

BAB I

PENDAHULUAN

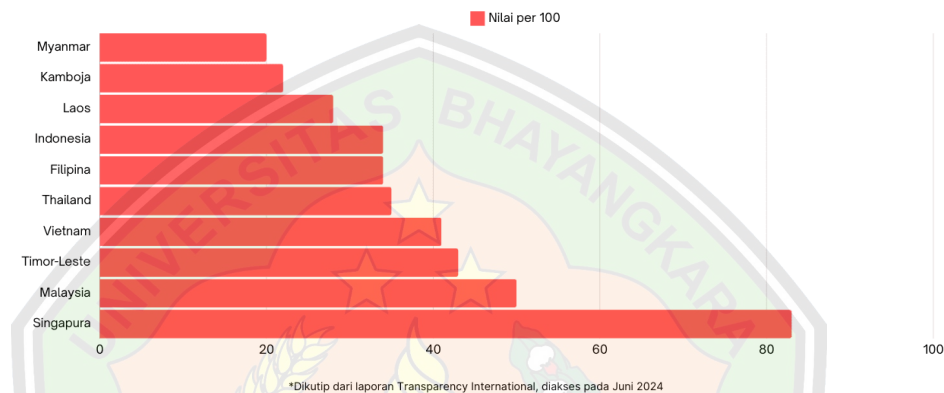
1.1 Latar Belakang

Korupsi merupakan fenomena yang sangat merusak tatanan pemerintahan dan kehidupan masyarakat. Perilaku ini melibatkan penyalahgunaan kekuasaan atau wewenang untuk kepentingan pribadi maupun kelompok, dan dampaknya sangat luas terhadap pembangunan sosial, politik, maupun ekonomi suatu negara. Di negara berkembang, korupsi sering menjadi penyebab utama masalah struktural seperti ketimpangan sosial dan kemiskinan. Di Indonesia, praktik ini sudah lama terjadi dan terus menjangkiti berbagai sektor, termasuk pemerintahan pusat, daerah, serta BUMN.

Tingkat korupsi di suatu negara secara global diukur menggunakan *Corruption Perceptions Index (CPI)* yang dirilis oleh *Transparency International*. Indeks ini menilai tingkat persepsi terhadap tingkat korupsi sektor publiknya, menggunakan rentang skor dari 0 hingga 100, di mana skor 0 menunjukkan tingkat korupsi yang sangat tinggi dan skor 100 menunjukkan sektor publik yang dianggap sangat bersih dari korupsi. *CPI* mencakup penilaian terhadap 180 negara di seluruh dunia [1]. Dalam beberapa tahun terakhir, skor Indonesia dalam *CPI* menunjukkan tren yang cenderung stagnan, bahkan mengalami penurunan, yang mencerminkan masih besarnya tantangan dalam upaya pemberantasan korupsi di tanah air. Jika dibandingkan dengan negara-negara tetangga di kawasan Asia Tenggara seperti Singapura, Malaysia, dan Vietnam, posisi Indonesia masih tertinggal cukup jauh.

Berdasarkan data pada Gambar 1.1, Indonesia bahkan termasuk dalam lima negara terkorup di ASEAN, menempati urutan keempat dari sepuluh negara ASEAN yang paling korup.

Negara Terkorup di Asia Tenggara



Gambar 1. 1 Posisi Indonesia Dalam Negara Terkorup di Asia Tenggara
Sumber: *Transparency International* [2]

Komisi Pemberantasan Korupsi (KPK) berperan penting dalam mengungkap berbagai kasus besar, praktik korupsi masih merambah institusi strategis, termasuk BUMN yang sejatinya memiliki tanggung jawab besar dalam mendukung perekonomian nasional [3]. PT Pertamina, sebagai perusahaan energi milik negara yang bertanggung jawab atas pengelolaan minyak dan gas nasional, kerap menjadi sorotan dalam berbagai dugaan kasus korupsi. Dalam beberapa dekade terakhir, Pertamina menghadapi beragam tuduhan penyimpangan keuangan yang mencerminkan lemahnya tata kelola, kurangnya transparansi, serta pengawasan internal dan eksternal yang belum optimal.

