

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sri Mulyani Indrawati sebagai Menteri Keuangan menyampaikan peningkatan jumlah wajib pajak sebanyak 37 juta [1]. Tetapi penerimaan pajak belum optimal jika tingkat kepatuhan wajib pajak yang masih cukup rendah [2]. Oleh karena itu proses layanan administrasi perpajakan seperti pendaftaran wajib pajak, pelaporan Surat Pemberitahuan Tahunan (SPT), pembayaran pajak, pemeriksaan dan penagihan pajak akan diintegrasikan pada Coretax System atau disebut sebagai Coretax. Pada 1 Januari 2025 Direktorat Jenderal Pajak (DJP) secara resmi meluncurkan Coretax kepada publik. Sistem ini menggabungkan seluruh layanan perpajakan, seperti *e-filing*, *e-billing*, dan pelaporan SPT ke dalam satu situs web. Coretax dirancang untuk menyederhanakan prosedur administratif yang sebelumnya terpisah pada *platform* DJP Online, sehingga diharapkan meningkatkan efisiensi dan transparansi tata kelola perpajakan di Indonesia. Namun, inovasi ini menuntut adaptasi yang cepat dari masyarakat, terutama rentang waktu sejak *soft launching* hanya berjarak satu bulan. Hal ini dapat mendorong timbulnya tantangan teknis dan sosial dalam masa transisi dari sistem DJP Online.

Pajak memegang peranan penting sebagai sumber pendapatan negara. Negara sebagai penyedia layanan publik [3] seperti perpajakan perlu menjamin kualitas pelayanan yang diberikan kepada masyarakat. Kepercayaan dan kepuasan publik secara signifikan dipengaruhi oleh kualitas layanan publik digital [4]. Sejak peluncuran Coretax secara resmi, muncul beragam respons dari masyarakat pribadi maupun sebagai pegawai korporasi. Coretax akan berdampak pada berbagai kalangan masyarakat yang memiliki kewajiban perpajakan baik dari individu hingga perusahaan besar sehingga keterlibatan publik menjadi sangat tinggi. Analisis sentimen dan pemodelan topik menjadi penting untuk memahami respons

wajib pajak dan mengukur kepuasan publik. Dengan analisis sentimen dan pemodelan topik dapat mengidentifikasi masalah potensial untuk pengambilan keputusan selanjutnya. Media sosial menjadi *platform* yang efektif untuk mengidentifikasi pola dan tren respon masyarakat [5]. Hal ini disebabkan oleh media sosial memfasilitasi penggunaannya mengekspresikan opini dan sentimen terhadap suatu organisasi, produk, atau jasa [6] termasuk pada layanan publik yang disediakan pemerintah [3]. Volume data yang dihasilkan mencapai jumlah tak hingga [7], [8] baik dalam bentuk unggahan berupa teks dan/atau foto atau interaksi antar pengguna melalui komentar [6]. Sejumlah penelitian memanfaatkan data dari X [9]–[13] dan Instagram [5], [14] dalam melakukan analisis sentimen serta pemodelan topik guna memahami respon dan opini publik.

Integrasi teknologi dalam layanan perpajakan seperti Coretax perlu diimbangi dengan kajian terhadap persepsi dan opini masyarakat. Analisis sentimen dengan data X dapat menjadi dasar pengambilan keputusan terhadap sistem yang ada saat ini karena Indonesia memiliki lebih dari 24 juta pengguna X [15]. Pemilihan media sosial X didasarkan karena karakteristiknya yang berbasis teks. Data berbasis teks memungkinkan ekstraksi sentimen dan identifikasi topik. X telah menghasilkan volume data dengan skala yang cukup besar sehingga dapat menjadi subjek yang cocok untuk melakukan analisis ilmiah [16]. Besarnya jumlah data ini membuat pendekatan secara manual menjadi tidak realistis terutama karena adanya keterbatasan waktu, konsistensi interpretasi, dan risiko subjektivitas. Kondisi ini mendorong pemanfaatan *Natural Language Processing* (NLP) untuk mempermudah proses analisis data dalam jumlah besar. NLP mampu mengidentifikasi pola bahasa, konteks emosional, dan opini dalam kumpulan *tweets*. Selain itu, X merupakan media sosial yang melibatkan diskusi terbuka sehingga dapat memfasilitasi pengumpulan data terkait reaksi dan opini publik terhadap Coretax.

Beberapa penelitian terdahulu telah melakukan analisis sentimen dalam berbagai tema pada data media sosial X termasuk sistem transaksi yang banyak digunakan masyarakat seperti QRIS. Namun penelitian dengan fokus utama terkait

sistem perpajakan di Indonesia belum dapat ditemukan, terutama karena Coretax baru saja diluncurkan pada tahun 2025. Dalam penelitian ini analisis sentimen diimplementasikan dengan model *hybrid* yang mengintegrasikan VADER, TF-IDF, BERT, dan Logistic Regression. VADER berperan sebagai leksikon untuk mengukur polaritas yang disesuaikan untuk media sosial [17]. TF-IDF sebagai metode statistik dengan memberikan bobot pada kata untuk mengetahui signifikansinya dalam *corpus* [18]. BERT sebagai model *deep learning* untuk membuat representasi *vector* dari teks menggunakan arsitektur *encoder* pada *transformers* [19]. Logistic Regression sebagai model *classifier* yang memanfaatkan fitur dari VADER, TF-IDF, dan BERT untuk memprediksi sentimen akhir. Sementara itu, pemodelan topik menggunakan BERTopic yang memanfaatkan *sentence transformers* untuk mengelompokkan *tweets* ke dalam topik [20]. Penelitian dengan analisis sentimen dan pemodelan topik ini dilakukan untuk mengetahui sentimen yang dominan terhadap Coretax disertai dengan topik yang didiskusikan oleh masyarakat. Diharapkan penelitian ini dapat berkontribusi bagi berbagai pihak seperti pengembang Coretax, wajib pajak, dan peneliti sistem pemerintahan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, terdapat beberapa permasalahan yang dapat dirumuskan sebagai berikut.

