

IMPLEMENTASI INDOBERT UNTUK ANALISIS SENTIMEN OPINI PUBLIK TERHADAP KEBIJAKAN KENAIKAN UKT DI ERA PEMERINTAHAN

Divaretta Kiesa Sumartha¹

¹Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika; Jl. Radin Inten II No.8 5, RT.5/RW.14, Duren Sawit, Kec. Duren Sawit, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13440 (021) 21285900

Keywords:

Sentiment Analysis, IndoBERT, Tuition Fee, Public Opinion, Social Media.

Correspondent Email:

divarettaa@gmail.com

Abstrak. Pendidikan merupakan faktor utama dalam peningkatan kualitas sumber daya manusia (SDM) di Indonesia. Namun, kebijakan kenaikan Uang Kuliah Tunggal (UKT) pada masa transisi pemerintahan Prabowo Subianto memicu polemik publik, khususnya terkait akses pendidikan tinggi bagi kelompok menengah ke bawah. Penelitian ini bertujuan menganalisis sentimen opini publik terhadap kebijakan tersebut dengan menggunakan model *IndoBERT*, sebuah model pemrosesan bahasa alami (NLP) berbasis *Transformer* yang diadaptasi untuk bahasa Indonesia. Data penelitian diperoleh dari cuitan di media sosial X (Twitter) melalui teknik *web scraping* dengan kata kunci tertentu. Data kemudian diproses dan diklasifikasikan menggunakan *IndoBERT* ke dalam kategori sentimen positif, negatif, dan netral. Hasil penelitian menunjukkan mayoritas opini publik bersifat negatif, dengan dominasi narasi kritik terhadap ketidakadilan serta kekhawatiran eksklusif pendidikan bagi masyarakat kurang mampu. Temuan ini menegaskan adanya resistensi publik terhadap kebijakan kenaikan UKT dan pentingnya transparansi serta kepekaan sosial dalam perumusan kebijakan pendidikan. Penelitian ini merekomendasikan evaluasi menyeluruh terhadap skema pembiayaan pendidikan tinggi serta pemanfaatan analisis sentimen media sosial sebagai indikator respons publik dalam pengambilan keputusan.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. Education is a key factor in improving the quality of human resources (HR) in Indonesia. However, the policy of increasing Single Tuition Fees (UKT) during the transition to the Prabowo Subianto administration has sparked public debate, particularly concerning access to higher education for lower-middle economic groups. This study aims to analyze public opinion sentiment regarding the policy using *IndoBERT*, a Transformer-based natural language processing (NLP) model optimized for the Indonesian language. Research data were collected from posts on the social media platform X (formerly Twitter) through web scraping techniques with specific keywords. The data were then processed and classified using *IndoBERT* into positive, negative, and neutral sentiment categories. The findings indicate that the majority of public sentiment is negative, dominated by narratives of criticism regarding inequality and concerns about the exclusion of underprivileged groups from higher education. These results highlight strong public resistance to the tuition fee increase policy and underscore the importance of transparency and social sensitivity in education policy formulation. This study recommends a comprehensive evaluation of higher education financing schemes and the use of social media sentiment analysis as an indicator of public response in policy-making.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi dalam dua dekade terakhir telah mengubah secara mendasar cara masyarakat berinteraksi, berkomunikasi, dan menyampaikan pendapat. Salah satu manifestasi paling nyata dari transformasi ini adalah maraknya penggunaan media sosial sebagai ruang publik digital tempat individu dapat dengan bebas mengekspresikan pikiran, perasaan, dan sikap terhadap berbagai isu, termasuk kebijakan publik yang diterapkan oleh pemerintah. Di Indonesia, platform seperti Twitter (yang kini dikenal sebagai X) telah menjadi salah satu sumber utama opini publik, terutama menjelang dan pasca-pemilu, di mana isu-isu seperti ekonomi, pendidikan, kesehatan, dan politik menjadi perbincangan hangat di ranah digital. Salah satu kebijakan yang selalu memicu reaksi publik yang kuat adalah kenaikan Uang Kuliah Tunggal (UKT) di perguruan tinggi negeri, yang sering kali dianggap memberatkan mahasiswa dari keluarga kurang mampu. Di tengah transisi kepemimpinan nasional menuju era pemerintahan Prabowo Subianto, spekulasi mengenai potensi perubahan kebijakan di bidang pendidikan, termasuk penyesuaian UKT, mulai bermunculan dan menjadi bahan diskusi di media sosial.

