Perbandingan Metode Indobert Dan Xlnet Dalam Mengukur Kemiripan Semantik Antara Tweet Dan IKP 2024

Stevren Christian Emor[1], Irene R. H. T. Tangkawarow[2], Audy A. Kenap[3]
[1,2,3] Universitas Negeri Manado
e-mail: [1]20210034@unima.ac.id, [2]irene.tangkawarouw@unima.ac.id, [3]audykenap@unima.ac.id

Diajukan: 01 Oktober 2025 ; Direvisi: 12 Oktober 2025; Diterima: 21 Oktober 2025

Abstrak

Pemilihan Umum (Pemilu) merupakan pilar demokrasi yang rentan terhadap polarisasi opini publik dan penyebaran disinformasi di media sosial. Untuk memahami dinamika tersebut, diperlukan pendekatan analisis semantik guna mengidentifikasi potensi kerawanan pemilu. Penelitian ini bertujuan membandingkan kinerja dua model berbasis transformer, IndoBERT dan XLNet, dalam mengukur kemiripan semantik antara komentar media sosial (Twitter/X) dan indikator Indeks Kerawanan Pemilu (IKP) 2024. Data diperoleh melalui teknik crawling sebanyak 574 tweet dengan kata kunci "Pemilu2024". Tahapan penelitian meliputi preprocessing (pembersihan data, tokenizing, normalisasi, dan stemming), analisis menggunakan model IndoBERT dan XLNet, perhitungan cosine similarity, serta evaluasi melalui Confusion Matrix (akurasi, presisi, recall, F1-score) dan Expert Judgment. Hasil menunjukkan bahwa XLNet memiliki performa lebih baik dengan akurasi 76%, presisi 69,5%, recall 80%, dan F1-score 74,3%, dibandingkan IndoBERT dengan akurasi 59,7%, presisi 54,3%, recall 60,9%, dan F1-score 57,3%. Penelitian ini berkontribusi dalam pengembangan metode komparatif berbasis Natural Language Processing (NLP) untuk mendeteksi potensi isu kerawanan pemilu melalui analisis opini publik di media sosial sebagai dukungan bagi pengawasan partisipatif Bawaslu.

Kata kunci: *IndoBERT, XLNet, Kemiripan Semantik, Media Sosial, Indeks Kerawanan Pemilu.*

Abstract

General elections are a vital pillar of democracy yet remain vulnerable to polarization of public opinion and the spread of disinformation on social media. To address this issue, a semantic analysis approach is needed to identify potential election vulnerabilities reflected in online discussions. This study aims to compare the performance of two transformer-based models, IndoBERT and XLNet, in measuring semantic similarity between social media comments (Twitter/X) and the 2024 Election Vulnerability Index (IKP) indicators. Data were collected through a crawling technique, resulting in 574 tweets with the keyword "Pemilu2024." The research stages included preprocessing (data cleaning, tokenizing, normalization, and stemming), semantic analysis using IndoBERT and XLNet, similarity calculation with cosine similarity, and evaluation through a Confusion Matrix (accuracy, precision, recall, F1-score) and Expert Judgment. The results indicate that XLNet performed better with 76% accuracy, 69.5% precision, 80% recall, and a 74.3% F1-score, compared to IndoBERT's 59.7% accuracy, 54.3% precision, 60.9% recall, and 57.3% F1-score. This research contributes to the development of a comparative Natural Language Processing (NLP) approach for detecting potential election vulnerability issues through public opinion analysis on social media, supporting Bawaslu's efforts in participatory electoral supervision.

Keywords*: IndoBERT, XLNet, semantic similarity, social media, Election Vulnerability Index..*

1. Pendahuluan

Pemilihan umum (pemilu) merupakan salah satu pilar utama dalam sistem demokrasi modern yang berfungsi sebagai mekanisme untuk memilih pemimpin dan perwakilan rakyat. Pemilu tidak hanya menjadi sarana legitimasi politik, tetapi juga instrumen untuk memperkuat partisipasi masyarakat dalam menentukan arah kebijakan negara Namun, di sisi lain, penyelenggaraan pemilu kerap menghadapi tantangan serius berupa kerawanan politik, sosial, dan keamanan. Potensi kerawanan ini dapat menimbulkan konflik, menurunkan kepercayaan publik, serta mengganggu stabilitas demokrasi[1]

Untuk memetakan potensi kerawanan, Badan Pengawas Pemilu (Bawaslu) mengembangkan Indeks Kerawanan Pemilu (IKP). Instrumen ini mengukur tingkat kerawanan suatu daerah berdasarkan data

historis, survei lapangan, serta pemetaan potensi konflik. IKP edisi terbaru (2024) mencakup empat dimensi utama, yaitu: konteks sosial-politik, penyelenggaraan pemilu, kontestasi, dan partisipasi masyarakat [2], [3]. IKP tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai instrumen pencegahan yang dapat membantu pemerintah dan penyelenggara pemilu mengambil langkah preventif [4].

