

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang sangat pesat telah membawa perubahan besar dalam kehidupan manusia modern. Teknologi kini menjadi bagian yang tidak terpisahkan dari berbagai aktivitas, terutama dalam hal komunikasi dan pertukaran informasi. Salah satu platform yang memegang peran sentral dalam komunikasi digital adalah X (sebelumnya *Twitter*) dan Instagram.

Indonesia termasuk negara dengan pengguna media sosial aktif terbesar di dunia. Berdasarkan laporan *Digital 2025 Global Overview Report* hasil kolaborasi *We Are Social* dan *Meltwater*, terdapat sekitar 143 juta identitas pengguna media sosial aktif di Indonesia pada Januari 2025. Angka tersebut setara dengan 50,2% dari total populasi Indonesia yang mencapai sekitar 285 juta jiwa di awal tahun 2025. Laporan ini memuat data yang diperoleh dari berbagai mitra seperti Ookla, Statista, dan DataReportal, menunjukkan bahwa lebih dari setengah penduduk Indonesia sangat aktif menggunakan berbagai platform media sosial sebagai sarana komunikasi, informasi, dan hiburan [1].

Media sosial X memungkinkan penggunanya untuk berbagi pesan singkat atau tweet hingga 280 karakter. Sejak didirikan pada tahun 2006 oleh Jack Dorsey, Noah Glass, Biz Stone, dan Evan Williams, X terus berkembang dengan berbagai fitur interaktif seperti *like*, *reply*, dan *retweet* [2]. Media sosial Instagram juga semakin populer bisa dijadikan alat ukur tingkat kepuasan masyarakat. Pada era sekarang, cara menyampaikan aspirasi masyarakat mengalami perkembangan yang cukup signifikan. Sebelumnya, penyampaian aspirasi hanya dilakukan melalui demonstrasi, audiensi dengan pemerintah, atau surat menyurat. Kini, dengan kemudahan akses, masyarakat lebih banyak memanfaatkan Instagram sebagai sarana cepat dan murah untuk menyampaikan aspirasi. Baik aspirasi yang ditujukan ke instansi terkait maupun langsung kepada pejabat pemerintah dapat disampaikan melalui fitur *direct message*, postingan, ataupun meninggalkan komentar di postingan terkait [3].

Salah satu isu publik yang banyak dibicarakan di media sosial adalah implementasi *Coretax*, sistem administrasi perpajakan terbaru yang diluncurkan oleh Direktorat Jenderal Pajak (DJP) dan mulai diimplementasikan secara resmi pada awal tahun 2025. Sistem ini dirancang untuk mengintegrasikan seluruh proses bisnis inti administrasi perpajakan, mulai dari pendaftaran wajib pajak,

pelaporan Surat Pemberitahuan (SPT), hingga pembayaran pajak, dengan tujuan meningkatkan efisiensi dan kemudahan bagi wajib pajak serta petugas pajak [4].

Namun, sejak peluncurannya, *Coretax* menghadapi berbagai kendala teknis dan operasional yang menimbulkan keluhan dari para pengguna. Beberapa masalah yang sering dilaporkan meliputi kesulitan akses sistem, sering terjadi *error* atau gangguan teknis, lambatnya respons server, serta ketidaksesuaian data dengan sistem lama yang memaksa pengguna melakukan pemutakhiran data secara manual [5], [6]. Selain itu, kompleksitas proses pelaporan yang sebelumnya tersebar di beberapa platform kini terkonsolidasi dalam *Coretax*, namun belum sepenuhnya berjalan mulus sehingga menimbulkan kebingungan dan hambatan dalam pelaporan pajak. Kondisi ini menyebabkan keterlambatan pelaporan dan potensi denda bagi wajib pajak, sehingga komentar dan keluhan terkait *Coretax* pun banyak bermunculan di media sosial [7].