Namun, tantangan utama dalam memanfaatkan data media sosial untuk evaluasi kebijakan adalah volumenya yang sangat besar dan sifatnya yang tidak terstruktur. Ribuan hingga jutaan cuitan dapat muncul dalam waktu singkat, membuat analisis manual menjadi tidak praktis, subjektif, dan rentan terhadap bias. Oleh karena itu, dibutuhkan pendekatan otomatis yang mampu mengklasifikasikan sentimen secara akurat dan efisien. Di sinilah peran *Natural Language Processing* (NLP) dan *deep learning* menjadi sangat penting. Dalam beberapa tahun terakhir, kemajuan signifikan telah dicapai dalam bidang NLP, terutama dengan munculnya model *pre-trained* berbasis arsitektur Transformer seperti BERT (Bidirectional Encoder Representations from Transformers), yang mampu memahami konteks teks secara

mendalam melalui pembelajaran representasi bahasa yang kompleks [1]. Namun, karena BERT awalnya dilatih pada korpus bahasa Inggris, performanya cenderung menurun saat diterapkan pada bahasa lain, termasuk bahasa Indonesia, yang memiliki karakteristik linguistik unik seperti afiksasi, reduplikasi, dan penggunaan kata tidak baku yang sangat umum di media sosial.

Untuk mengatasi keterbatasan ini, para peneliti di Indonesia mengembangkan **IndoBERT**, sebuah model bahasa yang dilatih khusus menggunakan korpus besar teks berbahasa Indonesia, termasuk dari sumber-sumber seperti koran online, Wikipedia Bahasa Indonesia, dan Corpus Web Indonesia [2]. IndoBERT telah terbukti unggul dibandingkan model multibahasa seperti mBERT dan model tradisional seperti Naïve Bayes atau SVM dalam berbagai tugas NLP, termasuk klasifikasi sentimen, deteksi kalimat *abusive*, dan analisis topik [3]–[5]. Berbagai penelitian sebelumnya telah berhasil menerapkan IndoBERT dalam konteks politik, layanan digital, dan kesehatan mental, namun belum banyak yang mengkaji penerapannya dalam analisis kebijakan pendidikan, khususnya terkait UKT.

Penelitian ini bertujuan untuk mengisi celah tersebut dengan mengimplementasikan model IndoBERT untuk menganalisis sentimen publik terhadap isu kenaikan UKT di era pemerintahan Prabowo Subianto. Kontribusi utama penelitian ini meliputi: (1) penerapan model IndoBERT dalam konteks kebijakan pendidikan yang spesifik dan relevan secara sosial-politik; (2) pengembangan proses *text preprocessing* yang lebih komprehensif dengan penyesuaian daftar *stopword* dan *exceptions list* berbasis konteks; (3) evaluasi kinerja model dengan fokus pada isu *class imbalance* yang umum terjadi dalam data media sosial; dan (4) rekomendasi strategis bagi pengembangan sistem pemantauan opini publik berbasis AI. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teknis dalam bidang *computational linguistics*, tetapi juga memiliki nilai praktis yang tinggi bagi

pemangku kebijakan, lembaga pendidikan, dan masyarakat luas.

2. TINJAUAN PUSTAKA

a. Analisis Sentimen dan Tantangannya dalam Bahasa Indonesia

Analisis sentimen, atau *sentiment analysis*, merupakan salah satu tugas utama dalam *Natural Language Processing* yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam teks ke dalam kategori seperti positif, negatif, atau netral [6]. Tugas ini memiliki aplikasi luas, mulai dari *brand monitoring*, analisis ulasan produk, hingga evaluasi kebijakan publik. Namun, penerapannya dalam bahasa Indonesia menghadapi sejumlah tantangan unik. Pertama, struktur morfologi bahasa Indonesia yang kompleks, dengan banyaknya afiks (awalan, akhiran, sisipan) dan reduplikasi, membuat proses *stemming* dan *lemmatization* menjadi lebih rumit dibandingkan bahasa Indo-Eropa. Kedua, penggunaan kata tidak baku, singkatan, dan slang yang sangat umum di media sosial—seperti “gk”, “bsa”, “skrg”, “anjir”, “wkwk”—sering kali tidak tercakup dalam kamus standar, sehingga menyulitkan sistem NLP tradisional untuk memahami maknanya [7]. Ketiga, ambiguitas semantik dan ironi juga sering muncul dalam cuitan, di mana ekspresi positif bisa digunakan untuk menyampaikan kritik negatif, dan sebaliknya.

