

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kebijakan Pajak Pertambahan Nilai (PPN) sebesar 12% yang mulai berlaku pada 1 Januari 2025 menjadi salah satu kebijakan fiskal yang mendapat perhatian luas dari masyarakat Indonesia. Meskipun pemerintah menyatakan bahwa kebijakan tersebut ditujukan untuk barang dan jasa tertentu, berbagai respons publik tetap muncul karena adanya kekhawatiran terhadap dampaknya pada daya beli dan kondisi ekonomi masyarakat [1]. Persepsi masyarakat terhadap kebijakan perpajakan juga diketahui berkaitan dengan tingkat kepercayaan kepada pemerintah dan persepsi keadilan pajak [2][3]. Berbagai respons tersebut banyak disampaikan melalui media sosial, khususnya X (sebelumnya Twitter), yang memungkinkan pengguna menyampaikan opini, kritik, maupun dukungan terhadap suatu isu secara terbuka melalui *cuitan*, *hashtag*, dan *thread* [4]. Sejumlah penelitian menunjukkan bahwa X mampu merepresentasikan opini publik terkait isu kebijakan dan politik serta menyediakan data teks dalam jumlah besar yang dapat dimanfaatkan untuk analisis sentimen [5][6]. Pada konteks kebijakan PPN 12%, penelitian sebelumnya menunjukkan bahwa diskusi publik di X cenderung didominasi oleh sentimen Negatif dan platform ini juga menjadi sumber data yang relevan untuk memahami persepsi masyarakat terhadap kebijakan tersebut [7][8].

Untuk memahami arah dan pola persepsi Masyarakat, analisis sentiment sebagai bagian dari *Natural Language Processing* digunakan untuk mengidentifikasi dan mengklasifikasikan opini ke dalam kategori Positif, Negatif, atau Netral [9]. Dalam penerapannya pada data media sosial masih menghadapi berbagai tantangan karena karakteristik teks yang tidak terstruktur, mengandung singkatan, bahasa gaul, eufemisme, ironi, serta variasi penulisan yang dapat memengaruhi pemahaman makna suatu teks [10][11]. Penelitian terdahulu mengenai analisis sentimen kebijakan PPN 12% telah dilakukan dan berhasil mengidentifikasi kecenderungan opini masyarakat terhadap kebijakan tersebut [7][8][12][13]. Akan tetapi, beberapa pendekatan klasifikasi yang digunakan masih memiliki keterbatasan dalam memahami konteks semantik, hubungan antarkata, serta ambiguitas bahasa yang sering ditemukan pada data media sosial [14]. Kondisi ini menunjukkan perlunya model yang mampu memahami konteks bahasa secara lebih mendalam untuk meningkatkan kualitas klasifikasi sentimen.

Salah satu pendekatan yang banyak digunakan dalam analisis sentimen berbahasa Indonesia adalah IndoBERT, yaitu model bahasa pralatih yang dikembangkan menggunakan korpus bahasa Indonesia dalam skala besar [15]. Berbagai penelitian menunjukkan bahwa IndoBERT memiliki performa yang baik pada berbagai tugas NLP, termasuk analisis sentiment [16][17][18][19]. Penelitian terkait sentimen kebijakan PPN 12% juga telah memanfaatkan IndoBERT sebagai model klasifikasi [20]. Tetapi penggunaan IndoBERT secara mandiri lebih berfokus pada pembentukan representasi kontekstual teks dan belum secara khusus mengoptimalkan pemodelan pola urutan kata dalam suatu kalimat [21]. Untuk mengatasi hal tersebut, penelitian ini mengusulkan penggunaan model hibrida IndoBERT–BiLSTM. BiLSTM memiliki kemampuan untuk mempelajari informasi dari dua arah sehingga dapat menangkap hubungan antarkata secara lebih komprehensif [22], sedangkan IndoBERT berperan dalam menghasilkan representasi kontekstual yang kaya terhadap teks bahasa Indonesia dibanding BERT dan mBERT [23][15]. Kombinasi keduanya memungkinkan integrasi antara pemahaman semantik dan pemodelan sekuensial sehingga berpotensi meningkatkan performa klasifikasi sentimen dibandingkan model tunggal [24][25][26][27].