Dalam konteks Indonesia, ancaman kerawanan pemilu semakin kompleks dengan hadirnya media sosial sebagai arena komunikasi politik. Twitter (X) secara khusus menjadi platform yang banyak digunakan untuk kampanye politik, penyebaran informasi, hingga mobilisasi opini public [5]. Analisis terhadap percakapan dan komentar di media sosial menunjukkan bahwa opini publik dapat terbentuk dan berkembang dengan cepat, seringkali tidak terlepas dari fenomena disinformasi, ujaran kebencian, dan polarisasi politik [6], [7]. Penelitian di berbagai negara membuktikan bahwa media sosial memiliki potensi besar dalam memengaruhi preferensi politik masyarakat dan bahkan hasil pemilu[8].

Fenomena polarisasi politik di media sosial Indonesia juga semakin marak, terutama dengan adanya aktivitas coordinated behavior melalui penggunaan hashtag terkoordinasi. Studi terbaru menemukan bahwa terdapat pola kampanye digital yang berpotensi memanipulasi opini publik serta meningkatkan kerawanan pemilu [9]. Oleh sebab itu, analisis semantik terhadap komentar media sosial menjadi penting untuk mengidentifikasi potensi kerawanan sekaligus memahami dinamika politik digital.

Selain itu, perkembangan media sosial sebagai ruang diskusi politik menjadikan analisis terhadap opini publik semakin penting untuk dilakukan. Media sosial seperti Twitter/X telah menjadi salah satu wadah utama masyarakat dalam menyampaikan pandangan terkait Pemilu 2024, sehingga dapat memberikan gambaran mengenai dinamika persepsi public [10]. Hal ini menegaskan perlunya pemanfaatan teknologi analisis data untuk memetakan opini masyarakat yang terus berubah dengan cepat.

Tidak hanya melalui media sosial, berita daring juga memiliki peran strategis dalam memberikan informasi dan merekam tanggapan publik terhadap tahapan pemilu. Kajian yang dilakukan menunjukkan bahwa informasi dari berita daring dapat digunakan untuk memahami sikap masyarakat, baik dalam bentuk opini positif maupun negatif, sehingga dapat mendukung pengawasan pemilu yang lebih efektif [11].

Lebih lanjut, analisis terhadap data publik menunjukkan adanya kecenderungan bahwa opini masyarakat di media sosial dapat dipetakan secara lebih akurat dengan memanfaatkan pendekatan komputasional [12]. Hal ini membuktikan bahwa data yang kompleks dan beragam dari media digital dapat diolah menjadi informasi yang berharga untuk memperkuat pemetaan kerawanan pemilu. Dengan demikian, penggabungan analisis semantik pada data media sosial dengan indikator kerawanan pemilu menjadi penting untuk memperkaya strategi pencegahan potensi gangguan demokrasi.

Selain pada ranah opini publik, analisis kemiripan juga digunakan untuk mengevaluasi proses bisnis internal lembaga pengawas pemilu. Hasil penelitian terkait Standar Operasional Prosedur (SOP) Bawaslu menunjukkan bahwa terdapat kesesuaian maupun perbedaan yang dapat diukur secara struktural maupun semantic [13]. Hal ini memberikan dasar penting bagi penyelarasan prosedur pengawasan agar lebih konsisten di semua tingkatan, baik pusat maupun daerah.

Untuk menganalisis opini publik di media sosial, digunakan teknik perhitungan kemiripan semantik, yaitu metode untuk mengukur kesamaan makna antara teks yang berbeda. Di sinilah teknologi Natural Language Processing (NLP) memainkan peran penting. Dua model bahasa berbasis transformer yang menonjol adalah IndoBERT dan XLNet. IndoBERT merupakan adaptasi dari arsitektur BERT untuk bahasa Indonesia, yang dilatih dengan pendekatan masked language modeling. Model ini mampu memahami konteks kata secara bidirectional dan telah menunjukkan kinerja unggul pada berbagai tugas NLP berbahasa Indonesia[14].

Sementara itu, XLNet dikembangkan dengan pendekatan berbeda, yaitu permutation language modeling, yang memungkinkan model memprediksi kata dengan mempertimbangkan semua kemungkinan urutan kata. Pendekatan ini memberikan keunggulan dalam menangkap relasi kata tanpa keterbatasan masking seperti pada BERT. Studi sebelumnya menunjukkan bahwa XLNet seringkali lebih baik dalam memahami konteks panjang dan kompleks, meskipun dengan biaya komputasi yang lebih tinggi.