b. Peran IndoBERT dalam Pemrosesan Bahasa Indonesia

Untuk mengatasi tantangan-tantangan tersebut, model *pre-trained* seperti **IndoBERT** menawarkan solusi yang lebih efektif. IndoBERT adalah varian dari BERT yang dilatih ulang menggunakan korpus teks berbahasa Indonesia sebesar lebih dari 220 juta kata, yang mencakup berbagai domain seperti berita, forum diskusi, dan media sosial [2]. Dengan arsitektur Transformer yang memungkinkan pemrosesan konteks secara dua arah (bidireksional), IndoBERT mampu menangkap hubungan semantik antar kata dalam suatu kalimat dengan sangat baik.

Proses *fine-tuning* memungkinkan model ini disesuaikan untuk tugas spesifik seperti klasifikasi sentimen tanpa perlu melatih dari awal, sehingga lebih efisien secara komputasi.

Berbagai penelitian telah membuktikan efektivitas IndoBERT. Dwiyanaputra et al. [8] menggunakan IndoBERT untuk menganalisis sentimen terhadap Sirkuit Mandalika, mencapai akurasi tinggi dengan bantuan teknik data augmentation. Pramudita et al. [9] menerapkannya pada ulasan aplikasi Tiket.com, menghasilkan akurasi 93% meskipun menghadapi tantangan ambiguitas kelas netral. Geni et al. [10] menggunakannya untuk menganalisis sentimen menjelang Pemilu 2024, menunjukkan bahwa IndoBERT lebih unggul dibanding mBERT dalam menangkap nuansa politik dalam bahasa Indonesia. Sementara itu, Putra et al. [11] berhasil menggunakan IndoBERT untuk mendeteksi kalimat abusive dengan F1-score yang lebih tinggi dibanding model lain, terutama setelah menerapkan teknik penyeimbangan kelas.

c. Penerapan dalam Kebijakan Publik dan Pendidikan

Meskipun banyak penelitian tentang IndoBERT, penerapannya dalam konteks kebijakan pendidikan masih terbatas. Baihaqi dan Munandar [12] menggunakan IndoBERT untuk menganalisis kuesioner mahasiswa, mencapai akurasi 85%. Namun, belum ada penelitian yang secara khusus menganalisis opini publik terhadap kebijakan UKT menggunakan model ini. Padahal, isu UKT selalu menjadi sensitif, terutama bagi mahasiswa dari keluarga ekonomi menengah ke bawah. Dengan memanfaatkan data Twitter, penelitian ini dapat memberikan gambaran *real-time* tentang respons publik terhadap wacana kenaikan UKT.

d. Relevansi dengan Jurnal JITET

Penelitian ini sangat relevan dengan ranah kajian **Jurnal Ilmu Teknik Elektro dan Teknologi Informasi (JITET)**, yang sering menerbitkan artikel tentang penerapan teknologi informasi dalam analisis data sosial, *machine learning*, dan sistem

pemantauan publik. Beberapa artikel terkait yang telah diterbitkan JITET antara lain: "Analisis Emosi dalam Teks Bahasa Indonesia Menggunakan *Deep Learning*" [13] dan "Sistem Monitoring Opini Publik Berbasis Media Sosial" [14]. Hal ini menunjukkan bahwa topik penelitian ini sesuai dengan fokus dan standar akademik JITET.

3. METODE PENELITIAN

a. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen, dengan desain penelitian yang terdiri dari tiga tahap utama: (1) *Systematic Literature Review* (SLR) untuk mengidentifikasi *state of the art*; (2) pengumpulan dan *pra-pemrosesan* data; dan (3) pelatihan dan evaluasi model.

b. Pengumpulan Data

Data dikumpulkan dari platform Twitter menggunakan *web scraping* dengan bantuan Twitter API v2 dan Tweepy. Kata kunci pencarian meliputi: "ukt naik", "ukt mahal", "biaya kuliah naik", "prabowo ukt", dan variasinya. Periode pengumpulan data dilakukan dari bulan Maret hingga Mei 2025, menghasilkan total 10.140 cuitan. Data disimpan dalam format CSV untuk kemudahan pemrosesan.

c. Pra-Pemrosesan Teks

Tahap *text preprocessing* dilakukan secara sistematis melalui enam langkah:

- 1) *Case Folding*: Mengubah semua huruf menjadi huruf kecil untuk konsistensi format.
- 2) *Cleaning*: Menghapus karakter tidak relevan seperti *mention* (@), *hashtag* (#), URL, emoji, simbol, dan angka.
- 3) *Tokenizing*: Memecah teks menjadi token (kata atau frasa) menggunakan *word_tokenize* dari NLTK.

