

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan kebutuhan informasi saat ini mengalami pertumbuhan yang pesat. Informasi yang dibutuhkan harus dapat diperoleh dengan cepat, akurat, dan sesuai dengan kebutuhan. Di era digital ini, internet menjadi sumber utama bagi banyak orang untuk mendapatkan informasi. Salah satu jenis informasi yang sering dicari adalah dokumen teks, dan fasilitas pencarian sering digunakan untuk menemukannya. Proses pencarian informasi dalam bentuk dokumen teks, yang dikenal sebagai Information Retrieval (IR), melibatkan pemisahan dokumen-dokumen yang dianggap relevan dari kumpulan dokumen yang tersedia [1]. Dalam tahap ini, stemming menjadi aspek krusial dalam Information Retrieval. Dalam konteks penelitian ini, stemming akan digunakan untuk memperoleh kata-kata dasar yang dapat mewakili opini, sentimen, dan emosi.

Stemming adalah proses pengubahan kata-kata menjadi bentuk dasarnya (root word) dengan menghilangkan imbuhan seperti awalan, akhiran, dan sisipan [2]. Selain itu, analisis sentimen juga menjadi penting dalam mengidentifikasi dan mengklasifikasikan opini, sentimen, dan emosi dalam teks. Analisis sentimen dapat dilakukan menggunakan berbagai metode, salah satunya adalah dengan menggunakan algoritma stemming. Saat ini, terdapat dua algoritma stemming bahasa Indonesia yang populer digunakan, yaitu algoritma stemming Nazief & Adriani dan algoritma stemming Arifin Setiono. Algoritma stemming Nazief & Adriani menggunakan kamus kata dasar sebagai acuan dan mendukung recoding, yakni penyusunan kembali kata-kata yang mengalami proses stemming berlebih [3]. Sementara itu, algoritma stemming Arifin Setiono menggunakan metode pemeriksaan dengan 12 kombinasi untuk mengatasi masalah overstemming, yang sering terjadi pada algoritma pemotongan imbuhan [4]. Namun, belum jelas algoritma stemming mana yang lebih efektif dalam menganalisis sentimen pengguna Twitter terhadap TikTok Shop.

Dalam konteks ini, penelitian ini bertujuan untuk membandingkan implementasi algoritma stemming Nazief & Adriani dan algoritma stemming Arifin Setiono dalam analisis sentimen terhadap TikTok Shop. Metode klasifikasi seperti Naive Bayes dan Support Vector Machine (SVM) akan digunakan sebagai pendukung analisis sentimen. Naive Bayes dipilih karena sifatnya yang sederhana, cepat dan mampu menghasilkan akurasi yang tinggi dalam klasifikasi teks, serta memiliki implementasi yang tersedia dalam bahasa pemrograman Python yang populer [5]. SVM dipilih karena kemampuannya menangani data dengan dimensi tinggi dan kelas yang tidak seimbang dengan baik, serta juga memiliki implementasi yang tersedia dalam Python [6]. Oleh karena itu, penggunaan kedua metode klasifikasi ini diharapkan dapat memberikan hasil yang komprehensif dan dapat diandalkan dalam analisis sentimen terhadap TikTok Shop.

Data penelitian akan berupa komentar pengguna Twitter terhadap TikTok Shop, yang kemudian akan dikelompokkan menjadi sentimen positif, negatif, dan netral. Selanjutnya, kedua algoritma stemming, yaitu Nazief & Adriani dan Arifin Setiono, akan diimplementasikan pada dataset tersebut. Untuk mengukur performa dari kedua metode tersebut, akan diukur waktu proses stemmingnya dan menggunakan metrik evaluasi seperti akurasi, presisi, recall dan F1-score. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan yang berharga mengenai efektivitas kedua algoritma stemming dalam konteks analisis sentimen pengguna Twitter terhadap TikTok Shop.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang masalah yang dikemukakan, maka dapat dirumuskan beberapa masalah sebagai berikut:

1. Bagaimana implementasi algoritma stemming Nazief & Adriani dalam menganalisis sentimen pengguna Twitter terhadap TikTok Shop menggunakan naive bayes dan SVM?