Sebelum proses pemodelan dilakukan, data akan melalui tahapan prapemrosesan untuk meningkatkan kualitas data teks dan mengurangi *noise* yang dapat memengaruhi hasil klasifikasi [28][29][30]. Berdasarkan uraian tersebut, penelitian ini mengangkat judul “ANALISIS SENTIMEN PUBLIK TERHADAP KEBIJAKAN PPN 12% PADA MEDIA SOSIAL X DENGAN MODEL HIBRIDA INDOBERT–BILSTM” yang diharapkan mampu mengklasifikasikan sentimen publik terhadap kebijakan PPN 12% ke dalam kategori Positif, Negatif, dan Netral serta mengevaluasi kemampuan model tersebut dalam memahami karakteristik bahasa Indonesia pada media sosial X.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, permasalahan dalam penelitian ini adalah sulitnya menganalisis dan mengklasifikasikan sentimen publik terhadap kebijakan PPN 12% pada media sosial X secara akurat ke dalam kategori Positif, Negatif, dan Netral dari data teks bebas yang berjumlah besar, karena karakteristik bahasanya yang memerlukan pemahaman terhadap konteks semantik dan hubungan sekuensial antarkata.

1.3 Tujuan.

Tujuan dari penelitian ini adalah menganalisis sentimen publik di platform X terhadap kebijakan PPN 12% serta mengimplementasikan model hibrida IndoBERT-BiLSTM untuk mengklasifikasikan sentimen publik ke dalam kategori Positif, Negatif, dan Netral, sehingga dapat menangkap konteks semantik dan hubungan sekuensial antarkata dalam proses klasifikasi, yang selanjutnya diimplementasikan ke dalam aplikasi berbasis web.

1.4 Manfaat

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Penelitian ini menyediakan aplikasi yang memanfaatkan model yang telah dilatih untuk klasifikasi dan penyajian hasil analisis opini publik dari media sosial, sehingga berpotensi dimanfaatkan dalam menganalisis berbagai isu kebijakan sejenis di masa depan.
2. Penelitian ini memberikan referensi tambahan bagi pengembangan penelitian analisis sentimen pada data media sosial dalam memahami opini publik khususnya pendekatan hibrida deep learning pada isu kebijakan pajak.

1.5 Ruang Lingkup

Ruang lingkup dari penelitian ini adalah sebagai berikut:

1. Penelitian ini difokuskan pada analisis sentimen publik terhadap kebijakan PPN 12% menggunakan data unggahan dari media sosial X (Twitter) berupa *tweet* berbahasa Indonesia yang relevan dengan topik kebijakan PPN 12%.
2. Pengambilan data dilakukan pada periode 02 Oktober 2024 hingga 31 Maret 2025 yang mencakup fase sebelum dan sesudah diberlakukannya tarif PPN 12% pada 1 Januari 2025 dengan jumlah data yang didapat sebesar 1913 data.
3. Data penelitian dihimpun melalui proses *crawling* dari media sosial X dengan menggunakan kata kunci yang berkaitan dengan kebijakan PPN 12% (“ppn12%”, “ppn 12%”, “ppn12”, “ppn 12”, “ppn 12 persen”, “ppn 12persen”, “TolakPPN12Persen”, “PPNMemperkuatEkonomi”).
4. Dataset yang diperoleh berformat CSV yang kemudian dikategorikan ke dalam tiga kelas sentimen, yaitu Positif, Netral, dan Negatif melalui tahap *manual labeling*.
5. Tahapan pemrosesan teks (*text preprocessing*) yang dilakukan meliputi *case folding*, *noise removal*, *demojize*, *handling repetitive characters*, serta *slang transformation*.

6. Pembagian data dilakukan dengan membagi data latih sebesar 80% (1.530 *tweet*), data validasi sebesar 10% (191 *tweet*) dan data uji sebesar 10% (192 *tweet*).
7. Konfigurasi *hyperparameter* yang digunakan dalam penelitian ini meliputi *learning rate* sebesar 1×10^{-5} dan 2×10^{-5} , *batch size* sebesar 16 dan 32, *hidden size* BiLSTM sebesar 256 dan 512, *dropout* sebesar 0,3 dan 0,4, serta *epoch* maksimum sebesar 10 dan 15, yang dikombinasikan menjadi lima konfigurasi (K1 hingga K5) pada tiga skenario penanganan ketidakseimbangan kelas, yaitu Baseline, Class Weight, dan Focal Loss.
8. Evaluasi performa model dilakukan menggunakan *confusion matrix* yang menghasilkan metrik evaluasi berupa F1-*macro*, *accuracy*, *precision*, *recall*, dan F1-*score* untuk menilai kinerja model yang dibangun.
9. Penelitian ini tidak mencakup analisis kebijakan fiskal secara mendalam maupun evaluasi dampak ekonomi dari penerapan kebijakan PPN 12%.
10. Penelitian ini dibatasi pada implementasi model analisis sentimen menggunakan bahasa pemrograman Python dengan beberapa *library* pendukung yang digunakan untuk pengumpulan data, prapemrosesan teks, pembangunan model hibrida IndoBERT–BiLSTM, visualisasi data, evaluasi performa model, serta *deployment model* dalam bentuk aplikasi berbasis web.