sebelumnya sebesar 94% untuk *pre-processing* tanpa normalisasi kata dan sebesar 98% untuk *preprocessing* dengan normalisasi kata yang dipadukan dengan *Information Gain* sebagai seleksi fitur [12]. Selanjutnya penelitian terkait Naive Bayes dikombinasikan dengan metode seleksi fitur yaitu *Information Gain* sebagai metode seleksi dalam memilih fitur-fitur yang berpengaruh untuk setiap label kelas menunjukkan bahwa, akurasi dan f-measure yang didapat pada klasifikasi dengan seleksi fitur *Information Gain* yaitu 91,33% dan 89,18% [13]. Kemudian penelitian terkait penggunaan spelling correction dengan metode Peter Norvig dan

N-Gram, hasil pengujian menunjukkan bahwa kombinasi dari kedua metode tersebut memberikan tingkat akurasi sebesar 73,684% dan tingkat kesuksesan 37,037% untuk total akurasi dari aplikasi ini adalah 69,09% [8]. Lalu penelitian terkait normalisasi kata menggunakan *Peter Norvig*, hasil penelitian tersebut menunjukkan normalisasi kata menggunakan *Peter Norvig* menghasilkan 0,903 untuk *precision*, 0,944 untuk *recall*, 0,922 untuk *f-measure*, dan 0,903 untuk *accuracy* [14].

Berdasarkan latar belakang yang telah diuraikan dari beberapa penelitian yang sudah dilakukan sebelumnya, maka penelitian ini akan berfokus pada analisis sentimen yang dilakukan pada level aspek menggunakan metode *Support Vector Machine* (SVM) dilanjutkan perbaikan pada kata yang tidak baku menggunakan algoritma *Peter Norvig* dengan Seleksi Fitur *Information Gain* untuk mengoptimalkan jumlah *feature*, kemudian dilakukan pembobotan data menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) yang akan digunakan pada analisis sentimen berbasis aspek terhadap *review* pariwisata pantai Malang Selatan.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan, merumuskan bagaimana pengaruh seleksi fitur *Information Gain* dan perbaikan kata dengan algoritma *Peter Norvig* terhadap tingkat akurasi metode *Support Vector Machine* (SVM) dalam mengklasifikasikan Analisis Sentimen Berbasis Aspek pada *review* pariwisata pantai Malang Selatan.

1.3. Maksud dan Tujuan Penelitian

Maksud dari penelitian ini yaitu mengimplementasikan metode *Support Vector Machine* dan pengaruh *Information Gain* dengan normalisasi kata *Peter Norvig* dalam melakukan analisis sentimen berbasis aspek. Adapun tujuan dilakukannya penelitian ini adalah untuk mengetahui performansi akurasi dari optimalisasi fitur *Information Gain* setelah metode *Support Vector Machine* (SVM) dikombinasikan dengan perbaikan pada kata yang tidak baku dengan *Peter*

Norvig pada analisis sentimen berbasis aspek terhadap *review* pariwisata pantai Malang Selatan.

1.4. Batasan Masalah

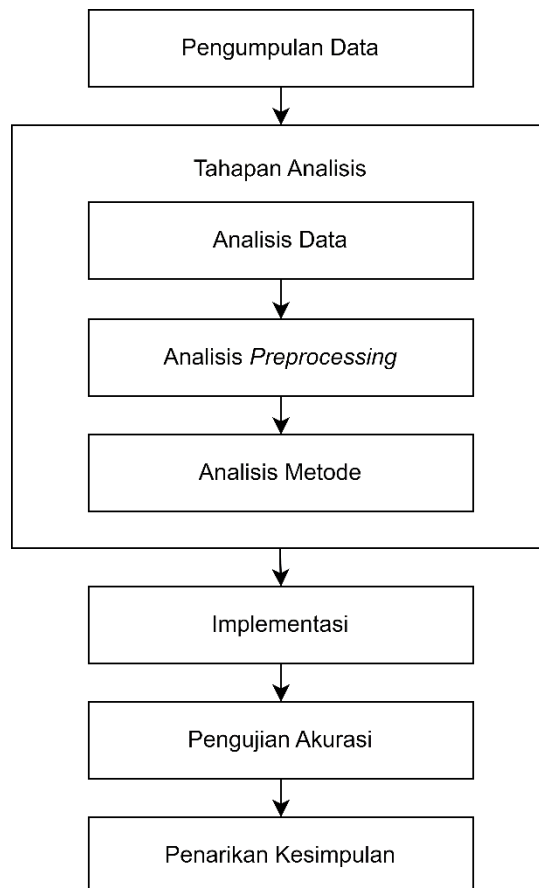
Berdasarkan masalah yang telah didapatkan, maka batasan yang ditentukan pada penelitian ini adalah sebagai berikut:.

