

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media jejaring sosial memberikan peran yang sangat besar bagi perkembangan teknologi khususnya pada teknologi komunikasi dan informasi. Melalui media sosial penyebaran informasi bisa didapatkan dengan mudah dan cepat. Berdasarkan data statistik dari Wearesocial di tahun 2019 pengguna media sosial Indonesia mencapai 150 juta atau sebesar 56% dari total populasi penduduk (Wearesocial, 2019). Twitter merupakan media sosial yang kini tren di masyarakat, Berdasarkan laporan We Are Social dan Hootsuite, ada 556 juta pengguna Twitter di seluruh dunia pada Januari 2023. Jumlah tersebut meningkat 27,4% dibandingkan pada periode yang sama tahun sebelumnya.. Tingginya popularitas media sosial Twitter banyak pengguna yang menggunakan sebagai media untuk mengungkapkan apapun, baik itu apa yang dia rasakan, apa yang sedang dialami ataupun kondisi yang terjadi disekitarnya. Dari data yang terdapat pada Twitter, dapat dimanfaatkan untuk memperoleh informasi tertentu. Hal ini menyebabkan banyak portal berita yang memanfaatkan Twitter sebagai media untuk menyampaikan informasi lalu lintas pada suatu daerah seperti kemacetan maupun kecelakaan.

Kota Surabaya merupakan kota metropolitan kedua di Indonesia setelah Jakarta. Banyaknya jumlah penduduk tidak sedikit permasalahan yang terjadi di Kota Surabaya, seperti yang kita ketahui berita tentang kemacetan lalu lintas, kecelakaan dan tindak criminal sering kita jumpai pada sejumlah media di Kota Surabaya. Angka kecelakaan lalu lintas masih tinggi, terutama di Jawa Timur.

Sepanjang bulan Januari hingga Februari 2023, berdasarkan data Dirlantas Polda Jawa Timur, tercatat ada 1.735 total kejadian kecelakaan lalu lintas, Surabaya sebagai ibu kota Jawa Timur menempati urutan kedua sebagai kota dengan angka kecelakaan lalu lintas terbesar di Jawa Timur. Begitu pula dengan tindak kriminal yang tercatat oleh Polrestabes Surabaya terdapat 939 kejadian sepanjang tahun 2023. Berbagai permasalahan tersebut kerap kita jumpai di Kota Surabaya setiap harinya.

Suara Surabaya merupakan salah satu portal berita yang ada di Surabaya yang secara aktif membagikan informasi lalu lintas area Surabaya dan sekitarnya melalui siaran radio (FM 100), website (suarasurabaya.net), dan juga media sosial Twitter (@e100ss). Informasi public yang di-post melalui media sosial Twitter sangat tinggi, terutama berita kemacetan di wilayah Surabaya hal ini ditunjukkan dengan jumlah rata-rata tweet dalam sehari dari akun @e100ss sejumlah 192,2, sementara jumlah post dalam sehari dari akun E100 sejumlah 126,18 post, yang mana kedua data tersebut belum termasuk pesan dari pendengar atau pengguna yang ditujukan kepada media sosial resmi Radio Suara Surabaya (Putra, 2019). Dengan ini media sosial berperan penting sebagai media penyebar informasi secara real-time. Namun, dengan banyaknya permasalahan yang terjadi di hari yang sama menjadikan informasi yang diberikan tidak tergambar dengan jelas. Hal tersebut terkadang menyulitkan masyarakat untuk memonitor dengan jelas. Data Tweet memuat informasi yang cukup jelas mengenai lokasi, waktu, dan siapa saja yang terlibat. Data tersebut dapat dimanfaatkan untuk menganalisa kejadian kecelakaan lalu lintas, kemacetan dan tindak Kriminal yang terjadi di Kota Surabaya.

Pengumpulan data Tweet dilakukan dengan cara scraping data dari Twitter menggunakan library `sntwitwe`, data yang dikumpulkan berupa isi Tweet serta tanggal dan waktu Tweet dengan rentang waktu September 2022 – Mei 2023. Karena data yang dimuat berbentuk Teks, maka perlu dilakukan pendekatan menggunakan Text Mining. Kirana, melakukan pendekatan Text Mining untuk menganalisis data berupa teks dari tingkat keakuratan serta menghitung potensi-potensi yang muncul dari data tersebut secara otomatis (Kirana, 2019). Abba menggunakan data dari twitter untuk dilakukan klasifikasi menggunakan support vector machine dan menggunakan pendekatan NER agar dapat mengidentifikasi informasi lokasi kejadian bencana. Hasil berupa visualisasi dalam bentuk peta dan terdapat titik lokasi terjadinya bencana (Abba Suganda, 2022). Setio Basuki menggunakan data twitter untuk mengklasifikasikan postingan twitter tindak kejahatan dan kecelakaan menggunakan metode Naïve Bayes (Setio Basuki, 2020). Metode Naïve Bayes dapat menghasilkan akurasi yang maksimal dengan data latih yang sedikit (Riri Nada, 2019). Berdasarkan penelitian-penelitian sebelumnya, maka pada penelitian ini akan menggunakan metode Naive Bayes untuk memprediksi kelas kecelakaan lalu lintas, kemacetan, dan tindak kriminal yang ada di Kota Surabaya berdasarkan data twitter yang diambil dari akun media sosial @e100ss, selain memprediksi kelas kejadian, penelitian ini juga akan mengekstraksi lokasi kejadian untuk ditampilkan di sistem dengan menggunakan pendekatan Name Entity Recognizer (NER).

