

BAB I PENDAHULUAN

1.1 Latar Belakang

Sistem Informasi Pelayanan atau yang sering disingkat *SIP* Disdukcapil Kabupaten Jember merupakan sebuah *website* yang disediakan oleh Disdukcapil Kabupaten Jember teruntuk warga Jember agar dapat mengakses layanan administrasi kependudukan secara daring. Pada *website* tersebut terdapat fitur *Indeks Kepuasan Masyarakat* yang menampung data ulasan dari pengguna. Namun dalam penerapannya Disdukcapil Kabupaten Jember mengalami kesulitan dalam melakukan analisis terhadap ulasan yang masuk secara manual, karena jumlah data yang terus meningkat dan membutuhkan waktu yang cukup lama. Kondisi ini menyebabkan opini masyarakat belum dapat dimanfaatkan secara optimal sebagai dasar evaluasi dan pengambilan keputusan kebijakan. Oleh karena itu, diperlukan suatu metode yang mampu menganalisis data ulasan secara cepat, sistematis, dan objektif. Analisis Sentimen yang dapat disebut juga dengan *opinion mining* merupakan sebuah bidang studi yang menganalisa opini, sentiment, evaluasi, penilaian, sikap, dan perasaan masyarakat terhadap entitas seperti produk, pelayanan, organisasi, individu, isu, peristiwa, topik, dan atributnya (Liu, 2022). Dengan melakukan analisis sentimen, Disdukcapil Kabupaten Jember dapat lebih mudah memperoleh informasi terkait kualitas layanan, khususnya dengan memberikan penekanan pada sentimen negatif. Sentimen negatif umumnya berisi seruan, kritik, maupun ajakan untuk melakukan pembenahan, sedangkan sentimen positif cenderung berisi apresiasi terhadap layanan yang diberikan. Oleh karena itu, dalam konteks evaluasi layanan, sentimen negatif dinilai lebih informatif dalam mengidentifikasi kelemahan dan kekurangan yang perlu diperbaiki. Sehingga kesalahan proses klasifikasi sentimen positif lebih dapat ditoleransi daripada kesalahan klasifikasi sentimen negatif, karena kesalahan klasifikasi sentimen negatif dapat berpotensi kehilangan informasi terkait hal-hal yang perlu diperbaiki.

Dalam melakukan Analisis Sentimen dibutuhkan pemilihan algoritma klasifikasi yang tepat agar mendapatkan hasil analisis yang akurat, adapun algoritma klasifikasi tersebut dapat berupa *Naïve Bayes* dan *K-Nearest Neighbor*.

Algoritma *Naïve Bayes* menggunakan teori keputusan Bayesian (Surohman, dkk., 2020). Teori keputusan *Bayesian* merupakan konsep fundamental dalam analisis pola dan pengenalan pola yang mengacu pada pendekatan pemodelan probabilitas untuk membuat keputusan. Berbeda dengan *Naïve Bayes*, algoritma *K-Nearest Neighbor* mengklasifikasikan data baru dengan mencari data-data lain (tetangga terdekat) yang paling mirip. Kelas yang paling sering muncul di antara tetangga terdekat tersebut menjadi kelas pada data yang diprediksi (Prasetyo, dkk., 2023). Dengan perbedaan algoritma dalam pemrosesan tersebut, maka hasil yang didapat dari setiap model tersebut dapat berbeda juga, sebagai contoh pada penelitian Surohman, dkk. (2020) yang menguji kinerja algoritma klasifikasi *Naïve Bayes*, dan *K-Nearest Neighbor* dalam Analisis Sentimen pada ulasan penggunaan *FinTech* di Aplikasi Dana, didapat akurasi *Naïve Bayes* sebesar 84,85% dan *K-Nearest Neighbor* sebesar 82,96%. Sebagai perbandingan lain, pada penelitian Hakim, dkk. (2024) dengan penelitian Analisis Sentimen data *tweet* dengan topik Kereta Cepat Jakarta – Bandung pada aplikasi *X* didapatkan hasil akurasi *Naïve Bayes* 99.49% dan *K-Nearest Neighbor* 99.76 %. Pada penelitian Prasetyo, dkk. (2021), terkait penelitian perbandingan metode *Naïve Bayes*, *K-Nearest Neighbor*, dalam Analisis Sentimen relokasi Ibu Kota Negara pada aplikasi *X*, didapatkan hasil akurasi *Naïve Bayes* sebesar 82.27% dan, *K-Nearest Neighbor* sebesar 88,12%.