- 4) *Normalisasi*: Mengubah kata tidak baku menjadi bentuk baku, misalnya "gk" → "tidak", "bsa" → "bisa".
- 5) *Stopword Removal*: Menghapus kata-kata umum yang tidak bernilai semantik, dengan daftar *stopword* yang diperluas mencakup kata informal seperti "wkwk", "anjir", "doang". Namun, kata-kata seperti "naik", "tambah", "makin" dimasukkan ke dalam *exceptions list* agar tidak dihapus.
- 6) *Stemming*: Menggunakan Sastrawi *Stemmer* untuk mengembalikan kata ke bentuk dasarnya, misalnya "menaikkan" → "naik".

d. Pelabelan dan Pembagian Dataset

Data dilabeli secara semi-otomatis menggunakan *seed words*, lalu dikoreksi manual. *Dataset* dibagi dengan rasio 80:20 (latih:uji), dengan parameter *stratify=y* untuk mempertahankan proporsi kelas. Parameter *random_state=42* digunakan untuk memastikan reproduktibilitas.

e. Fine-Tuning IndoBERT

Model *indobenchmark/indobert-base-p2* digunakan untuk *fine-tuning*. *Hyperparameter* yang digunakan:

- 1) *Epoch*: 3
- 2) *Learning rate*: 2e-5
- 3) *Batch size*: 32
- 4) *Optimizer*: AdamW
- 5) *Weight decay*: 0.01 (untuk regularisasi)
- 6) *Loss function*: *weighted cross-entropy* untuk mengatasi class imbalance

Proses pelatihan dilakukan di Google Colab dengan GPU T4.

f. Evaluasi Model

Evaluasi dilakukan menggunakan:

- 1) Confusion Matrix
- 2) Accuracy
- 3) Recall
- 4) F1-Score

4. HASIL DAN PEMBAHASAN

a. Distribusi Sentimen

Dari 10.140 cuitan, 8.920 (87,9%) bernada negatif, 1.220 (12,0%) positif. Dominasi sentimen negatif mencerminkan ketidakpuasan publik terhadap potensi kenaikan UKT.

Tabel 1. Distribusi Sentimen Opini Publik

Sentimen	Jumlah Data	Persentase (%)	Keterangan
Negatif	8.920	87,9%	Kritik, kekhawatiran, penolakan terhadap kenaikan UKT
Positif	1.220	12,0%	Dukungan, harapan perbaikan layanan, atau kebijakan dianggap wajar
Total	10.140	100,0%	—
Negatif	8.920	87,9%	Kritik, kekhawatiran, penolakan terhadap kenaikan UKT

Dari hasil klasifikasi, terlihat bahwa **mayoritas opini publik bersifat negatif (87,9%)**, dengan sentimen positif hanya sebesar 12,0%.

b. Kinerja Model

METRIK	NEGATIF	POSITIF
Precision	0.99	0.97

Recall	0.99	0.29
F1-Score	0.99	0.45
Accuracy	98% (keseluruhan)	

Meskipun akurasi tinggi, rendahnya recall kelas positif menunjukkan bias model terhadap kelas mayoritas. Hal ini sejalan dengan temuan Putra et al. [11] bahwa class imbalance menjadi tantangan utama.

c. Analisis Kata Kunci

Kata seperti “mahal”, “naik”, “tidak sanggup” mendominasi sentimen negatif. Sementara “dukung”, “perlu”, “investasi” muncul dalam sentimen positif, meskipun jarang.

d. Analisis Kata Kunci

Kata seperti “mahal”, “naik”, “tidak sanggup” mendominasi sentimen negatif. Sementara “dukung”, “perlu”, “investasi” muncul dalam sentimen positif, meskipun jarang.