Hasil klasifikasi dari kelas kecelakaan, kemacetan, dan tindak kriminal yang ada di Kota Surabaya akan divisualisasikan dalam bentuk grafik per hari dan

perbulan dari masing-masing kelas, serta peta persebaran terhadap lokasi yang diekstrak menggunakan metode NER ke OpenStreet Map untuk memvisualkan dalam bentuk peta persebaran. Dengan demikian penelitian ini diharapkan membantu masyarakat mempermudah dalam mendapatkan informasi tentang kemacetan, kecelakaan, dan tindak kriminal di Kota Surabaya serta mengetahui hasil evaluasi model klasifikasi terbaik dalam menentukan kelas kemacetan, kecelakaan, dan tindak kriminal.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas maka diperoleh suatu rumusan masalah yang akan dibahas pada penelitian ini yaitu :

1. Bagaimana memudahkan penyebaran informasi pada akun Twitter Suara Surabaya berdasarkn kelas kemacetan, kecelakan, dan tindak kriminal dengan menggunakan metode naïve bayes, serta
2. Bagaimana memvisualisasikan data tweet akun twitter suara Surabaya dalam bentuk grafik dan peta persebaran

1.3 Batasan Masalah

Agar penelitian ini tidak terlalu melebar dari topik yang sudah ditetapkan, maka dibuat batasan masalah sebagai berikut:

- a. Data Tweet kejadian terbatas di wilayah Surabaya Jawa Timur.
- b. Data yang diambil merupakan *tweet* dari akun Twitter @e100ss berdasarkan kata kunci “kecelakaan”, “kemacetan”, “macet”, “begal”, “copet”, “maling”.
- c. Kelas terbatas atas Kecelakaan lalu lintas, Kemacetan, Tindak kriminal.
- d. Hasil sebatas menentukan model klasifikasi terbaik belum sampai memprediksi data baru

- e. Rentang waktu data yang digunakan mulai dari September 2022 – Mei 2023
- f. Atribut waktu yang digunakan merujuk pada waktu postingan tweet

1.4 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan diatas, maka tujuan dari penelitian ini adalah untuk memudahkan penyebaran informasi berdasarkan kelas kemacetan, kecelakaan, dan tindak kriminal yang ada di Kota Surabaya dengan menggunakan data dari akun Twitter Suara Surabaya. Serta memvisualisasikan kejadian di wilayah tertentu dalam bentuk grafik dan peta persebaran berdasarkan tweet Suara Surabaya.

1.5 Manfaat

Adapun manfaat yang ingin dicapai dan diperoleh dari penelitian ini adalah:

1. Manfaat Praktisi
 - a) Mengetahui model klasifikasi terbaik dalam mengklasifikasikan kejadian darurat yang ada di Surabaya menggunakan data dari Twitter.
 - b) Menampilkan jumlah peristiwa kecelakaan, kemacetan dan tindak kriminal di Surabaya.
2. Manfaat Akademis
 - a) Bagi pengembangan ilmu pengetahuan, dapat memberikan suatu karya peneliti baru yang dapat mendukung dalam pengembangan sistem informasi.
 - b) Bagi peneliti dapat menambah wawasan dengan mengaplikasikan ilmu yang telah diperoleh secara teori di lapangan.

1.6 Sistematika Penulisan

Sistematika penulisan yang menjadi langkah-langkah dalam proses penyusunan skripsi ini berfungsi untuk membantu mencapai tujuan agar tidak menyimpang dari batasan masalah yang ditentukan.

BAB I PENDAHULUAN

Pada bab ini berisi mengenai latar belakang, rumusan masalah, batasan masalah, tujuan, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini akan dijelaskan tentang penelitian yang telah dilakukan sebelumnya, teori-teori mengenai twitter, Suara Surabaya, text mining, TF-IDF, *Named Entity Recognition*, dan Naïve Bayes.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang metode penelitian yang akan dibuat meliputi tahap pengumpulan data, analisis kebutuhan, pembuatan model.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil dari setiap tahapan yang ada pada metodologi penelitian serta pembahasan hasil dari pembuatan model klasifikasi tweet pada akun twitter Suara Surabaya @e100ss.

BAB V KESIMPULAN DAN SARAN

Pada Bab ini diuraikan mengenai kesimpulan yang diambil dari keseluruhan isi dari laporan serta saran yang disampaikan untuk penelitian selanjutnya.

DAFTAR PUSTAKA

Pada bagian ini akan dipaparkan tentang sumber-sumber literature yang digunakan dalam pembuatan laporan skripsi ini. Berisi tentang penjelasan dan pembahasan pendukung skripsi guna memahami isi laporan.