

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pajak Kendaraan Bermotor (PKB) merupakan salah satu sumber Pendapatan Asli Daerah (PAD) yang penting bagi pemerintah daerah [1]. Untuk meningkatkan kepatuhan wajib pajak, pemerintah daerah menerapkan program pemutihan pajak, yaitu penghapusan sanksi administratif bagi penunggak PKB sehingga mereka hanya membayar pokok pajak tanpa denda [1], [2]. Kebijakan ini terbukti efektif meningkatkan penerimaan pajak secara cepat—program pemutihan Provinsi Jawa Barat pada 2024 meningkatkan penerimaan PKB di Samsat Subang sebesar Rp18,55 miliar (naik 35%). Namun, penerapan yang terlalu sering berisiko menimbulkan moral hazard, karena masyarakat dapat sengaja menunda pembayaran pajak dengan harapan memperoleh pemutihan berikutnya [3]. Pelaksanaan program pemutihan pajak yang diumumkan melalui media sosial seperti X yang memiliki jumlah pengguna 25,2 juta dan menuai opini-opini terhadap program tersebut seperti pertanyaan, dukungan terhadap manfaat program, kecaman terhadap tujuan dari program ini dan lain sebagainya. Oleh karena itu, perlu dilakukan analisis sentimen untuk mengolah opini tersebut secara efektif dan memvisualisasi hasil dari keberagaman jenis sentimen terhadap topik ini.

Pengumuman program pemutihan pajak melalui media sosial X—dengan basis pengguna 25,2 juta di Indonesia—memicu beragam opini publik, mulai dari pertanyaan, dukungan, hingga kecaman. Volume dan keragaman opini ini sulit dipantau secara manual oleh pemangku kebijakan. Penelitian sentimen terkait layanan PKB sebelumnya umumnya masih menggunakan metode machine learning konvensional seperti Naïve Bayes, baik untuk opini Twitter terhadap layanan PKB [4] maupun ulasan aplikasi Samsat Mobile [5]. Kemampuan contextual embedding pada IndoBERT menjadikan model ini efektif dalam menangkap makna sentimen yang tidak dapat diperoleh hanya dari representasi kata biasa, sebagaimana ditunjukkan pada penelitian sentimen ulasan aplikasi mobile berbahasa Indonesia yang mencapai akurasi tinggi dengan pendekatan fine-tuning IndoBERT [6].

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan model hibrida IndoBERT-CNN-LSTM: IndoBERT menangkap konteks bahasa Indonesia secara mendalam, CNN mengekstraksi fitur lokal teks, dan LSTM mempelajari dependensi urutan kata. Kombinasi ini diharapkan menghasilkan klasifikasi sentimen yang lebih akurat

terhadap opini masyarakat tentang program pemutihan PKB di X, sekaligus menjadi bahan evaluasi bagi pemerintah daerah dalam merumuskan kebijakan pemutihan pajak yang lebih tepat sasaran.

Kombinasi CNN-LSTM digunakan untuk memperkuat proses ekstraksi fitur dan pemahaman pola urutan kata pada data teks. Convolutional Neural Network (CNN) berfungsi untuk mengekstraksi fitur penting dari hasil representasi teks IndoBERT, seperti pola kata atau frasa yang memiliki kecenderungan sentimen tertentu [7]. Setelah itu, Long Short-Term Memory (LSTM) digunakan untuk mempelajari hubungan berurutan antar kata dalam kalimat sehingga model dapat memahami konteks sentimen secara lebih mendalam [8]. Penggabungan IndoBERT dengan CNN-LSTM menghasilkan model yang mampu memanfaatkan keunggulan contextual understanding dari transformer serta kemampuan ekstraksi fitur dan pembelajaran sequence dari deep learning, sehingga lebih efektif dalam melakukan klasifikasi sentimen positif, negatif, dan netral. Hal ini sejalan dengan penelitian yang menunjukkan bahwa model hybrid IndoBERT + CNN-LSTM menghasilkan performa yang lebih unggul dibandingkan IndoBERT saja, dengan peningkatan akurasi dari 91,76% menjadi 94,92% dan F1-score dari 90,66% menjadi 95,41% pada kasus analisis sentimen berbahasa Indonesia [4], [9].

Sesuai dengan yang disampaikan, untuk penelitian ini judul yang tepat dipakai adalah **"Analisis Sentimen Pemutihan Pajak Kendaraan Bermotor di Aplikasi X dengan IndoBERT"**. Diharapkan dapat menghasilkan data yang bermanfaat untuk penelitian berbasis analisis sentimen dan menghasilkan data untuk topik pemutihan pajak kendaraan.

1.2. Rumusan Masalah

Mengacu pada latar belakang yang telah dipaparkan, khususnya menilai opini publik terhadap kebijakan pemutihan pajak kendaraan bermotor, maka rumusan masalah dalam penelitian ini dirumuskan bagaimana kinerja model hibrid IndoBERT dan CNN-LSTM dalam mengklasifikasikan sentimen positif, negatif, dan netral pada teks berbahasa Indonesia terkait program pemutihan pajak kendaraan bermotor.

1.3. Tujuan

Sejalan dengan rumusan masalah yang telah ditetapkan, penelitian ini bertujuan untuk:

1. Menganalisis sentimen masyarakat terhadap program pemutihan pajak kendaraan bermotor di aplikasi X.
2. Membangun dan mengevaluasi model analisis sentimen model hibrid IndoBERT CNN-LSTM untuk klasifikasi teks berbahasa Indonesia.

1.4. Manfaat

Hasil dari penelitian yang ingin dicapai dari penelitian ini adalah:

1. Memberi kontribusi dalam pengembangan model NLP berbahasa Indonesia dan model IndoBERT dalam analisis sentimen.
2. Memberikan visualisasi opini publik terhadap topik pemutihan pajak kendaraan.
3. Sebagai referensi untuk riset lanjutan mengenai terhadap topik pemutihan pajak kendaraan.

1.5. Ruang Lingkup

Untuk pengerjaan penelitian agar sesuai dengan tujuan yang diinginkan, maka lingkup penelitian yang dicakup berupa:

1. Data set yang dipakai diperoleh dari X dengan jangka waktu tanggal 5 Februari 2022 sampai 28 Juni 2025 berdasarkan data tren pencarian topik dan menggunakan program berbasis python untuk proses ekstraksi data yang berjumlah 6.864 postingan.
2. Data set yang telah diperoleh akan dilakukan *pre-processing* dan proses normalisasi karena memiliki *noise* seperti tautan, tag, simbol, dan *emoji*.
3. Data set untuk proses pelatihan model dataset dibagi 80% digunakan untuk proses *training* model dan 20% untuk proses *testing*.
4. Evaluasi kinerja model menggunakan *confusion matrix* dan nilai *accuracy*, *precision*, *recall* dan *f1-score*
5. Implementasi akhir penelitian ini adalah program yang dibuat menggunakan bahasa pemrograman python.