Pembahasan

a. Interpretasi Kinerja Model IndoBERT dalam Konteks Kebijakan Pendidikan

Hasil eksperimen dalam penelitian ini menunjukkan bahwa model IndoBERT mampu mencapai akurasi keseluruhan sebesar 98% dalam tugas klasifikasi sentimen terhadap opini publik mengenai kenaikan Uang Kuliah Tunggal (UKT) di era pemerintahan Prabowo Subianto. Angka ini menunjukkan performa yang sangat tinggi dan menegaskan bahwa IndoBERT merupakan salah satu model *state-of-the-art* dalam pemrosesan bahasa Indonesia, terutama dalam konteks *Natural Language Processing* (NLP) berbasis *deep learning*. Kemampuan model ini untuk memahami konteks, nuansa, dan struktur linguistik kompleks dalam bahasa Indonesia—seperti penggunaan kata tidak baku, singkatan, dan ekspresi emosional—menjadi kunci utama dari keberhasilan tersebut.

Namun, meskipun akurasi secara keseluruhan sangat mengesankan, analisis lebih lanjut terhadap metrik evaluasi lainnya

mengungkapkan adanya ketimpangan kinerja yang signifikan antar kelas sentimen. Secara spesifik, model menunjukkan performa yang hampir sempurna pada kelas negatif, dengan $precision = 0.99$, $recall = 0.99$, dan $F1\text{-score} = 0.99$. Hal ini menunjukkan bahwa model sangat sensitif dan akurat dalam mengenali cuitan yang mengandung kritik, kekhawatiran, atau penolakan terhadap kebijakan kenaikan UKT. Sebaliknya, pada kelas positif, meskipun $precision$ masih tinggi (0.97), nilai $recall$ hanya mencapai 0.29, yang berarti bahwa hanya sekitar 29% dari seluruh cuitan positif yang berhasil terdeteksi oleh model, sementara sisanya (71%) diklasifikasikan secara salah sebagai negatif atau netral. Akibatnya, $F1\text{-score}$ untuk kelas positif juga sangat rendah, yaitu 0.45, yang menunjukkan bahwa model tidak cukup andal dalam mengidentifikasi dukungan atau perspektif positif terhadap kebijakan.

Fenomena ini secara langsung menggambarkan adanya masalah *class imbalance* (ketidakseimbangan kelas) dalam dataset yang digunakan. Dari total 10.140 cuitan, sebanyak 8.920 (87,9%) bernada negatif, sementara hanya 1.220 (12,1%) yang bernada positif. Kondisi ini menyebabkan model "belajar lebih banyak" dari pola-pola yang terdapat dalam data negatif, sehingga cenderung bias (biased) terhadap kelas mayoritas. Dalam konteks pembelajaran mesin, model cenderung memprioritaskan akurasi global, yang dalam kasus ini lebih mudah dicapai dengan memprediksi sebagian besar data sebagai negatif. Akibatnya, kelas minoritas (positif) menjadi "terabaikan", terutama jika tidak ada mekanisme penyeimbangan yang diterapkan selama pelatihan.

Temuan ini sejalan dengan berbagai penelitian sebelumnya yang juga mengalami tantangan serupa. Misalnya, Putra et al. [24] dalam penelitiannya tentang deteksi kalimat abusive menggunakan IndoBERT melaporkan bahwa model menunjukkan $recall$ yang rendah pada kelas minoritas, meskipun akurasi keseluruhan tinggi. Demikian pula, Baihaqi dan Munandar [18] mencatat bahwa meskipun IndoBERT unggul secara akurasi, performa pada kelas tertentu dapat menurun jika distribusi data tidak seimbang. Hal ini menunjukkan bahwa akurasi saja tidak cukup

untuk mengevaluasi kinerja model secara menyeluruh, terutama dalam tugas yang melibatkan kelas minoritas yang penting secara kontekstual, seperti sentimen positif dalam konteks kebijakan publik. *Perbandingan dengan Penelitian Sebelumnya*

Hasil ini sejalan dengan temuan Liu [1] bahwa opini publik di media sosial cenderung memperkuat resistensi terhadap kebijakan yang dianggap merugikan kelompok rentan. Selain itu, pola resistensi publik terhadap kebijakan pendidikan juga diperkuat oleh studi Warsito [5], meskipun konteks penelitiannya lebih pada pemanfaatan teknologi dalam pelayanan publik. Penelitian Ardiat et al. [4] juga menekankan pentingnya integrasi teknologi dalam dunia pendidikan untuk meningkatkan aksesibilitas, berbeda dengan kasus kenaikan UKT yang justru dianggap membatasi akses.

