

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Indonesia dengan jumlah penduduk yang luas menyebabkan penggunaan internet berkembang secara dinamis dan meluas sangat pesat dengan dibuktikannya pemanfaatan internet di berbagai provider mengalami peningkatan dalam penggunaan kapasitas jaringan internet tertentu. Penggunaan internet ini digunakan dalam berbagai hal seperti dalam berbelanja, belajar bahkan memberikan pendapat diberbagai platform salah satunya dalam penggunaan media sosial yang semakin diminati oleh masyarakat untuk menyampaikan informasi. Diketahui bahwa media sosial merupakan media yang digunakan oleh pengguna untuk berbagi data dalam bentuk teks, gambar, suara dan lainnya. Berkembangnya media sosial membuat masyarakat lebih aktif untuk mengunggah segala sesuatu yang terjadi di lingkungan (Arifianto 2020)

Twitter adalah sebuah platform media sosial dan layanan *microblogging* yang memungkinkan penggunanya untuk mengirim dan membaca pesan teks secara real-time (Hadna, Paulus, and Winarno 2016). Twitter menjadi unggul sebagai media penyalur informasi tercepat karena menyajikan berbagai topik yang sedang tren dan menarik, sehingga memunculkan banyak opini dan komentar dari publik. Hal ini membuat Twitter potensial sebagai sumber data yang efisien untuk analisis sentimen. Analisis sentimen adalah studi komputasi yang bertujuan untuk memahami, mengekstrak, dan memproses data teks untuk mendapatkan informasi tentang sentimen dalam suatu kalimat. Algoritma *Naïve Bayes* adalah salah satu algoritma yang banyak digunakan untuk klasifikasi data teks dengan performa yang baik. Model *Bernoulli* dalam *Naïve Bayes* menggunakan informasi kejadian biner, di mana nilai 0 dan 1 mengindikasikan kehadiran atau ketiadaan suatu kata dalam dokumen (Putri, Ir. Mohammad Masjkur, M.S., and Indahwati 2023).

Korupsi telah lama menjadi permasalahan serius di Indonesia yang mempengaruhi kepercayaan publik terhadap pemerintah dan lembaga negara. Dalam era digital dan perkembangan media sosial, masyarakat memiliki akses lebih mudah untuk menyuarakan pendapat dan perasaan mereka terkait dengan isu korupsi (Putra and Linda 2022). Situs X, sebagai salah satu platform digital yang

populer di Indonesia, menjadi tempat di mana berbagai pendapat, opini, dan pandangan mengenai korupsi dapat diekspresikan secara luas dan real-time. Fenomena ini tidak hanya mempengaruhi pertumbuhan ekonomi negara, tetapi juga merusak kepercayaan publik terhadap institusi pemerintahan. Seiring dengan perkembangan teknologi dan internet, masyarakat kini memiliki akses lebih mudah untuk berbagi pandangan dan pendapat mereka tentang topik korupsi melalui platform daring, termasuk situs X.

Aparatur Sipil Negara (ASN) memiliki tanggung jawab strategis untuk menjaga integritas dan mencegah praktik korupsi (Putra and Linda 2022). Dalam menghadapi tantangan ini, ASN memiliki tanggung jawab moral dan profesional untuk bekerja sama dalam upaya pencegahan dan penindakan korupsi.

Analisis sentimen merupakan pendekatan komputasional yang dapat digunakan untuk menggali dan memahami perasaan serta opini masyarakat terhadap isu korupsi (Syakir and Hasan 2023). Dengan memanfaatkan algoritma *Bernoulli Naive Bayes*, metode ini dapat mengklasifikasikan teks-teks yang ditemukan di situs X ke dalam kategori sentimen positif atau negatif terkait dengan korupsi.

Penerapan algoritma *Bernoulli Naive Bayes* dalam konteks ini relevan karena model ini mampu mengatasi kelemahan data teks yang sparse, di mana setiap dokumen dianggap sebagai himpunan kata-kata yang diwakili oleh variabel biner (0 atau 1) (Putri et al. 2023). Terbukti pada penelitian yang dilakukan oleh (Saraswati dan Riminarsih 2020) dengan judul Analisis Sentimen Terhadap Pelayanan KRL *Commuterline* Berdasarkan Data Twitter Menggunakan Algoritma *Bernoulli Naive Bayes* dan penelitian oleh (Dewi, Jondri, and Lhaksmana 2021) dengan judul Prediksi *Retweet* Menggunakan Metode *Bernoulli Naive Bayes* dan *Gaussian Naive Bayes* di Media Sosial Twitter Dengan Topik Vaksinasi Covid 19. Hal ini memungkinkan untuk mengidentifikasi pola-pola penting dalam teks yang berkaitan dengan perasaan masyarakat terhadap korupsi.

Penelitian ini bertujuan untuk memberikan wawasan mendalam tentang bagaimana opini publik tentang korupsi diekspresikan di platform digital, serta untuk mengukur efektivitas algoritma *Bernoulli Naive Bayes* dalam melakukan analisis sentimen terhadap topik ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Berapa jumlah sentimen yang bernilai positif dan negatif pada klasifikasi analisis sentimen topik Korupsi di Indonesia?
2. Berapa akurasi, presisi dan *recall* yang diperoleh pada analisis sentimen topik Korupsi di Indonesia?

1.3 Tujuan penelitian

Tujuan dari penelitian ini adalah:

1. Mengidentifikasi dan menghitung jumlah sentimen positif dan negatif dalam analisis sentimen topik Korupsi di Indonesia
2. Mengevaluasi performa model analisis sentimen topik Korupsi di Indonesia berdasarkan akurasi, presisi, dan *recall*

1.4 Manfaat Penelitian

Manfaat penelitian ini adalah:

1. Bagi peneliti, sebagai bahan penyusunan tugas akhir dan juga sebagai bahan untuk mengembangkan pengetahuan yang diperoleh selama studi di perguruan tinggi.
2. Penelitian ini diharapkan dapat bermanfaat bagi peneliti lain dalam bidang analisis sentimen.

1.5 Batasan Masalah

Berdasarkan rumusan masalah yang telah di tetapkan, maka Batasan masalah penelitian ini sebagai berikut:

1. Media Sosial yang digunakan adalah X.
2. Pengambilan data dilakukan berdasarkan kata kunci korupsi Indonesia.
3. Data yang dipakai ada 504 data yang didapat dari scrapping di situs X.
4. Data yang dipakai hanya komentar yang menggunakan Bahasa Indonesia
5. Tools yang digunakan Bahasa Pemrograman Phyton