Untuk memperoleh data komentar yang relevan dan terkini, penelitian ini memanfaatkan beberapa metode pengumpulan data, yaitu input komentar secara manual, penggunaan dataset yang telah tersedia, serta pengambilan data langsung dari media sosial melalui tautan (link) postingan Instagram dan X. Proses pengambilan data dari media sosial dilakukan dengan memanfaatkan layanan Application Programming Interface (API) berbasis web melalui platform Apify. Platform ini memungkinkan proses scraping data komentar dari postingan publik secara otomatis berdasarkan tautan yang diberikan, sehingga data yang diperoleh lebih aktual dan sesuai dengan topik yang diteliti. Pemanfaatan Apify juga membantu mengurangi proses pengumpulan data secara manual yang memerlukan waktu dan tenaga yang lebih besar, serta memungkinkan pengambilan data dalam jumlah yang lebih banyak untuk kebutuhan analisis sentimen.

Namun, komentar-komentar tersebut bersifat tidak terstruktur dan memiliki volume yang sangat besar, sehingga sulit dianalisis secara manual. Hal ini membuat potensi wawasan dari opini publik tidak dapat dimanfaatkan secara optimal dalam evaluasi kebijakan maupun perbaikan sistem layanan perpajakan. Oleh karena itu, diperlukan pendekatan analitik berbasis teknologi untuk menyarikan pola dan tema utama dari komentar-komentar tersebut secara efisien dan objektif.

Salah satu solusi untuk mengatasi permasalahan ini adalah dengan menerapkan teknik pohon keputusan (*Decision Tree*) yang merupakan metode klasifikasi yang membagi data berdasarkan aturan keputusan yang terbentuk dari atribut data. Metode ini menyusun hubungan antar data dalam struktur pohon yang mudah dipahami, sehingga mempermudah analisis dan mendukung pengambilan keputusan. Dalam konteks ini, *Decision Tree* digunakan untuk

mengelompokkan komentar pengguna *Coretax* ke dalam kategori sentimen secara otomatis [8].

Penelitian terbaru menunjukkan bahwa *Decision Tree* banyak digunakan untuk klasifikasi data di berbagai bidang dan mampu menghasilkan akurasi yang tinggi, seperti pada studi prediksi cuaca di Kota Semarang menggunakan algoritma C4.5 yang mencapai akurasi 99,12% [8], pengembangan sistem prediksi siswa terbaik berbasis web[9], serta optimalisasi model *Decision Tree* dengan algoritma genetika untuk meningkatkan akurasi prediksi hasil kelulusan mahasiswa [10]. Selanjutnya, penelitian oleh Maulidiyyah et al. (2024) [11] membandingkan performa *Decision Tree* dan Random Forest dalam klasifikasi data kesehatan, di mana *Decision Tree* berhasil mencapai rata-rata akurasi 93% dalam prediksi penyakit diabetes mellitus. Namun demikian, *Decision Tree* memiliki kelemahan, antara lain rentan terhadap *overfitting*, sensitif terhadap variasi data, serta kurang stabil ketika berhadapan dengan data yang besar dan kompleks[3].

Untuk mengatasi keterbatasan tersebut, penelitian ini mengoptimasi *Decision Tree* dengan algoritma *Adaboost* (*Adaptive Boosting*). *Adaboost* merupakan salah satu metode *ensemble learning* yang menggabungkan beberapa model *Decision Tree* sederhana (*weak learners*) menjadi satu model yang lebih kuat (*strong learner*) sehingga dapat meningkatkan akurasi, mengurangi kesalahan klasifikasi, dan menghasilkan prediksi yang lebih stabil [12].

Beberapa studi sebelumnya membuktikan bahwa kombinasi *Adaboost* dan *Decision Tree* efektif digunakan dalam berbagai bidang, termasuk analisis sentimen, seperti *Penerapan Metode Adaptive Boosting pada Analisis Sentimen* (Faizi, 2023) yang berhasil mengklasifikasikan tweet positif/negatif dengan akurasi sekitar 86,8% [13], serta penelitian *Implementasi Ensemble Learning Adaboost pada Decision Tree dan SVM untuk Klasifikasi SMS* (Rosyidi & Rochmawati, 2024) yang menunjukkan peningkatan akurasi base model setelah diterapkan *Adaboost* [14].