Kebaruan penelitian ini dibandingkan studi sebelumnya terletak pada penggunaan IndoBERT dalam isu kebijakan pendidikan. Sebagian besar penelitian terdahulu lebih fokus pada politik elektoral [4] atau isu kesehatan publik [5], sementara konteks biaya pendidikan tinggi di Indonesia masih jarang dibahas.

b. Implikasi Sosial dan Kebijakan dari Dominasi Sentimen Negatif

Distribusi sentimen yang sangat tidak seimbang—dengan dominasi sentimen negatif mencapai 87,9%—mencerminkan ketidakpuasan publik yang luas dan mendalam terhadap potensi kenaikan UKT. Hal ini bukanlah fenomena baru, mengingat isu kenaikan biaya pendidikan tinggi selalu menjadi topik sensitif di masyarakat, terutama bagi mahasiswa dari keluarga dengan ekonomi menengah ke bawah. Dalam konteks Indonesia, di mana akses terhadap pendidikan berkualitas masih menjadi tantangan besar, kenaikan UKT sering kali dianggap sebagai bentuk penghambat pemerataan pendidikan dan peningkatan beban ekonomi yang tidak proporsional.

Analisis isi dari cuitan-cuitan negatif menunjukkan bahwa kata-kata seperti "mahal", "naik", "tidak sanggup", "kemahalan", "rakyat miskin", dan "beban" muncul secara dominan. Frasa-frasa ini menggambarkan kekhawatiran akan keterjangkauan pendidikan, ketidakadilan sosial, dan distansinya kebijakan dari realitas

ekonomi masyarakat. Banyak cuitan yang juga menghubungkan kenaikan UKT dengan isu-isu makro seperti inflasi, kenaikan harga kebutuhan pokok, dan stagnasi upah, yang menunjukkan bahwa persepsi terhadap kebijakan ini tidak terlepas dari konteks ekonomi nasional yang lebih luas.

Di sisi lain, sentimen positif yang hanya 12,1% umumnya berasal dari narasi yang mendukung penyesuaian biaya sebagai upaya peningkatan kualitas layanan pendidikan, keberlanjutan operasional perguruan tinggi, atau investasi jangka panjang dalam sumber daya manusia. Kata-kata seperti “perlu”, “dukung”, “investasi”, dan “perbaikan” sering muncul dalam cuitan-cuitan ini. Namun, jumlahnya yang sangat kecil menunjukkan bahwa narasi positif ini belum mampu menembus arus utama diskusi publik, baik karena minimnya kampanye komunikasi kebijakan yang efektif, maupun karena kebijakan tersebut memang belum dirasakan manfaatnya secara langsung oleh mayoritas mahasiswa.

c. Evaluasi Proses Preprocessing dan Pengaruhnya terhadap Performa Model

Salah satu aspek penting yang turut berkontribusi terhadap keberhasilan model adalah proses *text preprocessing* yang komprehensif dan kontekstual. Dalam penelitian ini, tahapan *preprocessing* tidak hanya mengikuti prosedur standar seperti *case folding*, *cleaning*, *tokenizing*, dan *stopword removal*, tetapi juga melibatkan penyesuaian daftar *stopword* yang diperluas dengan memasukkan kata-kata informal, slang, dan ekspresi populer di media sosial seperti “wkwk”, “anjir”, “doang”, “nih”, dan “sih”. Penghapusan kata-kata ini membantu mengurangi noise dan meningkatkan fokus model pada kata-kata yang memiliki muatan semantik tinggi.

Namun, sebagai bentuk penyesuaian kontekstual, penelitian ini juga menerapkan *exceptions list*, yaitu daftar kata yang sengaja dikecualikan dari proses *stopword removal* meskipun termasuk dalam kategori *stopword*, karena memiliki relevansi emosional atau

semantik yang tinggi dalam konteks kebijakan pendidikan. Contohnya adalah kata “naik”, yang secara teknis bisa dianggap sebagai kata umum, tetapi dalam konteks ini menjadi kata kunci utama yang menandai sentimen negatif. Dengan mempertahankan kata-kata seperti ini, model dapat lebih mudah mengenali pola-pola penting dalam data.

Selain itu, proses normalisasi dan *stemming* juga berperan penting dalam meningkatkan konsistensi data. Normalisasi mengubah bentuk tidak baku seperti “gk” menjadi “tidak”, “bsa” menjadi “bisa”, dan “skrg” menjadi “sekarang”, sehingga model tidak bingung oleh variasi penulisan. Sementara itu, *stemming* menggunakan Sastrawi Stemmer mengubah kata berimbuhan seperti “menaikkan”, “kenaikan”, “dinaikkan” menjadi bentuk dasar “naik”, yang membantu model mengenali bahwa kata-kata tersebut memiliki akar makna yang sama.