Perbandingan kedua model ini dalam konteks analisis semantik di media sosial Indonesia masih relatif terbatas. Mengingat kompleksitas bahasa Indonesia yang kaya akan morfologi, variasi dialek, serta penggunaan bahasa gaul di media sosial, penelitian ini menjadi relevan untuk melihat sejauh mana IndoBERT dan XLNet mampu mendeteksi kesamaan makna komentar publik dengan indikator kerawanan pemilu. Perbandingan ini penting dilakukan karena kedua model tersebut memiliki karakteristik arsitektur dan pendekatan pelatihan yang berbeda, sehingga dapat menghasilkan performa yang bervariasi dalam memahami konteks semantik bahasa Indonesia. IndoBERT yang dilatih secara *masked language modeling* lebih fokus pada pemahaman konteks dua arah (*bidirectional context understanding*), sedangkan XLNet dengan pendekatan *permutation language modeling* mampu mempertimbangkan seluruh kemungkinan urutan kata dalam kalimat. Dengan membandingkan kedua metode ini, penelitian ini bertujuan untuk

mengidentifikasi model yang paling efektif dan akurat dalam menangkap makna semantik teks tidak terstruktur seperti komentar media sosial.

Hasil perbandingan ini diharapkan dapat memberikan kontribusi empiris terhadap pengembangan model pemrosesan bahasa alami (NLP) untuk bahasa Indonesia serta menjadi acuan bagi lembaga pengawas pemilu, seperti Bawaslu, dalam menentukan metode analisis yang paling tepat untuk mendeteksi potensi kerawanan politik berdasarkan opini publik di ruang digital.

Dengan demikian, penelitian ini bertujuan untuk mengkaji dan membandingkan efektivitas IndoBERT dan XLNet dalam perhitungan kemiripan semantik antara komentar media sosial Twitter (X) dengan indikator IKP Pemilu 2024. Hasil penelitian diharapkan dapat memberikan kontribusi signifikan terhadap strategi mitigasi risiko kerawanan pemilu serta memperluas pemahaman akademis mengenai interaksi antara media sosial, opini publik, dan stabilitas demokrasi di Indonesia.

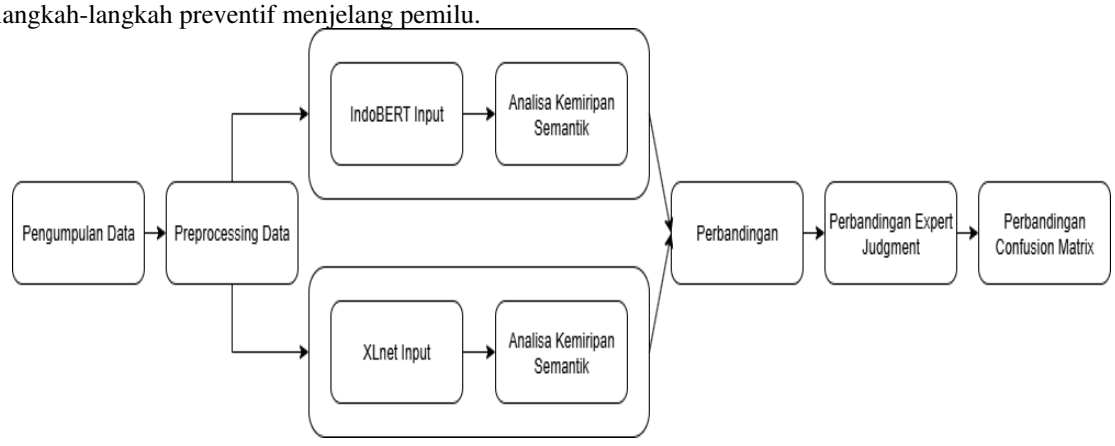
2. Metode Penelitian

Data tweet yang berhubungan dengan pemilu 2024 . Pengambilan sampel data tweet dilakukan secara acak pada tanggal 29 Agustus 2024. Pengambilan sampel data menggunakan kata kunci “pemilu 2024” dan mendapatkan data tweet sebanyak 574 tweet.

Data Indeks Kerawanan Pemilu 2024 ini mencakup indikator-indikator yang digunakan untuk mengukur kerawanan pemilu, seperti stabilitas politik, partisipasi pemilih, aspek teknis pemilu, dan kondisi sosial. Indikator-indikator ini diambil dari laporan resmi dan survei yang dilakukan oleh lembaga seperti Komisi Pemilihan Umum (KPU), Badan Pengawas Pemilihan Umum (Bawaslu), dan organisasi non-pemerintah yang memantau pemilu[3].

Crawling Data Proses pengambilan data menggunakan library Tweepy. Penggunaan library ini membutuhkan akses berupa OAuth 1 yang bisa didapat dari website Twitter Developer. Penggunaan OAuth 1 pada library ini mendapatkan izin untuk melakukan akses 10 menggunakan API yang tersedia. Library ini memanfaatkan API Twiiter Berupa search yang bisa mencari tweet yang memiliki kecocokan dengan kata kunci yang diberikan. Peneliti memberikan kata kunci berupa pemilu 2024, Dari hasil pengambilan data tersebut yang berjumlah 574 Data tweet yang kemudian disimpan dalam format CSV