Berdasarkan hasil dari penelitian terdahulu tersebut, demi mendapatkan hasil Analisis Sentimen yang baik, maka tujuan dari penelitian ini adalah melakukan perbandingan Analisis Sentimen terhadap model algoritma klasifikasi *Naïve Bayes* dengan *K-Nearest Neighbor* dengan menggunakan *dataset* ulasan layanan daring aplikasi *SIP* Disdukcapil Kabupaten Jember. Pada penelitian ini model dapat dikatakan lebih baik apabila model dapat melakukan klasifikasi pada sentimen negatif dengan baik. Data penelitian yang digunakan adalah *dataset* ulasan pengguna pada tahun 2024. Hasil dari penelitian ini adalah dapat membandingkan kinerja kedua model tersebut dengan membandingkan kinerjanya berdasarkan hasil *Recall*, *F2-Score*, dan *Confusion Matrix*. Ketiga matrik tersebut dipilih dikarenakan memiliki kemampuan untuk mengukur ketepatan proses klasifikasi terhadap data aktual, sehingga tetap sejalan dengan tujuan penelitian dikarenakan kehilangan sentimen negatif dapat berpotensi mengaburkan permasalahan layanan yang nyata.

1.2 Rumusan Masalah

Berdasarkan latar belakang telah diuraikan sebelumnya, peneliti merinci rumusan masalah sebagai berikut:

1. Bagaimana melakukan Analisis Sentimen terhadap data ulasan pengguna aplikasi *SIP* Disdukcapil Kabupaten Jember?
2. Bagaimana cara menerapkan model algoritma klasifikasi *Naïve Bayes* dan algoritma klasifikasi *K-Nearest Neighbor*?
3. Bagaimana cara mengetahui model algoritma klasifikasi mana yang lebih baik dalam melakukan klasifikasi sentimen negatif berdasarkan nilai *Recall*, *F2-Score*, dan *Confusion Matrix*?

1.3 Tujuan Penelitian

Berdasarkan masalah yang telah dirumuskan penelitian memiliki tujuan sebagai berikut:

1. Melakukan Analisis Sentimen layanan daring aplikasi *SIP* Disdukcapil Kabupaten Jember agar mendapatkan informasi terkait kualitas layanan, khususnya pada ulasan negatif yang berisikan kritik atau saran yang sangat penting guna proses evaluasi layanan.
2. Menerapkan model algoritma klasifikasi *Naïve Bayes* dan algoritma klasifikasi *K-Nearest Neighbor* pada dataset ulasan pengguna aplikasi *SIP* Disdukcapil Kabupaten Jember.
3. Melakukan analisis perbandingan, untuk menentukan hasil algoritma klasifikasi terbaik dalam melakukan klasifikasi kelas negatif, berdasarkan *Recall*, *F2-Score*, dan *Confusion Matrix*.

1.4 Manfaat Penelitian

Berdasarkan uraian diatas adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Mendapatkan informasi-informasi hasil Analisis Sentimen khususnya sentimen negatif pada data ulasan pengguna aplikasi *SIP* Disdukcapil Kabupaten Jember, yang mana proses analisa data secara manual akan sulit diterapkan pada *dataset* yang cukup besar.

2. Mendapatkan hasil proses analisa terbaik dengan membandingkan kinerja model algoritma klasifikasi *Naïve Bayes* dengan *K-Nearest Neighbor* berdasarkan hasil *Recall*, *F2-Score*, dan *Confusion Matrix*.

1.5 Batasan Penelitian

Dalam penelitian ini peneliti menetapkan beberapa batasan masalah dengan uraian sebagai berikut:

1. Model algoritma klasifikasi yang digunakan adalah model algoritma klasifikasi *Naïve Bayes* dan *K-Nearest Neighbor*.
2. Data penelitian yang digunakan adalah data *Indeks Kepuasan Masyarakat* pada aplikasi *SIP* Dinas Kependudukan dan Pencatatan Sipil dengan kurun waktu bulan Januari hingga Desember pada tahun 2024.
3. Proses pelabelan pada *dataset* penelitian dilakukan oleh Yhoni Restian, S.Sos., MM. selaku Kepala Bidang Pengelolaan Informasi Administrasi Kependudukan, Disdukcapil Kabupaten Jember.
4. Metode analisa perbandingan yang digunakan adalah *Recall*, *F2-Score*, dan *Confusion Matrix*.