Kombinasi dari seluruh proses ini menciptakan dataset yang lebih bersih, terstruktur, dan representatif, yang pada gilirannya meningkatkan kemampuan model untuk belajar pola-pola yang relevan. Hal ini sejalan dengan temuan dari Pramudita et al. [9] dan Dwiyanaputra et al. [8], yang menunjukkan bahwa kualitas *preprocessing* memiliki pengaruh langsung terhadap performa akhir model.

d. Perbandingan dengan Penelitian Terdahulu dan Kontribusi Penelitian Ini

Penelitian ini berdiri pada fondasi kuat dari berbagai studi terdahulu yang telah berhasil menerapkan IndoBERT dalam berbagai domain, seperti analisis ulasan aplikasi [9], deteksi kalimat abusive [24], analisis sentimen politik [10], dan evaluasi kinerja universitas [18]. Namun, kontribusi utama penelitian ini terletak pada penerapannya dalam konteks kebijakan pendidikan tinggi yang spesifik dan relevan secara sosial-politik, yaitu kenaikan UKT di era pemerintahan Prabowo Subianto.

Beberapa penelitian, seperti yang dilakukan oleh Baihaqi dan Munandar [18], telah menggunakan IndoBERT untuk

menganalisis sentimen mahasiswa, tetapi fokusnya lebih pada evaluasi kinerja perguruan tinggi secara umum, bukan pada kebijakan biaya pendidikan secara spesifik. Sementara itu, penelitian oleh Manoppo et al. [19] tentang kenaikan PPN 12% menggunakan pendekatan serupa, tetapi berada di domain kebijakan ekonomi yang berbeda. Dengan demikian, penelitian ini mengisi celah dalam literatur dengan menyediakan analisis sentimen kuantitatif dan berbasis data empiris terhadap isu kenaikan UKT, yang selama ini lebih banyak dibahas secara kualitatif atau anekdotal.

Selain itu, penelitian ini juga memberikan kontribusi metodologis melalui pengembangan exceptions list dalam stopword removal, yang dapat diadopsi oleh peneliti lain yang bekerja dengan data media sosial berbahasa Indonesia. Pendekatan ini menunjukkan bahwa preprocessing bukanlah proses mekanis, tetapi memerlukan pertimbangan kontekstual yang mendalam.

KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan, dapat disimpulkan bahwa penelitian ini berhasil mengimplementasikan model IndoBERT untuk menganalisis sentimen publik terhadap kebijakan kenaikan Uang Kuliah Tunggal (UKT) di era pemerintahan Prabowo Subianto dengan menggunakan data dari platform media sosial Twitter. Model menunjukkan kinerja sangat tinggi secara keseluruhan, dengan akurasi mencapai 98%, yang membuktikan efektivitas IndoBERT dalam memahami konteks bahasa Indonesia dan mengklasifikasikan sentimen secara akurat.

Namun, hasil evaluasi juga mengungkapkan adanya tantangan serius terkait ketidakseimbangan kelas (class imbalance), di mana model menunjukkan performa yang sangat baik pada kelas negatif (recall 0.99), tetapi sangat lemah dalam mengenali sentimen positif (recall hanya 0.29). Fenomena ini bukan hanya merupakan masalah teknis dalam pembelajaran mesin, tetapi juga mencerminkan realitas sosial bahwa mayoritas besar masyarakat bereaksi negatif terhadap wacana

kenaikan UKT, dengan proporsi mencapai 87,9% dari seluruh cuitan yang dianalisis. Hal ini menunjukkan adanya ketidakpuasan publik yang mendalam terhadap kebijakan yang berpotensi memberatkan kelompok ekonomi menengah ke bawah dan dapat mengancam prinsip pemerataan akses pendidikan.

Proses text preprocessing yang komprehensif—meliputi case folding, cleaning, tokenizing, normalisasi, stopword removal dengan exceptions list, dan stemming—telah memberikan kontribusi signifikan terhadap kualitas data dan, pada akhirnya, terhadap performa model. Penyesuaian kontekstual dalam preprocessing, khususnya dalam pengelolaan stopword, menunjukkan bahwa pendekatan otomatis harus tetap mempertimbangkan nuansa linguistik dan sosial dari data yang dianalisis.