Meskipun demikian, penelitian yang secara khusus mengaplikasikan *Decision Tree* yang dioptimasi dengan *Adaboost* untuk analisis sentimen pada aplikasi perpajakan digital seperti *Coretax* masih sangat terbatas, khususnya dalam konteks Indonesia. Selain itu, belum banyak penelitian yang mengembangkan sistem pengambilan data komentar media sosial secara langsung menggunakan tautan (*link*) postingan tertentu sebagai sumber data, khususnya dengan memanfaatkan kata kunci “*Coretax*” untuk memperoleh komentar yang relevan dan kontekstual. Penelitian ini juga menghadirkan kebaruan dengan

membangun sistem visualisasi kluster berbasis web yang menyajikan hasil analisis sentimen secara interaktif, sehingga dapat membantu pemangku kebijakan dalam memahami pola opini publik secara real-time sebagai dasar evaluasi dan perbaikan layanan perpajakan digital.

Berdasarkan latar belakang permasalahan tersebut, penelitian ini berjudul: “Analisis Sentimen Publik Pada Media Sosial Menggunakan *Decision Tree* dan *Adaptive Boosting*”, dengan tujuan untuk mengelompokkan komentar pengguna *Coretax* serta membangun sistem visualisasi kluster berbasis web sebagai alat bantu evaluasi dan perbaikan layanan.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini mencakup bagaimana proses *preprocessing* dilakukan untuk menyiapkan komentar pengguna *Coretax* agar siap digunakan dalam analisis sentimen, serta bagaimana algoritma *Decision Tree* dapat dioptimasi menggunakan metode *Adaboost* untuk meningkatkan akurasi dan kestabilan dalam mengklasifikasikan komentar berdasarkan sentimen positif, negatif, dan netral.

1.3 Batasan Masalah

1. Data yang digunakan adalah komentar pengguna aplikasi *Coretax* yang berbahasa Indonesia dan diperoleh dari platform media sosial X (sebelumnya *Twitter*) dan Instagram.
2. Analisis komentar hanya menggunakan metode text mining dengan teknik *preprocessing* teks, serta algoritma klasifikasi *Decision Tree* yang dioptimasi menggunakan *Adaboost*.
3. Klasifikasi dilakukan berdasarkan pengelompokan sentimen komentar ke dalam tiga kategori, yaitu negatif, positif, dan netral.
4. Sistem yang dikembangkan berupa aplikasi berbasis web untuk hasil klasifikasi komentar secara interaktif.

1.4 Tujuan Penelitian

1. Melakukan *preprocessing* teks untuk menyiapkan data komentar sebagai dasar analisis sentimen.
2. Menerapkan algoritma *Decision Tree* yang dioptimasi dengan metode *Adaboost* untuk mengklasifikasikan komentar berdasarkan sentimen negatif, positif, dan netral secara lebih akurat.
3. Mengembangkan sistem berbasis web yang dapat menampilkan hasil klasifikasi sentimen komentar secara interaktif sebagai alat bantu evaluasi dan perbaikan layanan.

1.5 Manfaat Penelitian

1.5.1 Teoretis

Penelitian ini memberikan kontribusi terhadap pengembangan literatur di bidang *text mining* dan klasifikasi data teks melalui penerapan algoritma *Decision Tree* yang dioptimasi dengan metode *AdaBoost* dalam pengelompokan sentimen. Hasil penelitian ini dapat dimanfaatkan sebagai referensi dalam pengembangan analisis opini publik dan visualisasi data berbasis web.

1.5.2 Praktis

1. Membantu Direktorat Jenderal Pajak dan pengembang *Coretax* dalam memahami pola komentar pengguna secara cepat dan efisien.
2. Menjadi alat bantu pengambilan keputusan untuk meningkatkan kualitas layanan perpajakan digital berdasarkan hasil analisis sentimen.
3. Memberikan hasil klasifikasi sentimen yang terstruktur sebagai bahan evaluasi terhadap respons dan persepsi pengguna terhadap sistem *Coretax*.