

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, data pribadi menjadi aset yang sangat berharga. Masyarakat *modern* semakin bergantung pada *internet* dan teknologi untuk berbagai aktivitas sehari-hari. Namun dibalik kenyamanan tersebut muncul tantangan besar terkait keamanan data pribadi. Kebocoran data telah menjadi isu yang sering diperbincangkan. Salah satu kebocoran data yang diduga terjadi pada tahun 2024 melibatkan Kementerian Komunikasi dan Informatika (Kominfo) Indonesia. Dilaporkan bahwa data sensitif yang diduga milik Kominfo telah dibeli dengan harga Rp 1,9 miliar di *BreachForums* [1]. Hal ini menimbulkan berbagai reaksi dari masyarakat, yang diungkapkan melalui media sosial seperti *Twitter*.

Twitter telah menjadi platform bagi masyarakat Indonesia untuk menanggapi dan berbagi pendapatnya mengenai isu terkini. Dengan kemudahan akses jangkauannya yang luas, Twitter memungkinkan pengguna untuk menyampaikan opini dengan cepat [2]. Tagar sering digunakan untuk membangun diskusi dan memulai percakapan yang lebih luas. Sehingga memungkinkan terbentuknya opini publik.

Oleh karena itu, analisis sentimen mengenai kebocoran data di Twitter perlu dilakukan untuk memahami bagaimana reaksi masyarakat terhadap kasus kebocoran data Kominfo pada pengguna media sosial Twitter. Analisis sentimen adalah proses yang bertujuan untuk mengidentifikasi sentimen dalam data teks, yang kemudian dikategorikan ke dalam tiga kelompok yaitu positif, negatif dan netral [3]. Dalam konteks ini, analisis sentimen terhadap kebocoran data Kominfo memberikan gambaran yang jelas mengenai pandangan masyarakat terhadap isu ini, agar memperkuat kebijakan perlindungan data.

Pada Penelitian sebelumnya menunjukkan berbagai model klasifikasi analisis sentimen. Penelitian pada tahun 2022 menggunakan metode Naïve Bayes terhadap sentimen Kasus kebocoran data *Facebook* dan *Cambridge Analytica* pada

media sosial Twitter menghasilkan akurasi sebesar 83.06% [4]. Penelitian tahun 2020 menggunakan metode Naïve Bayes, Random Forest, dan Support Vector Machine terhadap aplikasi Ruangguru dengan hasil akurasi terbesar dimiliki Random Forest sebesar 97.16% [5]. Penelitian tahun 2022 menggunakan metode Logistic Regression terhadap Covid-19 melalui media sosial Twitter menghasilkan akurasi sebesar 82% [6].

Penelitian ini menggunakan Model Klasifikasi seperti, Naïve Bayes, Random Forest, dan Logistic Regression. Setiap model akan dievaluasi berdasarkan tingkat akurasi yang dihasilkan dalam mengklasifikasikan data. Akurasi menjadi fokus utama untuk menentukan sejauh mana model mampu memprediksi sentimen dengan tepat. Hasil akurasi dari setiap model akan dibandingkan untuk menentukan model yang paling optimal, dengan harapan dapat memberikan gambaran yang paling akurat mengenai sentimen publik terhadap kebocoran data Kominfo.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan diatas, maka rumusan masalah pada penelitian ini, yaitu:

1. Berapa nilai akurasi tertinggi dari model klasifikasi terhadap analisis sentimen publik berdasarkan data tweet kebocoran data Kominfo?
2. Bagaimana sentiment masyarakat terhadap kebocoran data Kominfo pada Twitter?

1.3 Batasan Masalah

1. Data Tweet yang digunakan berbahasa Indonesia
2. Menggunakan Bahasa pemrograman Python
3. Dataset diperoleh melalui proses scraping pada Twitter
4. Sentimen pada tweet dikategorikan kedalam tiga kelompok yaitu positif, negatif, dan netral
5. Nilai performa model algoritma yang dihitung adalah akurasi

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah untuk

menganalisis sentimen publik terhadap kebocoran data Kominfo berdasarkan data Twitter dan untuk mengetahui model klasifikasi yang memberikan akurasi tertinggi dalam sentimen publik di Twitter terhadap Kasus kebocoran data Kominfo.

1.5 Manfaat Penelitian

1. Memberikan gambaran mengenai opini publik terhadap kebocoran data Kominfo yang menjadi masukan dalam terhadap sistem keamanan data.
2. Memberikan wawasan yang diharapkan dapat digunakan untuk pengambilan kebijakan terkait perlindungan data

1.6 Sistematika Penulisan

Berisi sistematika penulisan yang memuat uraian secara garis besar isi untuk tiap-tiap bab. Berikut ini penulisannya :

BAB I PENDAHULUAN, bab ini menjelaskan mengenai latar belakang masalah, rumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA, bab ini berisi mengenai studi literatur, dan dasar-dasar teori yang digunakan untuk mendukung penelitian.

BAB III METODE PENELITIAN, bab ini berisi mengenai metode pengumpulan data, metode analisis, dan alur penelitian.

BAB IV HASIL DAN PEMBAHASAN, bab ini berisi mengenai penjelasan hasil penelitian dan pembahasan

BAB V PENUTUP, bab ini berisi kesimpulan dan saran untuk penelitian selanjutnya.