Indeks Kerawanan Pemilu (IKP) Identifikasi indikator yang relevan dan representatif untuk mengukur tingkat kerawanan suatu daerah dalam menghadapi pemilu. Indikator dapat mencakup aspek sosial, politik, dan keamanan. Merancang metode pengumpulan data yang sesuai untuk setiap indikator. Metode ini dapat meliputi survei lapangan, analisis data historis, dan pemetaan potensi konflik. Melakukan survei lapangan untuk mengumpulkan data primer yang diperlukan untuk mengukur indikator-indikator yang telah ditentukan sebelumnya. Survei lapangan dapat dilakukan melalui wawancara, kuesioner, atau observasi langsung. Mengumpulkan data sekunder yang relevan dari sumber-sumber yang tersedia, seperti lembaga statistik, laporan pemerintah, dan riset terdahulu. Menganalisis data yang telah terkumpul untuk mengukur tingkat kerawanan pemilu pada setiap daerah. Analisis ini dapat dilakukan menggunakan metode statistik atau analisis kualitatif, tergantung pada jenis data yang digunakan.Memetakan data yang telah dianalisis ke dalam sebuah indeks yang dapat mengukur tingkat kerawanan pemilu pada setiap daerah. Indeks ini dapat disusun menggunakan metode tertentu, seperti analisis faktor atau weighting. Melakukan validasi terhadap indeks yang telah disusun untuk memastikan bahwa indeks tersebut dapat memberikan gambaran yang akurat tentang tingkat kerawanan pemilu pada setiap daerah. Menyajikan hasil indeks kerawanan pemilu kepada publik dan pemangku kepentingan terkait untuk digunakan sebagai acuan dalam mengambil langkah-langkah preventif menjelang pemilu.



Gambar 1. Tahapan Penelitian.

Proses analisis kemiripan semantik dilakukan dengan menginput *dataset* yang akan dibandingkan. Setelah melalui tahapan *preprocessing*, teks hasil pembersihan dimasukkan ke dalam model IndoBERT untuk menghasilkan representasi vektor dari setiap teks. Representasi vektor ini menggambarkan makna kontekstual dari setiap kata maupun kalimat, sehingga dapat digunakan untuk mengukur tingkat kemiripan makna (*semantic similarity*) antara komentar media sosial dan indikator pada Indeks Kerawanan Pemilu (IKP). Model IndoBERT bekerja dengan pendekatan *masked language modeling*, di mana sistem mempelajari hubungan antar-kata dalam dua arah sekaligus (*bidirectional context*), sehingga lebih mampu memahami makna yang tersembunyi di antara konteks kalimat.

Langkah yang sama juga diterapkan pada model XLNet, di mana teks hasil *preprocessing* diubah menjadi vektor numerik melalui mekanisme *permutation language modeling*. Berbeda dengan IndoBERT yang berfokus pada konteks dua arah, XLNet mempertimbangkan seluruh kemungkinan urutan kata dalam kalimat untuk memprediksi kata berikutnya. Pendekatan ini memungkinkan model memahami konteks yang lebih luas dan menangkap hubungan semantik yang lebih kompleks antar kata dalam kalimat. Dengan demikian, kedua model menghasilkan representasi semantik yang berbeda meskipun berasal dari data yang sama, sehingga perbandingan hasilnya dapat memberikan gambaran kinerja yang lebih komprehensif.

Setelah representasi vektor diperoleh, perhitungan kemiripan semantik dilakukan menggunakan metode *Cosine Similarity*. Metode ini dipilih karena mampu mengukur sudut kemiripan antara dua vektor tanpa terpengaruh oleh panjang vektor masing-masing. Nilai *cosine similarity* berada pada rentang 0 hingga 1, di mana nilai mendekati 1 menunjukkan tingkat kemiripan semantik yang tinggi, sedangkan nilai mendekati 0 menunjukkan tidak adanya kemiripan. Rumus dasar yang digunakan dalam penelitian ini adalah:

$$Cosine\ Similarity = \frac{A\ .\ B}{||_A|\ |\ |_B||}$$

dengan A dan B masing-masing merupakan vektor hasil representasi dari model IndoBERT dan XLNet.

Nilai kemiripan yang dihasilkan oleh model kemudian dikategorikan berdasarkan ambang batas (*threshold*) tertentu. Misalnya, apabila nilai *cosine similarity* ≥ 0,5 maka teks dianggap "mirip", sedangkan jika nilainya < 0,5 maka dianggap "tidak mirip". Proses ini menghasilkan data klasifikasi biner yang menunjukkan apakah sebuah komentar memiliki kemiripan makna dengan indikator kerawanan pemilu atau tidak. Klasifikasi ini diperlukan untuk memudahkan proses evaluasi dan perbandingan kinerja antar model.

Confusion Matrix atau Evaluasi bertujuan untuk mengukur kinerja model yang dihasilkan dari proses fine-tuning. Evaluasi dilakukan dengan menggunakan confusion matrix.Metode confusion matrixini terdiri dari 4 jenis nilai yaitu TP (True Positive), FP (False Positive), FN (False Negative), dan TN (True Negative). Nilai-nilai tersebut kemudian digunakan untuk menganalisis hasil kerja model dengan menghitung nilai accuracy, precision, recall, dan f1-score.