Korupsi di Pertamina tidak hanya mencoreng citra perusahaan sebagai BUMN strategis, tetapi juga berdampak pada stabilitas ekonomi dan kesejahteraan masyarakat. Di sisi lain, perkembangan teknologi dan media sosial seperti Twitter (X) memberi ruang bagi publik untuk mengakses informasi dan menyuarakan opini secara bebas [4]. Salah satu kasus yang tengah menjadi sorotan publik adalah skandal korupsi di anak perusahaan PT Pertamina, yang diduga mengakibatkan kerugian negara hingga Rp193,7 triliun. Skandal ini melibatkan Direktur Utama dan mencakup berbagai pelanggaran yang terjadi sejak tahun 2020 hingga Februari 2025. Kasus tersebut semakin memperkuat perhatian publik terhadap krisis integritas di tubuh perusahaan, sebagaimana ditampilkan pada Gambar 1.2 dibawah ini:



Kasus Korupsi di Lingkungan Pertamina Indonesia (2020-2025)			
Tahun	Tersangka	Kerugian Negara	Modus
2020	Andriyanto	Rp4 miliar	Penyalahgunaan pengelolaan cash card
2021	Betty Halim	Rp1,4 triliun	Tindak pidana pencucian uang dana pensiun (Dapen) PT Pertamina
2022	Dedi Susanto	Rp8,1 miliar	Proyek fiktif
2023	Karen Agustawan	Rp1,77 triliun	Korupsi pengadaan gas alam cair (LNG) di PT Pertamina
2024	Luhur Budi Djatmiko	Rp348,6 miliar	Korupsi pembelian tanah di Rasuna Epicentrum, Jakarta Selatan
2025	Riva Siahaan	Rp193,7 triliun	Menurunkan produksi minyak domestik dan mencampur impor minyak RON 90 dengan kualitas lebih rendah menjadi RON 92

Sumber: KPK dan Polda

GoodStats

Gambar 1. 2 Kasus Korupsi di Lingkungan Pertamina Indonesia
Sumber: KPK dan Polda [5]

Dalam kasus korupsi yang melibatkan Pertamina, Kejaksaan menemukan indikasi penyalahgunaan wewenang antara pejabat *subholding* Pertamina dan kontraktor kontrak kerja sama (KKKS). Diduga, tiga direktur subholding secara

sengaja mengatur rapat optimasi hilir agar Pertamina tidak menyerap minyak mentah domestik. Akibatnya, KKKS mendapat keuntungan besar dari ekspor, sementara Pertamina harus mengimpor minyak dengan harga lebih tinggi. Selain itu, juga ditemukan dugaan praktik pengoplosan BBM dalam kasus ini [6].

Naïve Bayes dan *SVM* dipilih sebagai algoritma yang akan dibandingkan dalam analisis teks pada penelitian kasus korupsi di PT Pertamina karena keduanya memiliki karakteristik dan keunggulan berbeda dalam klasifikasi data teks. *Naïve Bayes* mengklasifikasikan data berdasarkan perhitungan probabilitas dengan asumsi bahwa setiap fitur bersifat independen. Metode ini menghitung probabilitas gabungan dari masing-masing fitur secara terpisah, sehingga membuat proses klasifikasi lebih sederhana dan efisien. Algoritma ini sangat cocok untuk data berukuran besar dan banyak digunakan dalam analisis teks seperti deteksi spam dan analisis sentimen, karena kesederhanaan serta kecepatan pemrosesannya [7]. Sementara itu, *Support Vector Machine (SVM)* unggul dalam menangani data berdimensi tinggi dan mampu membentuk model akurat untuk kasus linier maupun non-linier. *SVM* bekerja dengan mencari *hyperplane* optimal yang memisahkan data berdasarkan kelasnya. Berkat keandalannya, *SVM* banyak digunakan dalam pengenalan pola dan analisis teks [8].