1. Bagaimana hasil sentimen pengguna X terhadap Coretax di Indonesia menggunakan model *hybrid* VADER, TF-IDF, dan BERT pada Logistic Regression?
2. Bagaimana performa model *hybrid* VADER, TF-IDF, dan BERT pada Logistic Regression dalam mengklasifikasikan sentimen?
3. Bagaimana hasil pemodelan topik pada masing-masing sentimen terhadap Coretax di Indonesia menggunakan BERTopic?
4. Bagaimana performa model BERTopic dalam pemodelan topik masing-masing sentimen?

1.3 Batasan Masalah

Agar penelitian lebih terarah dan fokus, penelitian ini dibatasi oleh beberapa hal sebagai berikut.

1. Data hanya diambil dari media sosial X.
2. Data *tweets* yang diambil menggunakan kata kunci Coretax
3. Data *tweets* hanya menggunakan Bahasa Indonesia.
4. Data dianalisis berdasarkan *tweets* pada rentang tanggal 1 Januari 2025 hingga 28 Februari 2025.
5. Proses model *hybrid* analisis sentimen menggunakan algoritma VADER dengan TF-IDF, BERT, pada Logistic Regression.
6. Proses pemodelan topik dilakukan menggunakan algoritma BERTopic.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

1. Mengklasifikasikan sentimen terkait Coretax di Indonesia pada media sosial X dengan model *hybrid* VADER, TF-IDF, dan BERT pada Logistic Regression.
2. Mengukur performa model *hybrid* VADER, TF-IDF, dan BERT pada Logistic Regression dalam analisis sentimen.
3. Melakukan pemodelan topik pada masing-masing sentimen terkait Coretax di Indonesia pada media sosial X dengan model BERTopic.
4. Mengukur performa model BERTopic dalam pemodelan topik.

1.4.2 Manfaat Penelitian

1. Bagi pengembang Coretax
 - a. Menyediakan informasi terkini mengenai isu aktual terkait Coretax bagi pemerintah untuk mendukung proses pengambilan kebijakan berbasis data.
 - b. Mengeksplorasi perspektif baru dari dampak implementasi sistem perpajakan terhadap respon dan opini masyarakat di media sosial seperti X.

2. Bagi wajib pajak
 - a. Menyediakan gambaran besar sentimen dan topik yang dibahas pada media sosial X terhadap Coretax.
3. Bagi peneliti sistem pemerintahan
 - a. Menguji efektivitas model *hybrid* dalam konteks analisis sentimen dan pemodelan topik layanan digital perpajakan yang dapat menjadi referensi pengembangan NLP pada topik serupa.
 - b. Memberikan pengetahuan sebagai literatur dinamika sosial, politik, atau ekonomi yang muncul terkait perpajakan digital untuk penelitian selanjutnya.

1.5 Sistematika Penulisan

Penulisan pada skripsi ini disusun dalam lima bab dengan sistematika masing-masing bab sebagai berikut.

BAB I PENDAHULUAN

Pada bab satu berisikan penjelasan tentang latar belakang mengenai digitalisasi layanan perpajakan dan Coretax System, rumusan masalah yang berisi beberapa pertanyaan permasalahan terkait analisis sentimen dan pemodelan topik yang ingin dilakukan, batasan masalah, tujuan dari penelitian, manfaat penelitian dan sistematika penulisan penelitian.

BAB II LANDASAN TEORI

Pada bab dua berisikan penjelasan tentang teori yang menjadi landasan untuk penulisan penelitian ini, diantaranya berkaitan dengan topik, *framework*, algoritma, dan juga *tools* yang digunakan. Selain itu juga terdapat tabel penelitian terkait berdasarkan hasil dari *literature review*.

BAB III METODE PENELITIAN

Pada bab tiga dijelaskan mengenai *tweets* pada X sebagai objek penelitian, metode penelitian menggunakan KDD. Variabel independen pada analisis sentimen ialah data *tweets* dan variabel dependennya merupakan klasifikasi sentimen positif dan negatif oleh VADER. Sedangkan pemodelan topik memiliki variabel independen

berupa *tweets* yang telah diberikan label sentimen dan hasil ekstraksi topik sebagai variabel dependen. Selain itu pada bab ini dijelaskan teknik pengumpulan data dan teknik analisis data.

BAB IV ANALISIS DAN HASIL PENELITIAN

Pada bab empat dijelaskan tahap penelitian sesuai dengan *framework* KDD. Terdapat lima tahap diantaranya, *data selection*, *data cleaning*, *data transformation*, *data mining*, dan *interpretation* atau *evaluation*. Selain itu, terdapat diskusi yang membandingkan hasil penelitian ini dengan penelitian terkait.

BAB V SIMPULAN DAN SARAN

Pada bab lima dijelaskan kesimpulan berdasarkan rumusan masalah yang disebutkan pada bab satu disertai limitasi dan saran untuk penelitian selanjutnya.

