

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perpajakan merupakan sumber utama penerimaan negara yang memiliki peran penting dalam membiayai pembangunan nasional serta mendukung penyelenggaraan pemerintahan. Upaya peningkatan penerimaan pajak tidak hanya bergantung pada regulasi yang diterapkan, tetapi juga pada tingkat kepatuhan wajib pajak dalam menjalankan kewajibannya. Untuk mendukung peningkatan kepatuhan sekaligus menyempurnakan sistem administrasi perpajakan, Direktorat Jenderal Pajak (JP) mengembangkan Coretax sebagai bentuk modernisasi pengelolaan perpajakan [1]. Sistem ini dirancang untuk menyederhanakan proses administrasi, termasuk pembuatan faktur pajak dan pelaporan Surat Pemberitahuan Tahunan (SPT) melalui mekanisme yang terintegrasi. Implementasi Coretax diharapkan mampu meningkatkan efisiensi administrasi perpajakan, mendukung integrasi layanan, serta memperkuat pengawasan dan kepatuhan wajib pajak melalui sistem yang lebih terintegrasi. Meskipun demikian, implementasi Coretax di lapangan masih menghadapi berbagai kendala teknis, seperti gangguan sistem, kegagalan penyimpanan data faktur, serta keterbatasan akses pada waktu tertentu. Permasalahan tersebut tidak hanya menghambat pemenuhan kewajiban perpajakan, tetapi juga memunculkan berbagai tanggapan dari pengguna. Di satu sisi, sistem ini memperoleh apresiasi sebagai upaya modernisasi layanan perpajakan, sedangkan di sisi lain masih banyak pengguna yang menyampaikan keluhan akibat kendala teknis yang dialami [2], [3]. Beragam tanggapan tersebut banyak disampaikan melalui media sosial yang kini berfungsi sebagai ruang publik digital untuk menyampaikan opini, kritik, maupun saran terhadap layanan pemerintah [4]. Kondisi tersebut menunjukkan bahwa persepsi pengguna terhadap Coretax bersifat beragam sehingga diperlukan analisis sentimen untuk mengetahui kecenderungan persepsi tersebut secara lebih mendalam. Analisis sentimen merupakan teknik pengolahan teks yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi yang terkandung dalam data digital ke dalam kategori tertentu, seperti positif dan negatif [5].

Beberapa penelitian sebelumnya telah menggunakan analisis sentimen untuk mengetahui pendapat pengguna terhadap layanan digital dan aplikasi publik. Menganalisis sentimen opini masyarakat terhadap suatu isu di platform X (Twitter) menggunakan metode *Bernoulli Naive Bayes* akurasi sebesar 78% [6]. Meskipun demikian, nilai *recall* pada kelas

positif hanya mencapai 0,677 lebih rendah dibandingkan kelas negatif yang sebesar 0,856. Penelitian lain juga menganalisis sentimen ulasan Coretax di Twitter menggunakan *Naive Bayes* dan SVM dengan akurasi masing-masing sebesar 92% [4]. Meskipun angka tersebut tergolong tinggi, hasil evaluasi menunjukkan bahwa model masih kurang optimal dalam mendeteksi sentimen positif, terlihat dari nilai *recall* yang hanya mencapai 80% pada *Naive Bayes* dan 88% pada SVM. Berbeda dengan pendekatan konvensional tersebut, Penelitian oleh Rohman dan Vrawati [7] menggunakan pendekatan *deep learning* untuk menganalisis sentimen terhadap isu Coretax di platform X dengan model IndoBERT. Berdasarkan hasil pengujian, model tersebut mencapai nilai *recall* sebesar 90,17%, dengan akurasi 88,89%. Capaian *recall* yang tinggi menunjukkan bahwa IndoBERT mampu mengidentifikasi kelas sentimen secara akurat serta memahami konteks bahasa yang digunakan dalam percakapan di media sosial. Hal serupa juga disampaikan oleh penelitian Sumartha [8]. Model IndoBERT terbukti efektif dalam memahami konteks bahasa Indonesia pada teks media sosial, sehingga mampu menghasilkan klasifikasi sentimen yang lebih baik dibandingkan pendekatan tradisional. Hal ini ditunjukkan oleh hasil pengujian yang mencapai akurasi sebesar 98%, dengan nilai *precision* hingga 0,99 dan *recall* 0,99 pada kelas mayoritas (negatif). Namun, pada kelas minoritas (positif), model menunjukkan keterbatasan dengan *recall* hanya 0,29 dan F1-score sebesar 0,45, yang mengindikasikan adanya pengaruh ketidakseimbangan data (*class imbalance*) dalam proses klasifikasi.