Akurasi bertujuan untukmengukur persentase sampel yang diprediksi dengan benar dalam data tertentu. Perhitungan accuracydapat dilihat pada Persamaan 1

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \quad (1)$$

Precision adalah tingkat ketepatan antara kasus yang diprediksi positif dan hasil yang positif benar sesuai data sebenarnya. Perhitungan precisiondapat dilihat pada Persamaan

$$Precision = \frac{Tp}{Tp+Fp} \quad (2)$$

Recall adalah tingkat keberhasilan kasus positif dari data sebenarnya diprediksi secara positif dengan benar. Perhitungan recalldapat dilihat pada Persamaan

$$Recall = \frac{Tp}{Tp+Fp} \quad (3)$$

F1-score adalah metrik evaluasi yang menggabungkan nilai precisiondan recallsecara bersamaan. Perhitungan f1-scoredapat dilihat pada Persamaan

$$F1\ Score = 2x\frac{Recall\ x\ Precision}{Recall\ +Precision} \quad (4)$$

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data yang berhubungan dengan komentar publik tentang Pemilu 2024 yang diambil melalui teknik Crawling Data. Pengambilan sampel data dilakukan secara acak pada tanggal 29 Agustus 2024 menggunakan kata kunci "Pemilu2024". Dari proses tersebut

diperoleh sebanyak 574 data tweet yang kemudian disimpan dalam format CSV untuk tahap pengolahan lebih lanjut.

Teknik Crawling Data merupakan metode pengumpulan data otomatis yang digunakan untuk mengambil informasi dari situs web atau media sosial menggunakan skrip atau program tertentu. Dalam penelitian ini, proses crawling dilakukan dengan memanfaatkan Application Programming Interface (API) dari platform Twitter (sekarang X), yang memungkinkan peneliti mengakses data publik secara terstruktur. Proses crawling dilakukan dengan menentukan beberapa parameter, seperti kata kunci (keyword), rentang waktu pengambilan data, serta batas jumlah data (limit).

Setiap tweet yang diperoleh melalui proses crawling mencakup beberapa elemen penting seperti isi teks komentar, tanggal unggahan, username anonim, serta informasi tambahan seperti jumlah suka dan retweet. Data mentah tersebut kemudian dikumpulkan dalam satu berkas dengan format Comma Separated Values (CSV) untuk memudahkan proses preprocessing dan analisis selanjutnya.

Penggunaan teknik Crawling Data dipilih karena mampu mengumpulkan data dalam jumlah besar secara cepat dan efisien, sekaligus memastikan bahwa data yang digunakan bersumber langsung dari opini publik yang aktual di media sosial. Dengan demikian, data yang diperoleh dapat merepresentasikan persepsi masyarakat secara real time terhadap isu Pemilu 2024, sehingga relevan untuk digunakan dalam analisis kemiripan semantik dengan Indeks Kerawanan Pemilu.

3.1. Preprocessing Data

Selanjutnya adalah preprocessing pada tahapan ini data hasil dari tahapan proses crawling data akan diproses hingga menjadi data yang bersih, Berikut tahapan dalam melakukan preprocessing:

3.1.1 Pembersihan Data

Pembersihan data bertujuan untuk menghilangkan karakter yang Mengurangi kualitas data, seperti link, nama akun X (Twitter) Seseorang dan tanda hashtag. Tabel 1 menunjukan hasil tahapan Pembersihan Data.

Tabel 1. Contoh Hasil Pembersihan Data

Teks	Hasil Pembersihan Data
Pilkada Serentak 2024 adalah momen untuk memilih masa depan bukan untuk perpecahan jagapersatuan untuknkri #Pilkada2024 #Pemilu2024 #PilkadaAman https://t.co/VuwLrQW613	pilkada serentak adalah momen untuk memilih masa depan bukan untuk perpecahan jagapersatuan untuknkri pilkada pemilu pilkadaaman

3.1.2 Tokenizing

Tokenizing adalah tahapan pemisahan teks kalimat menjadi potongan per kata sesuai spasi dalam teks. Tabel 2. menunjukan hasil tahapan tokenizing.

Tabel 2. Contoh Hasil Tokenizing

Teks	Hasil Pembersihan Data
pilkada serentak adalah momen untuk memilih masa depan bukan untuk perpecahan jagapersatuan untuknkri pilkada pemilu pilkadaaman https://t.co/VuwLrQW613	['pilkada', 'serentak', 'adalah', 'momen', 'untuk', 'memilih', 'masa', 'depan', 'bukan', 'untuk', 'perpecahan', 'jagapersatuan', 'untuknkri', 'pilkada', 'pemilu', 'pilkadaaman']

3.1.3 Normalisasi

Normalisasi adalah proses yang berkaitan dengan model data relational untuk mengorganisasi himpunan data dengan ketergantungan dan keterkaitan yang tinggi atau erat Tabel 3. menunjukan hasil Normalisasi:

