

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara kepulauan yang memiliki potensi pariwisata yang besar karena Indonesia memiliki aneka ragam budaya, suku, kerajinan tangan, kekayaan alam, dan sebagainya. Dengan banyaknya kekayaan alam dan budaya yang ada serta didukung dengan iklim tropis, membuat wisatawan asing berbondong bondong datang ke Indonesia. Berdasarkan data dari Kementerian Pariwisata dan Ekonomi Kreatif (KEMENPAREKRAF) kunjungan wisatawan mancanegara per juli 2021- juli 2022 meningkat sebesar 6.396,46% [1]. Dengan meningkatnya kunjungan wisatawan secara tidak langsung pertumbuhan ekonomi daerah wisata juga meningkat terutama di wilayah Jawa Timur, menurut data Badan Pusat Statistik (BPS) kunjungan wisatawan nusantara ke daerah jawa timur per tahun 2021 sebanyak 155.114.120 kunjungan [2]. Berdasarkan penelitian yang dilakukan oleh Chairunisalda [3], jumlah kunjungan wisatawan memiliki pengaruh positif dan sangat besar terhadap pendapatan asli daerah (PAD) Jawa Timur, dimana setiap kunjungan wisata mengalami kenaikan sebesar 1% maka kenaikan pendapatan asli daerah akan naik sebesar 35950,25.

Sebelum wisatawan berkunjung ke objek wisata, wisatawan akan melakukan survey terlebih dahulu terhadap objek wisata yang ada. Survey mengenai objek wisata dapat dari berita, media sosial, dan lain sebagainya. Namun biasanya wisatawan akan mencari di Google tentang wisata yang ada di daerah yang mereka kunjungi, dari Google akan di tampilkan artikel / berita tentang rekomendasi atau berita tentang yang mereka cari. Jika wisatawan membaca berita mengenai objek wisata yang negatif seperti “KKP Kumpulkan 1,4 Ton Sampah dari Pantai Teleng Ria Pacitan” maka dia akan berpikir dua kali untuk mengunjunginya dan pemasokan pemerintah daerah dari segi pariwisata akan berkurang begitu juga dengan masyarakat di sekitar wilayah pariwisata yang mengandalkan wisatawan untuk mencari nafkah. Sedangkan jika wisatawan membaca mengenai objek wisata yang positif seperti “10 Air Terjun di Jawa Timur, Menyegarkan Raga Memanjakan Mata” maka akan menimbulkan kesan bahwa dia harus mengunjungi wisata itu dan

tempat wisata tersebut akan ramai akan wisatawan yang datang dan meningkatkan kesejahteraan masyarakat di sekitar wilayah pariwisata.

Berdasarkan permasalahan yang diuraikan diatas maka diperlukan adanya suatu solusi untuk menganalisis sentimen dari berita tersebut. Salah satu metode untuk menganalisis sentimen yang populer adalah Sentimen Analysis, Sentimen Analysis atau bisa disebut dengan *opinion mining* adalah proses menganalisis suatu text guna mengetahui nada atau sentimen dari penulis, apakah sentimen dari tulisannya cenderung ke positif, negatif, atau netral [4]. Sentimen analysis sering digunakan untuk melakukan analisa terhadap suatu topik tertentu, review terhadap sebuah produk, dan lain sebagainya. Dengan menggunakan sentimen analysis kita dapat mengetahui sentimen dari berita yang ada, dan bisa membantu pemerintah daerah Jawa Timur untuk menampilkan berita yang positif saja terhadap berita pariwisata yang ada, sedangkan berita yang negatif akan di sembunyikan.

Banyak metode yang dapat digunakan untuk melakukan sentimen analysis seperti *Naïve Bayes*, *K-Nearest Neighbor*, *Logistic regression*, *Maximum entropy*, SVM dan sebagainya. Tetapi metode yang paling sering digunakan untuk melakukan Sentimen Analysis adalah metode *K-Nearest Neighbor* (KNN), *Naïve Bayes*, dan *Support Vector Machine* (SVM) [5]. Metode KNN, *Naïve Bayes* dan SVM tentu memiliki kelebihan dan kekurangannya masing-masing, metode KNN memiliki kelebihan tangguh dalam proses training yang noise dan memiliki kekurangan yaitu perlu menentukan nilai dari parameter K atau jumlah dari tetangga terdekat, *Naïve Bayes* memiliki kelebihan yaitu hanya membutuhkan sedikit data training untuk dapat mengestimasi parameter yang dibutuhkan untuk klasifikasi dan memiliki kekurangan yaitu klasifikasi *Naïve Bayes* tidak akan berlaku jika probabilitas kondisinya bernilai 0 karena jika nilai probabilitas bernilai 0 maka probabilitas prediksinya juga bernilai 0, SVM memiliki kelebihan yaitu mampu mengklasifikasikan teks dalam data yang banyak serta lengkap dan memiliki kekurangan yaitu proses training memakan waktu yang cukup lama [6]. Pada penelitian yang dilakukan oleh Fajar Sodik, dkk [7], dilakukan perbandingan metode SVM, *Naïve Bayes*, dan KNN dengan data sebanyak 10.000 data dan didapatkan hasil bahwa metode SVM lebih baik dari metode *Naïve Bayes* dan KNN

dengan hasil akurasi sebesar 90,1%, *Naïve Bayes* memperoleh nilai akurasi sebesar 79,2%, dan KNN memperoleh nilai akurasi sebesar 62,1%.