2. Bagaimana implementasi algoritma stemming Arifin Setiono dalam menganalisis sentimen pengguna Twitter terhadap TikTok Shop menggunakan naive bayes dan SVM?
3. Bagaimana perbandingan implementasi algoritma stemming Nazief & Adriani dan algoritma stemming Arifin Setiono dalam menganalisis sentimen pengguna Twitter terhadap TikTok Shop menggunakan naive bayes dan SVM?

1.3 Batasan Masalah

Batasan masalah yang akan digunakan untuk mempersempit masalah sebagai berikut:

1. Penelitian ini hanya membandingkan implementasi dua algoritma stemming, yaitu algoritma stemming Nazief & Adriani dan algoritma stemming Arifin Setiono, dalam menganalisis sentimen pengguna Twitter terhadap TikTok Shop.
2. Data yang digunakan adalah komentar pengguna Twitter terhadap TikTok Shop.
3. Metode klasifikasi yang digunakan adalah Naive Bayes dan Support Vector Machine (SVM).
4. Penelitian ini tidak membahas algoritma stemming lainnya atau metode klasifikasi selain Naive Bayes dan SVM.
5. Analisis sentimen yang dilakukan akan mengelompokkan komentar pengguna menjadi tiga kategori, yaitu positif, negatif, dan netral.
6. Penelitian ini fokus pada sentimen pengguna Twitter terhadap TikTok Shop, dengan pengguna Twitter yang memiliki bahasa dalam Bahasa Indonesia sebagai subjek penelitian.
7. Bahasa pemrograman yang digunakan adalah Python serta google Collaboratory sebagai platform.

1.4 Tujuan Penelitian

Tujuan penelitian yang dilakukan adalah sebagai berikut:

1. Membandingkan efektivitas implementasi algoritma stemming Nazief & Adriani dan algoritma stemming Arifin Setiono dalam menganalisis sentimen pengguna Twitter terhadap TikTok Shop menggunakan metode klasifikasi Naive Bayes dan Support Vector Machine (SVM).
2. Memberikan wawasan yang berharga tentang mana algoritma stemming yang lebih efektif dalam konteks analisis sentimen terhadap TikTok Shop berdasarkan hasil perbandingan dan evaluasi menggunakan metode klasifikasi Naive Bayes dan Support Vector Machine (SVM).

1.5 Manfaat Penelitian

Manfaat dari penelitian yang dilakukan adalah sebagai berikut:

1. Mengetahui dan memahami cara mengimplementasi algoritma stemming Nazief & Adriani dengan algoritma Arifin Setiono dalam analisis sentimen.
2. Memberikan pengetahuan tentang efektivitas algoritma stemming Nazief & Adriani dan algoritma stemming Arifin Setiono dalam analisis sentimen.
3. Membantu penelitian selanjutnya dalam memilih algoritma stemming yang tepat untuk analisis sentimen.
4. Memberikan bahan referensi untuk penelitian selanjutnya tentang analisis sentimen.

1.6 Sistematika Penulisan

Penulisan skripsi ini ditulis secara sistematis yang terdiri dari beberapa sub bab. Berikut uraian sistematika penulisan

1.6.1 BAB I : Pendahuluan

Pada bab ini akan dijelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan

1.6.2 BAB II : Tinjauan Pustaka

Pada bab ini akan dijelaskan tentang kajian pustaka, dasar teori dan kutipan penelitian sebelumnya

1.6.3 BAB III : Metode Penelitian

Pada bab ini akan menganalisis tentang desain sistem Algoritma Nazief & Adriani dan Algoritma Arifin Setiono.

1.6.4 BAB IV : Implementasi dan Pembahasan

Pada bab ini akan membahas tentang hasil analisis dan penerapan Algoritma Nazief & Adriani dan Algoritma Arifin Setiono dalam proses stemming pada sentimen analisis.

1.6.5 BAB V : Penutup

Pada bab ini akan berisikan kesimpulan dan saran dari perumusan masalah yang disampaikan.