Tabel 3. Contoh Hasil Normalisasi

Teks	Hasil Pembersihan Data
['pilkada', 'serentak', 'adalah', 'momen', 'untuk', 'memilih', 'masa', 'depan', 'bukan', 'untuk', 'perpecahan', 'jagapersatuan', 'untuknkri', 'pilkada', 'pemilu', 'pilkadaaman']	['pilkada', 'serentak', 'adalah', 'momen', 'untuk', 'memilih', 'masa', 'depan', 'bukan', 'untuk', 'perpecahan', 'jagapersatuan', 'untuknkri', 'pilkada', 'pemilu', 'pilkadaaman']

3.1.4 Stemming

Stemming adalah tahap perubahan sebuah kata ke dalam bentuk kata dasarnya. Jadi pada tahap ini, kata-kata pada data yang sudah di tokenisasi dan dinormalisasi akan diubah ke dalam bentuk dasarnya dengan cara menghilangkan imbuhan pada kata tersebut. Tabel 4. menunjukan hasil Stemming**:**

Tabel 4. Contoh Hasil Stemming

Teks	Hasil Pembersihan Data
['pilkada', 'serentak', 'adalah', 'momen', 'untuk', 'memilih', 'masa', 'depan', 'bukan', 'untuk', 'perpecahan', 'jagapersatuan', 'untuknkri', 'pilkada', 'pemilu', 'pilkadaaman']	pilkada serentak adalah momen untuk pilih masa depan bukan untuk pecah jagapersatuan untuknkri pilkada milu pilkadaaman

3.2 Perhitungan Kemiripan Semantik IndoBERT

Menginput proses data yang akan direview, data yang digunakan untuk perhitungan semantik adalah Data yang telah melalui preprocessing setelah data diproses Data dimasukkan ke dalam perhitungan metode IndoBERT untuk mendapatkan hasil kemiripan semantik dari teks dan indikator indeks kerawanan Pemilu 2024 (IKP 2024) lalu datanya di simpan dalam tabel di file excel dengan format xlsx. Seperti pada Input IndoBERT di bawah ini :

1. Gunakan library pandas, transformers, torch, dan scipy untuk pemrosesan data dan perhitungan jarak cosine.
2. Inisialisasi model dan tokenizer IndoBERT
   a. Tentukan model_name = 'indobenchmark/indobert-base-p1'
   b. Muat tokenizer dari model_name
   c. Muat model dari model_name
3. Buat fungsi get_embedding(teks)
   a. Tokenisasi teks dan ubah ke bentuk tensor
   b. Lakukan forward pass ke model IndoBERT tanpa perhitungan gradien
   c. Ambil rata-rata dari last hidden state sebagai representasi vektor (embedding)
   d. Kembalikan embedding sebagai hasil fungsi
4. Buat fungsi calculate_similarity(embedding1, embedding2)
   a. Hitung nilai cosine similarity antara kedua embedding
   b. Kembalikan nilai similarity
5. Baca file Excel (misalnya 'stemm.xlsx') untuk mendapatkan teks utama
   a. Ambil kolom 'clean_stemm' dan ubah ke dalam bentuk daftar (list)
   b. Hapus entri kosong (NaN)
6. Tentukan daftar teks pembanding (misalnya “61 Indikator Indeks Kerawanan Pemilu 2024.”)

7. Siapkan variabel:
   a. processed_texts → untuk menyimpan teks yang sudah diproses
   b. results → untuk menyimpan hasil kemiripan
8. Untuk setiap teks utama:
   a. Jika teks sudah pernah diproses, lewati langkah ini
   b. Tandai teks sebagai sudah diproses
   c. Dapatkan embedding dari teks utama
   d. Bandingkan dengan setiap teks pembanding:
      i. Dapatkan embedding teks pembanding
      ii. Hitung skor cosine similarity
      iii. Simpan teks pembanding dengan skor kemiripan tertinggi
9. Simpan hasil ke dalam tabel (DataFrame) yang berisi:
   - Teks Utama
   - Teks Paling Mirip
   - Skor Kemiripan
10. Ekspor hasil ke file Excel (hasil_kemiripan.xlsx)")

Setelah dilakukannya Perhitungan Semantik dengan IndoBert dan di sortir data tweet yang sama maka data yang sebelumnya berjumlah 574 berkurang menjadi 229 dan langsung otomatis memunculkan juga Skor kemiripanya pada tabel 5. Menunujukan Hasil Kemiripan Semantik yang didapat.

Tabel 5. Skor Kemiripan Semantik IndoBERT

Indikator	Tweet	Skor Kemiripan Semantik
Adanya penghitungan suara ulang di Pemilu/Pilkada	pilkada serentak adalah momen untuk pilih masa depan bukan untuk pecah jagapersatuan untuknkri pilkada milu pilkadaaman	0,519894791

3.3 Perhitungan Kemiripan Semantik XLnet

Menginput proses data yang akan direview, data yang digunakan untuk perhitungan XLnet sama dengan data yang digunakan dalam perhitungan Bert yang telah melalui tahap preprocessing setelah data diproses Data dimasukkan ke dalam perhitungan metode XLnet untuk mendapatkan hasil kemiripan semantik dari teks dan indikator indeks kerawanan Pemilu 2024 (IKP 2024) lalu datanya di simpan dalam tabel di file excel dengan format xlsx. Seperti Input XLnet di bawah ini.