Berdasarkan permasalahan yang telah diuraikan, penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap implementasi sistem Coretax dengan menggunakan model IndoBERT berdasarkan data yang diperoleh dari media sosial TikTok. TikTok merupakan platform media sosial dengan tingkat interaksi pengguna yang tinggi, yang memungkinkan pengguna untuk menyampaikan kritik sosial dan pendapat terhadap berbagai isu publik melalui konten video serta fitur interaksi yang tersedia, sehingga berpotensi mempengaruhi pembentukan opini publik [9]. Meskipun beberapa penelitian sebelumnya telah melakukan analisis sentimen terhadap sistem Coretax menggunakan data dari platform X (Twitter), pemanfaatan data komentar pengguna pada platform TikTok sebagai sumber analisis sentimen terhadap sistem Coretax masih sangat terbatas. Komentar pengguna pada TikTok mencerminkan berbagai pengalaman, keluhan, serta tanggapan terhadap penggunaan sistem Coretax. Dengan demikian, penelitian ini difokuskan pada analisis sentimen pengguna terhadap sistem Coretax berdasarkan komentar yang terdapat pada aplikasi TikTok. Berdasarkan fokus tersebut, penelitian ini berjudul “**Analisis**

Sentimen Ulasan Pengguna Coretax di Aplikasi TikTok Menggunakan Metode IndoBERT.”

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah Bagaimana kecenderungan sentimen masyarakat yang terkandung dalam teks media sosial TikTok terhadap implementasi sistem Coretax.

1.3 Tujuan

Tujuan penelitian ini adalah Untuk menganalisis kecenderungan sentimen masyarakat terhadap implementasi sistem Coretax berdasarkan data komentar pengguna pada platform TikTok menggunakan model IndoBERT.

1.4 Manfaat

Manfaat dari penelitian ini adalah :

1. Memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang analisis sentimen dan pemrosesan bahasa alami (*Natural Language Processing*) NLP, khususnya dalam penerapan model bahasa berbasis transformer IndoBERT pada data media sosial berbahasa Indonesia.
2. Menjadi referensi bagi penelitian selanjutnya yang membahas analisis sentimen pada layanan publik berbasis digital.

1.5 Ruang Lingkup

Berikut beberapa ruang lingkup dari penelitian ini

1. Data yang digunakan adalah data hasil crawling dari media sosial TikTok yang diambil dari bulan Januari 2025 sampai Maret 2026. Data dikumpulkan dari 18 video TikTok yang membahas sistem Coretax dengan total 13.164 komentar hasil *scraping* yang selanjutnya digunakan sebagai data awal penelitian.
2. Data penelitian berupa komentar berbahasa Indonesia yang dikumpulkan dari media sosial TikTok, yang berisi opini, keluhan, serta respons pengguna terhadap penggunaan sistem Coretax.
3. Proses pelabelan data sentimen dilakukan secara manual oleh peneliti dengan mengelompokkan komentar ke dalam kategori sentimen positif dan sentimen negatif berdasarkan konteks kalimat yang terkandung pada setiap komentar.

4. Output dari analisis sentimen ini adalah sentimen positif dan sentimen negatif.
5. *Keywords* yang digunakan saat *crawling* data meliputi: “Coretax”, “Coretax DJP”, “aplikasi Coretax” dan “sistem Coretax”.
6. Kinerja model dievaluasi dengan menggunakan *confusion matrix* serta pengukuran metrik akurasi, presisi, *recall*, dan *F1-score*.

