

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Opini adalah suatu istilah konsep yang luas, mencakup aspek-aspek seperti sentimen, evaluasi, penilaian atau sikap serta informasi mengenai suatu pendapat individu atau kelompok. Sentimen dapat dianalisis dengan memahami, menganalisis dan mengekstraksi suatu pendapat, sehingga dapat mengetahui opini yang diungkapkan tersirat sentimen positif, negatif atau netral [1]. Analisis sentimen dikenal sebagai suatu penambangan opini, yang telah menjadi salah satu bidang penelitian aktif dalam metode bahasa alami. Analisis sentimen memiliki tujuan untuk menafsirkan suatu sistem otomatis yang dapat mengekstrak informasi tekstual dalam bahasa alami, seperti sebuah opini dan sentimen, sehingga menciptakan pengetahuan yang terstruktur dan dapat dimanfaatkan sebagai sistem pendukung pengambilan keputusan [2].

Sentimen dapat dilihat melalui berbagai jenis platform media sosial yang faktanya saat ini siapapun dapat menggunakan media sosial melalui perangkat *smartphone* yang dilengkapi beragam fitur canggih dan komponen pendukung seperti koneksi internet [3]. Media sosial memberi kebebasan untuk memungkinkan pengguna dalam melakukan banyak hal, sehingga dapat memberikan akses untuk pengguna mengumpulkan banyak informasi yang berharga, memberi pendapat, menjalin interaksi dan juga berekspresi tanpa batasan tempat dan waktu, salah satunya yaitu *X* yang merupakan jejaring sosial dengan memberi kebebasan bagi penggunanya untuk dapat melakukan interaksi secara terbuka kepada lembaga maupun masyarakat melalui pesan singkat yang dikenal dengan kicauan atau *tweet* [4].

Sentimen yang dihasilkan melalui *tweet* akan melewati proses analisis dengan tujuan untuk melakukan pengklasifikasian teks. Permasalahan dalam penelitian ini

yaitu mencari algoritma pengklasifikasian yang terbaik untuk *dataset X*, sehingga diperlukan suatu tindakan pengujian atau seleksi untuk mengevaluasi performa pada beberapa algoritma pengklasifikasian [5]. Ada sejumlah pendekatan algoritma untuk melakukan analisis sentimen, diantaranya yaitu *Naïve Bayes*, *SVM* dan *Random Forest*.

Algoritma *Naïve Bayes classifier* atau disebut dengan NBC adalah algoritma mengklasifikasikan probabilitas berdasarkan *Teorema Bayes*. Dalam kasus analisis sentimen algoritma ini bertujuan untuk mengklasifikasi sebuah teks atau dokumen ke dalam kelas positif, negatif maupun netral [2]. Penelitian sebelumnya yang dilakukan oleh Ade Tiara S, Nur Anjeni L, dan Puput Alpria dengan mengangkat topik analisis sentiment public pada *twitter* terhadap boikot produk Israel menggunakan algoritma *Naïve Bayes* dengan 303 *tweets* menghasilkan akurasi dengan nilai *accuracy* 95%, *precision* 96%, *recall* 95% dan *f1-score* 95%, sehingga dapat disimpulkan bahwa tingginya tingkat keberhasilan dalam mengidentifikasi sentimen melalui *tweet* menggunakan algoritma *Naïve Bayes* [6].

Algoritma *Support Vector Machine* atau disebut SVM adalah teknik yang beroperasi bertujuan untuk menentukan *hyperplane* yang optimal dengan memisahkan jarak antara masing-masing titik yang disebut kelas, *Hyperplane* ini berperean sebagai pemisah antar kelas dalam proses klasifikasi [7]. Penelitian sebelumnya yang dilakukan oleh Primandani Arsi dan Retno Waluyo dengan mengangkat topik analisis sentimen wacana pemindahan Ibu Kota Indonesia menggunakan algoritma *Support Vector Machine* (SVM) yang menghasilkan nilai *accuracy* 96,68%, *precision* 95.82%, *recall* 94.04% dan AUC = 0,979 dengan 1.236 *tweets* (404 positif dan 832 negatif), sehingga dapat disimpulkan SVM terbukti mampu diterapkan dengan baik dengan berbagai teknik pemrosesan dan penentuan suatu nilai *hyperparameter* yang tepat [8].