Pada penelitian sebelumnya, menurut Faradilla Izzah Rahmadani, Rizki Rahmawati [9] yang membahas tentang analisis sentimen pengguna Twitter (X) dengan membandingkan tiga algoritma, yaitu *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine*. Hasil analisis menunjukkan bahwa sentimen positif sebesar 16,3%, negatif 52,3%, dan netral 31,4%. Dari ketiga metode yang digunakan, algoritma *SVM*

menghasilkan akurasi tertinggi, yaitu sebesar 72%. Selain itu, penelitian juga telah dilakukan oleh Bintang Adhiyaksa Maulana ,Rifqi Arul Fauzi ,Rissa Ilmia Agustin ,Siti Alia Azhaar ,Tatang Rohana [10] menggunakan algoritma *Naïve Bayes* dan *SVM* untuk menganalisis sentimen Twitter (X) pada korupsi bansos pada masa pandemi. Diperoleh akurasi sebesar 60,61% untuk *Naïve Bayes* dan 66,67% untuk *Support Vector Machine (SVM)*. Hasil ini menunjukkan bahwa *SVM* memiliki performa yang lebih baik. Selain itu, analisis sentimen menunjukkan 41,21% data bersentimen positif dan 58,79% bersentimen negatif.

Penelitian ini menunjukkan bahwa algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* mampu mengklasifikasikan sentimen dalam data teks dengan baik. Fokus utama penelitian adalah mengklasifikasikan sentimen terkait kasus korupsi PT Pertamina ke dalam tiga kategori: positif, negatif, dan netral, menggunakan kedua algoritma tersebut. Pemilihan algoritma terbaik dilakukan dengan membandingkan kinerja berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-Score* terhadap data yang diperoleh dari Twitter (X). Pendekatan yang digunakan menggabungkan teknik *text mining* berbasis *machine learning* dan *lexicon-based*, serta menggunakan *SMOTE* untuk menangani data yang tidak seimbang. Evaluasi model dilakukan menggunakan *Confusion Matrix*, dan seluruh proses dianalisis secara sistematis melalui pendekatan *Knowledge Discovery in Database (KDD)*.

Berdasarkan latar belakang yang telah diuraikan, peneliti memutuskan untuk melakukan penelitian dengan judul **“Perbandingan Algoritma *Naïve Bayes* dan**

***Support Vector Machine* Pada Analisis Sentimen Komentar Pengguna Twitter (X) Terhadap Korupsi Pada PT Pertamina”.**

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan yang dapat diidentifikasi adalah:

1. Belum diketahuinya algoritma mana yang memberikan hasil akurasi paling tinggi antara metode *Naïve Bayes* dan *Support Vector Machine* terkait kasus korupsi PT Pertamina pada komentar pengguna Twitter (X).
2. Diperlukan metode otomatis untuk mengklasifikasi sentimen dari komentar pengguna media sosial Twitter (X) pada kasus korupsi PT Pertamina menggunakan algoritma *Naïve Bayes* dan *SVM*.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah tersebut, permasalahan yang dapat dirumuskan adalah:

1. Bagaimana perbandingan akurasi dan performa pada metode *Naïve Bayes* dan *Support Vector Machine* pada kasus korupsi PT Pertamina terhadap komentar pengguna Twitter (X)?
2. Bagaimana cara mengklasifikasikan sentimen pengguna Twitter (X) terhadap kasus pada PT Pertamina menggunakan algoritma *Naïve Bayes* dan *SVM*?