1. Gunakan library pandas, transformers, torch, dan scipy untuk pemrosesan data serta perhitungan jarak cosine.
2. Inisialisasi model dan tokenizer XLNet
   a. Tentukan model_name = 'xlnet-base-cased'
   b. Muat tokenizer dari model_name
   c. Muat model dari model_name
3. Buat fungsi get_embedding(teks)
   a. Tokenisasi teks dan ubah ke bentuk tensor
   b. Lakukan forward pass ke model XLNet tanpa perhitungan gradien
   c. Ambil rata-rata (mean) dari last hidden state sebagai representasi vektor (embedding)
   d. Kembalikan embedding sebagai hasil fungsi
4. Buat fungsi calculate_similarity(embedding1, embedding2)
   a. Hitung nilai cosine similarity antara kedua embedding menggunakan rumus:
      similarity = 1 - cosine_distance(embedding1, embedding2)
   b. Kembalikan nilai similarity
5. Baca file Excel (misalnya 'stemm.xlsx') untuk mendapatkan teks utama
   a. Ambil kolom 'clean_stemm' dan ubah ke dalam bentuk daftar (list)
   b. Hapus entri kosong (NaN) untuk memastikan tidak ada data kosong
6. Tentukan daftar teks pembanding
   a. Misalnya: “61 Indikator Indeks Kerawanan Pemilu 2024.”
7. Siapkan variabel:

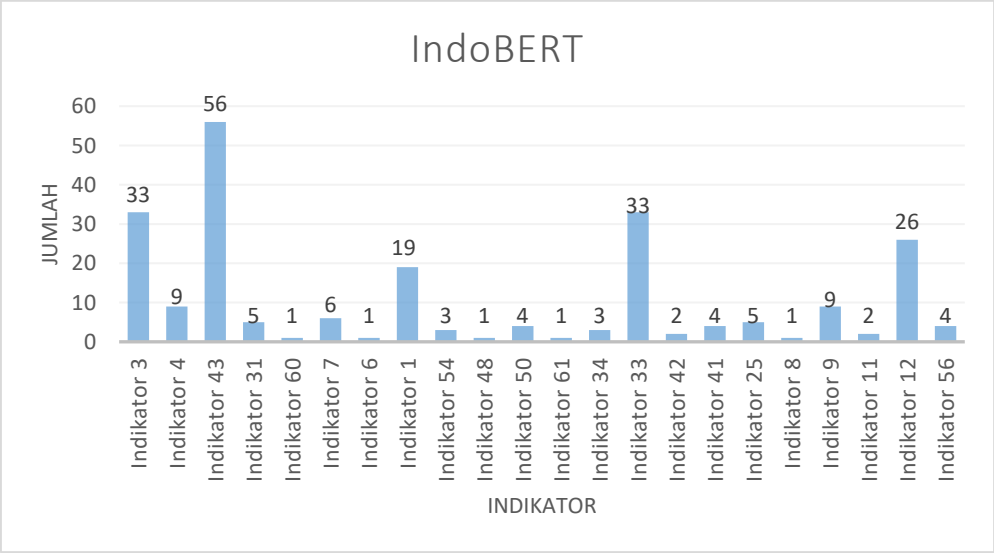
a. processed_texts → untuk menyimpan teks yang sudah diproses
b. results → untuk menyimpan hasil perbandingan kemiripan semantik
8. Untuk setiap teks utama dalam daftar teks:
a. Jika teks sudah pernah diproses, lewati langkah ini
b. Tandai teks sebagai sudah diproses
c. Dapatkan embedding dari teks utama menggunakan fungsi get_embedding()
d. Bandingkan dengan setiap teks pembanding:
i. Dapatkan embedding teks pembanding
ii. Hitung skor cosine similarity antara dua embedding
iii. Simpan teks pembanding dengan skor kemiripan tertinggi
9. Simpan hasil ke dalam tabel (DataFrame) dengan kolom:
- Teks Utama
- Teks Paling Mirip
- Skor Kemiripan
10. Ekspor tabel hasil ke file Excel (hasil_kemiripan_xlnet.xlsx)

Setelah dilakukannya Perhitungan Semantik dengan XLnet dan di sortir data tweet yang sama maka data yang sebelumnya berjumlah 574 berkurang menjadi 229 dan langsung otomatis memunculkan juga Skor kemiripanya pada Tabel 6. Hasil Skor Kemiripan Semantik yang didapat