Algoritma *Random Forest*, sesuai dengan namanya merupakan serangkaian pohon (*tree*) yang menyatu membentuk hutan (*forest*) atau algoritma berdasarkan

pohon keputusan dan menerapkan *bagging*. Pohon dibentuk untuk menemukan prediksi serangkaian aturan pemisah. Pohon-pohon tersebut tumbuh dengan menggunakan algoritma *classification and regression trees* atau biasa disebut dengan CART [9]. Penelitian sebelumnya yang dilakukan oleh Thifal Fadiyah B, Dian Eka R, dan Issa Arwani dengan mengangkat topik analisis sentimen pengguna *twitter* terhadap pembayaran *cashless* menggunakan *Shopeepay* dengan algoritma *Random Forest* menghasilkan nilai *accuracy* 95%, *recall* 94%, *precision* 95% dan *f1-score* 95%, sehingga disimpulkan bahwa sistem algoritma *Random Forest* (RF) dapat mengidentifikasi *dataset twitter* dengan efektif dan menghasilkan nilai performa yang baik dengan jumlah *tree* 300 serta kedalaman *tree* 55 [10]

Tujuan dari penelitian ini adalah untuk melakukan klasifikasi sentimen menggunakan konteks Ibu Kota Nusantara (IKN), memanfaatkan tiga algoritma yaitu *Naïve Bayes*, SVM dan *Random Forest*, berdasarkan tiga kelas positif, negatif dan netral. Dengan dilakukan seleksi untuk mencari algoritma pengklasifikasian yang terbaik dengan membandingkan tingkat performa algoritma menggunakan *dataset X* meliputi *accuracy*, *recall*, *precision* dan *f1-Score* yang lebih optimal, serta dengan bantuan pendekatan *text mining* yaitu *machine learning* dan berbasis *lexicon based*, yang terfokus pada analisis kekuatan sentimen. Selanjutnya menerapkan teknik *k-fold cross validation* dan SMOTE serta dilakukan pengujian dengan mengevaluasi kinerja model menggunakan *Confusion Matrix*. Penelitian ini menggunakan framework CRISP-DM untuk memastikan pendekatan terstruktur dan sistematis dalam proses analisis data.

Dari banyaknya opini atau sentimen di media sosial *X* salah satunya yaitu konteks mengenai Ibu Kota Nusantara (IKN) yang memicu beragam tanggapan mendukung dan menentang. Dari pemilihan letak hingga pengesahan Undang-Undang Nomor 3 Tahun 2022 yang mengacu pada Ibu Kota Nusantara (IKN) oleh Pemerintah Republik Indonesia dan DPR RI dianggap terlalu terburu-buru dan belum lama ini pemerintah juga turut mengajak masyarakat untuk ikut serta dalam pengumpulan dana (*Crowdfunding*) dalam proses pembangunan Ibu Kota

Nusantara (IKN) yang sedang berjalan. Perancangan Undang-Undang IKN diketahui juga sangat kurang dalam melibatkan partisipasi warga Indonesia, dimana hanya sebagian kecil warga yang ikut serta, hal tersebut dinilai Undang-Undang Ibu Kota Nusantara (IKN) sangat kurang akan keterlibatan masyarakat Indonesia sehingga tidak sejalan dengan prinsip demokrasi yang menjadi dasar ideologi negara [11].

1.2 Rumusan Masalah

Berikut kesimpulan dari rumusan masalah yang dapat diambil berdasarkan uraian latar belakang diatas yaitu:

1. Bagaimana melakukan perbandingan klasifikasi performa pada tiap algoritma dari tanggapan pengguna media sosial *X* terhadap Ibu Kota Nusantara (IKN) menggunakan algoritma *Naïve Bayes*, *SVM* dan *Random Forest* meliputi *accuracy*, *precision*, *recall* dan *f-1 score*?
2. Algoritma manakah yang memiliki performa terbaik dalam pengklasifikasian data sentimen pengguna media sosial *X*?