Meskipun hasil akurasi yang didapatkan oleh metode SVM memiliki nilai tertinggi, metode SVM memiliki kelemahan ketika data yang digunakan memiliki dimensi yang tinggi. Menurut Yu dan Liu [8] data yang berdimensi tinggi dapat berisi informasi yang tidak relevan dan mengakibatkan menurunnya kinerja dari SVM. Berdasarkan permasalahan tersebut maka dibutuhkan optimasi pada ekstraksi fitur. Penelitian yang dilakukan oleh Fitriyani, dkk [9], menggunakan N-Gram dengan SVM sebagai classifier mendapatkan hasil, bahwa penambahan N-Gram berhasil menambah akurasi sebesar 1% dengan menggunakan data sebanyak 1.200. Penelitian yang dilakukan oleh Mita Oktafiani, dkk [10] tentang menganalisis sentimen komentar pada aplikasi tanda tangan digital menggunakan n-gram dan svm, didapatkan n-gram meningkatkan akurasi sebesar 2,54% dengan nilai akurasi sebesar 98,90%. Berdasarkan penelitian terdahulu didapatkan metode N-gram dapat mengatasi kelemahan SVM dengan cara optimisasi pada ekstraksi fitur.

Berdasarkan uraian latar belakang, maka penelitian ini akan mencoba mengkombinasikan antara N-Gram dengan SVM. Dengan harapan N-Gram mampu meningkatkan akurasi dari classifier SVM.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Berdasarkan latar belakang diatas, didapatkan permasalahan bahwa metode SVM dengan data yang berdimensi tinggi akan menurunkan kinerja dari SVM. Sehingga membutuhkan metode untuk optimasi pada ekstraksi fitur.

1.2.2 Solusi Permasalahan

Berdasarkan permasalahan yang ada, penelitian ini mengusulkan penambahan metode N-Gram untuk membantu SVM dalam melakukan optimisasi pada ekstraksi fitur.

1.2.3 Pertanyaan Penelitian

1. Bagaimana pengaruh dari penerapan N-Gram terhadap sistem klasifikasi sentimen menggunakan SVM?

2. Kernel SVM mana yang terbaik untuk kasus Sentiment Analysis Berita?

1.3 Batasan Masalah

1. Web yang digunakan untuk *crawling* berita hanya dari web Detik dengan url www.detik.com, Antara News dengan url www.jatim.antaranews.com, Berita Jatim dengan url www.beritajatim.com, Jatim Now dengan url www.jatimnow.com, Media Madura dengan url www.mediamadura.com
2. Berita yang digunakan hanya berita pariwisata Jawa Timur berbahasa Indonesia yang terbit dari bulan Januari 2021 - Oktober 2022

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

1. Mengetahui perubahan akurasi dari penerapan N-Gram dengan $n=1$ sampai dengan $n=3$ terhadap sistem klasifikasi sentimen menggunakan SVM.
2. Mengetahui kernel SVM yang terbaik dalam kasus Sentiment Analysis Berita

1.4.2 Manfaat

1. Membantu pemerintah Jawa Timur dalam mengetahui sektor pariwisata yang sudah baik atau yang buruk pada berita yang ada.
2. Sebagai bahan rujukan untuk penelitian selanjutnya

1.5 Sistematika Penulisan

Sistematika penulisan pada laporan tugas akhir ini adalah sebagai berikut:

Bab 1 membahas latar belakang mengenai ruang lingkup masalah yang dibahas, solusi yang diusulkan untuk menyelesaikan permasalahan, tujuan serta manfaat dengan diterapkannya solusi, dan sistematika penulisan.

Bab 2 menjelaskan mengenai materi yang ada dalam penelitian ini dan penelitian terkait yang dijadikan dasar dalam melakukan penelitian.

Pada bab 3 dijelaskan tentang arsitektur sistem, *dataset* yang akan digunakan, tahapan pengujian, dan jadwal pengerjaan.

Bab 4 menjelaskan terkait hasil klasifikasi sentimen yang dihasilkan menggunakan metode SVM dan N-Gram

Bab 5 memuat mengenai kesimpulan yang diperoleh dari penelitian yang telah dilakukan serta saran untuk mengembangkan penelitian lebih lanjut