1.4 Batasan Masalah

Berdasarkan masalah yang telah dirumuskan, maka permasalahan dibatasi pada hal-hal berikut:

1. Metode klasifikasi yang digunakan hanya *Naïve Bayes* dan *Support Vector Machine*.
2. Hanya menggunakan media sosial Twitter (X) untuk melakukan klasifikasi terhadap kasus korupsi pada PT Pertamina.
3. Pemrosesan data, pemodelan data, dan pengevaluasian dilakukan menggunakan *tools Google Colaboratory* dengan bahasa pemrograman *Phyton*.
4. Data yang dikumpulkan dalam penelitian ini adalah *tweet* tentang korupsi pada PT Pertamina pada bulan Februari sampai dengan Maret 2025 dengan jumlah data 1,946 *tweets*.
5. Tahapan metode yang digunakan adalah KDD
6. Pengklasifikasian sentimen hanya dilakukan pada kelas data positif, netral dan negatif.
7. Data yang digunakan merupakan data komentar berbahasa Indonesia.

1.5 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah:

1. Untuk membandingkan akurasi algoritma mana yang memiliki performa terbaik dalam pengklasifikasian sentimen terhadap kasus korupsi di PT Pertamina, berdasarkan analisis hasil evaluasi dari *dataset* yang diperoleh melalui Twitter (X).
2. Membangun model klasifikasi sentimen untuk menganalisis komentar pengguna Twitter (X) menggunakan algoritma *Naïve Bayes* dan *SVM*

1.6 Manfaat Penelitian

Dengan tercapainya tujuan penelitian, maka penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Akademik

- a. Dapat memperdalam pemahaman di bidang *Machine Learning*, khususnya dalam penerapan *text mining* menggunakan algoritma *Naïve Bayes* dan *SVM* untuk membandingkan kinerja masing-masing dalam klasifikasi sentimen pada media sosial Twitter (X) pada kasus korupsi PT Pertamina.
- b. Memberikan kontribusi sebagai referensi pada penelitian selanjutnya, khususnya dalam analisis sentimen kasus korupsi Pertamina dengan pendekatan perbandingan dua algoritma tersebut: *Naïve Bayes* dan *SVM*.

2. Manfaat Praktis

- a. Memberikan gambaran umum terhadap persepsi atau opini masyarakat tentang isu korupsi yang terjadi di PT Pertamina berdasarkan data yang diambil secara real-time dari Twitter (X).
- b. Memberikan kontribusi dalam bentuk bahan uji dan evaluasi yang dapat digunakan untuk memberikan wawasan serta informasi yang bermanfaat bagi pembuat kebijakan dalam menilai dampak sosial dan ekonomi dari kasus korupsi Pertamina ini.
- c. Memberikan insight awal terhadap reaksi masyarakat yang dapat dijadikan indikator dalam mengukur reputasi institusi terkait.

1.7 Sistematika Penulisan

Sistematika penulisan dibagi secara garis besar dalam beberapa bab, penjelasan masing-masing bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, identifikasi masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Berisi tentang berbagai teori-teori, konsep, metode, sistem, atau analisis dan pustaka ilmiah yang berkaitan dengan permasalahan yang sedang diteliti, serta membahas terkait penelitian-penelitian terdahulu yang relevan dengan topik penelitian ini.

BAB III METODOLOGI PENELITIAN

Merupakan bab yang menjelaskan tentang objek serta metode penelitian yang digunakan dengan penggambaran menggunakan kerangka penelitian dan terdapat penjelasan awal dari setiap tahapan-tahapan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjabarkan hasil analisis sentimen yang telah dilakukan menggunakan metode *Naive Bayes* dan *SVM* terhadap data yang telah dikumpulkan dari media sosial Twitter (X) pada kasus Korupsi PT Pertamina.

BAB V PENUTUP

Bab terakhir menjelaskan tentang rangkuman dari penelitian yang telah dilakukan dengan memberikan kesimpulan dari hasil analisis serta saran-saran untuk penelitian selanjutnya.