1.3 Batasan Masalah

Untuk memastikan penelitian ini terfokus, maka ditetapkan batasan masalah Batasan sebagai berikut:

1. Data diperoleh dari sumber di media sosial *X* melalui *crawling data* dari 2 Februari 2024 hingga 26 April 2024 dengan jumlah data yang 10.000 *tweet*.
2. Hasil data yang akan diklasifikasi hanya 3 yaitu positif, negatif dan netral.
3. Data *tweet* yang diambil sesuai tanggapan pengguna media sosial *X* dengan kata kunci yang memiliki hubungan mengenai Ibu Kota Nusantara (IKN)
4. *Dataset* yang digunakan pada penelitian berbahasa Indonesia.

1.4 Tujuan

Berikut ini adalah tujuan dari penelitian, yaitu sebagai berikut:

1. Menganalisis dan membandingkan performa algoritma *Naïve Bayes*, *SVM* dan *Random Forest* dalam mengklasifikasikan sentimen dari tanggapan pengguna media sosial *X* terhadap Ibu Kota Nusantara (IKN) dengan menggunakan metrik-metrik evaluasi meliputi *accuracy*, *precision*, *recall*, dan *f-1 score*.
2. Menentukan algoritma yang memiliki performa terbaik dalam mengklasifikasikan data sentimen terhadap Ibu Kota Nusantara (IKN) berdasarkan hasil evaluasi menggunakan dataset yang telah dikumpulkan dari media sosial *X*

1.5 Manfaat Penelitian

Adapun manfaat yang didapatkan dari penelitian analisis sentimen ini sebagai berikut:

1. Manfaat Akademik
 - a. Mengembangkan dan memperluas pengetahuan dalam salah satu cabang ilmu pengetahuan, yakni *Machine Learning*, terutama dalam konteks penerapan teks mining menggunakan *Naïve Bayes*, *SVM* dan *Random Forest*, untuk membandingkan performa algoritma dan melakukan klasifikasi sentimen melalui platform media sosial *X*
 - b. Dapat menjadi landasan bagi penelitian lanjutan yang lebih spesifik mengenai aspek tertentu dengan melakukan analisis sentimen dengan membandingkan tiga algoritma yakni, *Naïve Bayes*, *SVM* dan *Random Forest*.
2. Manfaat Praktis
 - a. Untuk memahami tanggapan atau sentimen pengguna *X* khususnya masyarakat Indonesia mengenai Ibu Kota Nusantara (IKN) Indonesia
 - b. Dapat digunakan sebagai kontribusi dalam bahan uji dan evaluasi yang berpotensi untuk memberi wawasan dan juga informasi yang bermanfaat untuk pembuat kebijakan yang dilakukan oleh pemerintah Indonesia terhadap kinerja, analisis dampak sosial dan ekonomi

- c. Membantu dalam menentukan algoritma yang paling akurat untuk analisis sentimen pengguna media sosial X mengenai IKN.

1.6 Sistematika Penulisan

Berikut Sistematika penulisan laporan skripsi ini

BAB 1 PENDAHULUAN

Di bagian ini, terdapat penjelasan terkait konteks umum meliputi latar belakang, rumusan permasalahan dalam penelitian, batasan-batasan masalah yang diterapkan, tujuan penelitian, manfaat akademik dan praktis serta sistematika penulisan yang dapat dilihat pada bagian ini

BAB 2 TINJAUAN PUSTAKA

Di bagian ini, terdapat tinjauan terhadap pustaka (buku atau publikasi) yang berkaitan dengan tema penelitian, teori, paradigma, cara pandang; tinjauan literatur dari penelitian terdahulu yang ada kaitan dengan tema penelitian.

BAB 3 METODOLOGI PENELITIAN

Di bagian ini, terdapat pembahasan mengenai jenis penelitian, metode penelitian, instrumen penelitian, objek penelitian (termasuk profil perusahaan tempat observasi) dan kerangka penelitian (*research framework*).

BAB IV HASIL DAN PEMBAHASAN

Di bagian ini, terdapat hasil dan uraian dari penelitian sesuai dengan metode yang telah ditentukan sebelumnya.

BAB V SIMPULAN DAN SARAN

Di bagian ini, terdapat uraian mengenai simpulan yang mencakup inti dari hasil penelitian berdasarkan perumusan masalah dan tujuan penelitian serta saran yang mencakup rekomendasi untuk pengembangan penelitian lanjutan.