1. Data opini yang diklasifikasikan adalah opini berbahasa Indonesia.
2. Data bersumber dari situs TripAdvisor menggunakan teknik *Web Scraping* untuk mendapatkan teks opini dengan menggunakan kata kunci pantai Malang Selatan pada tanggal 1 Januari 2013 sampai 1 Juni 2022.
3. Proses pelabelan kata dilakukan secara manual.
4. Dataset dilabelkan ke dalam kelas positif, negatif, dan netral.
5. Aspek yang digunakan pada penelitian ini meliputi aspek umum, aspek kebersihan, aspek keramaian, aspek akses jalan, dan aspek kondisi ombak [10].
6. Metode yang dipakai untuk melakukan klasifikasi adalah Support Vector Machine (SVM), kemudian *Term Frequency-Invers Document Frequency* (TF-IDF) digunakan dalam pembobotan kata.
7. Perbaikan kata tidak baku diselesaikan menggunakan pencocokan dengan kamus SS yang disimpan dalam format .csv lalu diberi nama *slang words* dan singkatan [22].
8. *Spelling Corrector* yang diperbaiki merupakan *non-word* dengan metode *Peter Norvig* menggunakan korpus.txt [23].
9. Peter Norvig hanya dapat mengoreksi 1 huruf yang salah [8]
10. Pada *Peter Norvig* dilakukan dengan cara memeriksa semua kata yang ada berdasarkan korpus, diperoleh dari korpus PAN-BPPT Localization Bahasa Indonesia, korpus berita tempo dan korpus *Leipzig Corpora Collection* [16].

1.5. Metodologi Penelitian

Dalam penelitian ini, metode penelitian yang digunakan adalah metode penelitian kuantitatif dimana masing-masing tahapan dilakukan secara berurutan

dengan tujuan untuk menguji hipotesis yang telah ditetapkan. Berikut merupakan proses tahapan penelitian dengan alur seperti pada Gambar 1.1.



Gambar 1. 1. Metodologi Penelitian

1.5.1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Metode Studi Literatur merupakan teknik pengumpulan data dengan cara mengumpulkan jurnal, paper, atau buku yang berkaitan dengan analisis sentimen berbasis aspek dan pengaruh normalisasi dengan seleksi fitur yang diterbitkan dari tahun 2018 sampai 2022.

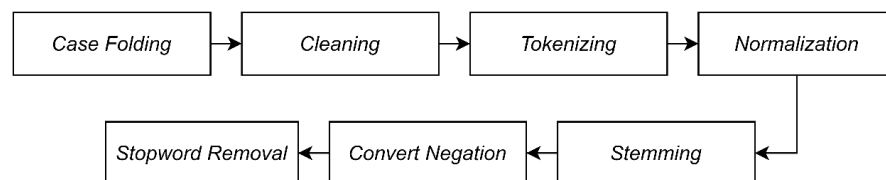
2. Pengumpulan Dataset

Pengumpulan data diambil dari situs TripAdvisor dalam bentuk opini yang diberikan oleh pelanggan terhadap objek wisata pantai Malang Selatan menggunakan metode Web Scraping.