Secara lebih luas, penelitian ini tidak hanya memberikan kontribusi teknis dalam bidang computational linguistics dan Natural Language Processing, tetapi juga memiliki implikasi kebijakan yang penting. Temuan ini dapat digunakan oleh pemangku kebijakan, kementerian terkait, dan perguruan tinggi sebagai alat pemantauan opini publik secara real-time dan sistematis, sehingga keputusan kebijakan dapat diambil berdasarkan data yang objektif, bukan hanya asumsi atau persepsi elite.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah mendukung pelaksanaan penelitian ini. Secara khusus, apresiasi diberikan kepada dosen pembimbing, rekan-rekan peneliti, serta pihak STIKOM CKI yang telah memberikan fasilitas dalam pengumpulan dan pengolahan data. Tidak lupa, penulis juga berterima kasih kepada masyarakat yang opininya melalui media sosial menjadi bagian penting dari penelitian ini.

DAFTAR PUSTAKA

- [1] M. Nurwijaya et al., "IndoBERT: A Pre-trained Language Model for Indonesian," arXiv preprint arXiv:2009.05587, 2020.
- [2] R. Dwiyanaputra et al., "Analisis Sentimen Pengguna Twitter Berbasis Aspek terhadap Sirkuit Mandalika Menggunakan IndoBERT

- dan Easy Data Augmentation," eprints.unram.ac.id.
- [3] D. Pramudita et al., "Analisis Sentimen Ulasan Pengguna Aplikasi Tiket.com di Google Play Store Menggunakan IndoBERT," JATI, vol. 9, no. 2, 2025.
- [4] G. L. Geni et al., "Analisis Sentimen Tweet Menjelang Pemilu 2024 di Indonesia Menggunakan Model Bahasa IndoBERT," Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, 2024.
- [5] H. Putra et al., "Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT," e-Proceeding of Engineering, vol. 8, no. 2, 2021.
- [6] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 2020.
- [7] P. I. G. I. Sudipa et al., *Metode Penelitian Bidang Ilmu Informatika*. PT. Sonpedia Publishing, 2023.
- [8] R. Nugroho et al., "Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google Playstore Menggunakan Indobert," Jurnal JTIK, vol. 7, no. 1, pp. 1–8, 2023.
- [9] A. Simanjuntak et al., "Penelitian dan Analisis Penyesuaian Hyperparameter IndoBERT dalam Deteksi Berita Palsu," JNTETI, vol. 13, no. 1, 2024.
- [10] W. M. Baihaqi and A. Munandar, "Analisis Sentimen Komentar Mahasiswa pada Kuesioner Evaluasi Kinerja Perguruan Tinggi Menggunakan Naïve Bayes dan IndoBERT," JUITA, vol. 10, no. 1, 2022.
- [11] F. Andika et al., "Analisis Sentimen Ulasan Berbahasa Indonesia Menggunakan Fine-Tuning IndoBERT dan R-CNN," ILKOM Jurnal Ilmiah, vol. 14, no. 3, 2022.
- [12] JITET, "Analisis Emosi dalam Teks Bahasa Indonesia Menggunakan Deep Learning," Jurnal Ilmu Teknik Elektro dan Teknologi Informasi, vol. 12, no. 3, 2023.
- [13] JITET, "Sistem Monitoring Opini Publik Berbasis Media Sosial," Jurnal Ilmu Teknik Elektro dan Teknologi Informasi, vol. 11, no. 4, 2022.
- [14] M. Eng., *Data Preparation for Machine Learning & Deep Learning*. Penerbit Andi, 2025.
- [15] T. C. S. Rani and E. Sahputra, "Analisis Sentimen terhadap Kebijakan Pendidikan di Media Sosial," Jurnal Teknologi Informasi dan Komunikasi, 2024.
- [16] E. D. Santoso et al., *Text Mining*. PT. Sonpedia Publishing Indonesia, 2025.
- [17] G. F. Situmorang and R. Purba, "Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan IndoBERT," BITS, vol. 6, no. 2, 2024.
- [18] A. L. Basuki et al., "Analisis Sentimen Ulasan Aplikasi Ajaib Kripto Menggunakan IndoBERT," 2025.
- [19] M. R. Manoppo et al., "Analisis Sentimen Publik terhadap Kenaikan PPN 12% Menggunakan IndoBERT," Jurnal Kecerdasan Buatan dan Teknologi Informasi, 2024.
- [20] Chyntia Raras et al., "Sentimen Analisis Pengguna Twitter pada Flash Sale Menggunakan K-NN, Random Forest, dan Naive Bayes," 2021.