1.5.2. Tahapan Analisis

Berikut merupakan tahapan analisis pada penelitian ini adalah sebagai berikut:

1. **Analisis Data**, menganalisis pola kalimat. Data yang dianalisis adalah kalimat dari opini yang diberikan oleh pelanggan terhadap objek wisata pantai Malang Selatan dari situs TripAdvisor dengan kata kunci pantai Malang Selatan.
2. **Analisis Text Pre-processing**, kalimat dari opini yang terkumpul akan diproses sehingga data yang didapat menjadi lebih terstruktur dan mudah untuk diolah. Tahapan pre-processing yang dilakukan terdiri dari *Case Folding*, *Cleaning*, *Tokenizing*, *Normalization*, *Stemming*, *Convert Negation*, *Stopword Removal*.



Gambar 1. 2. Tahapan Preprocessing

3. **Analisis Metode**, analisis klasifikasi menggunakan *Support Vector Machine* (SVM) meliputi metode perbaikan penggunaan kata tidak baku yaitu *Peter Norvig* dan metode *Information Gain* sebagai metode seleksi fitur.

1.5.3. Implementasi

Pada tahapan implementasi yaitu melakukan implementasi yang diawali dengan memasukkan dan membaca data ulasan yang telah dilabeli secara manual, kemudian melakukan implementasi preprocessing yaitu: *Case Folding*, *Cleaning*, *Tokenizing*, *Normalization*, *Stemming*, *Convert Negation*, *Stopword Removal*, setelah data masukkan dipraproses. Kemudian data yang sudah melalui

preprocessing perlu diubah menjadi bentuk numerik, sehingga dilakukan pembobotan kata dengan TF-IDF. Lalu melakukan proses seleksi fitur *Information Gain* untuk menentukan batas dari kepentingan sebuah atribut. Tahap selanjutnya yaitu membangun model *Support Vector Machine* (SVM) yang nantinya akan dilakukan proses pelatihan dan pengujian berupa klasifikasi aspek juga sentimen..

1.5.4. Pengujian Akurasi

Dalam tahapan ini akan dilakukan pengujian akurasi dari metode *Support Vector Machine* (SVM). Berikutnya metode *Support Vector Machine* (SVM) dikombinasikan dengan seleksi fitur *Information Gain*. Kemudian metode *Support Vector Machine* (SVM) dikombinasikan dengan normalisasi kata Peter Norvig. Selanjutnya metode *Support Vector Machine* (SVM) dikombinasikan dengan seleksi fitur *Information Gain* dan normalisasi kata *Peter Norvig* yang akan digunakan pada analisis sentimen berbasis aspek terhadap *review* pariwisata pantai Malang Selatan.

1.5.5. Penarikan Kesimpulan

Pada tahapan ini dilakukan penarikan kesimpulan berdasarkan hasil dari tahapan pengujian dalam mengoptimalkan jumlah feature dengan seleksi fitur *Information Gain* dan normalisasi kata *Peter Norvig* juga klasifikasi menggunakan metode *Support Vector Machine* (SVM) pada analisis sentimen berbasis aspek terhadap *review* pariwisata pantai Malang Selatan.

1.6. Sistematika Penulisan

Adapun sistematika penulisan laporan penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan laporan penelitian ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian serta sistematika penulisan untuk menggambarkan keseluruhan mengenai penelitian ini.

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai berbagai konsep dan dasar-dasar teori yang menunjang dan berkaitan dengan penelitian ini.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini berisikan tentang analisis terhadap sistem yang akan dibangun meliputi analisis masalah, analisis proses, dan analisis pada kebutuhan *non-fungsional*.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisikan hasil implementasi dari analisis dan pengujian terhadap sistem yang sudah dibangun.

BAB 5 KESIMPULAN DAN SARAN

Pada bab ini membahas tentang kesimpulan dari penelitian ini dan saran yang diberikan untuk proses pengembangan perangkat lunak ini agar lebih baik.