Tabel 6. Hasil Skor Kemiripan Semantik Xlnet

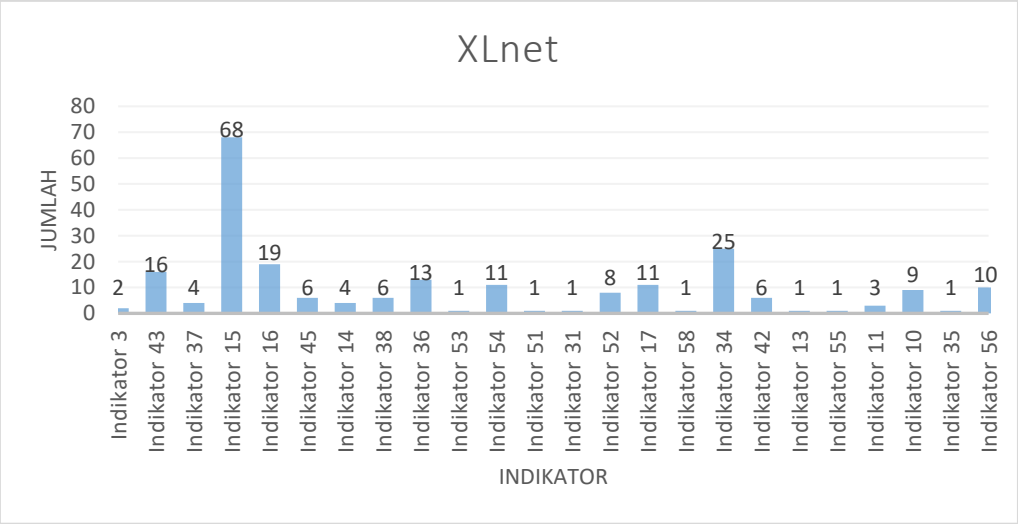
Indikator	Tweet	Skor Kemiripan Semantik
Adanya himbauan dan/atau tindakan untuk menolak calon tertentu dari pemerintah lokal atau tokoh masyarakat	pilkada serentak adalah momen untuk pilih masa depan bukan untuk pecah jagapersatuan untuknkri pilkada milu pilkadaaman	0,991106984

3.4 Perbandingan



Gambar 4. IndoBERT

Pada Gambar 4. IndoBERT menunjukan berapa banyak jumlah kemunculan tiap indikator dan indikator apa saja yang muncul pada Metode IndoBert



Gambar 5. Xlnet

Pada Gambar 5. XLnet menunjukan berapa banyak jumlah lemunculan tiap indikator dan indikator apa saja yang muncul pada Metode XLnet.

3.5 Perbandingan Expert Judgment

Data yang telah dibandingkan akan di lihat mirip tidaknya oleh Expert Judgment apakah data dari kedua metode tersebut benar – benar mirip atau tidak data tweet(Teks Utama) dan IKP 2024(Teks Mirip dan Tidak Mirip). Data yang di periksa adalah 20% dari keseluruhan data. Bisa di lihat pada Tabel 7. IndoBERT dan Tabel 8. XLnet

Tabel 7. IndoBERT

Teks Utama	Teks Paling Mirip	Skor Kemiripan Tertinggi	Teks Tidak Mirip	Skor Kemiripan Terendah	Mirip	Tidak Mirip
pilkada serentak adalah momen untuk pilih masa depan bukan untuk pecah jagapersatuan untuknkri pilkada milu pilkadaaman	Adanya penghitungan suara ulang di Pemilu/Pilkada	0,519894838	Adanya himbauan dan/atau tindakan untuk menolak calon tertentu dari pemerintah lokal atau tokoh masyarakat	0,305020928	Tidak	Tidak

Tabel 8. Xlnet

Teks Utama	Teks Paling Mirip	Skor Kemiripan Tertinggi	Teks Tidak Mirip	Skor Kemiripan Terendah	Mirip	Tidak Mirip
pilkada serentak adalah momen untuk	Adanya himbauan dan/atau tindakan untuk menolak	0,991106868	Adanya materi kampanye Hoax di	0,968133807	Mirip	Tidak

pilih masa depan bukan untuk pecah jagapersatuan untuknkri pilkada milu pilkadaaman	calon tertentu dari pemerintah lokal atau tokoh masyarakat	tempat umum

3.6 Confusion Matrix

Tabel 9. Hasil Confusion Matrix

	IndoBERT	**XLnet**
TP	25	32
TN	30	38
FP	21	14
FN	16	8

Dari Hasil di atas maka kita bisa menghitung Accuracy, Recall, Precision dan F1 Scorenya seperti di bawah ini :

IndoBERT

Akurasi keseluruhan dihitung untuk menilai seberapa baik Metode IndoBERT dalam melakukan perhitungan kemiripan semantik antar kata. Akurasi keseluruhan ini dinyatakan sebagai persentase dari total prediksi yang benar dibandingkan dengan jumlah total data. Dengan True Positives (TP) untuk positif = 25, True Negatif (TN) = 30, False Positives (FP) = 21, dan False Negatif (FN) =16 . Untuk menghitung accuracy dapat dilihat pada persamaan di bawah ini :

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} = \frac{25 + 30}{25 + 30 + 21 + 16} = \frac{55}{92} = 0{,}597\ x\ 100\% = 59{,}7\%$$

Hasil akurasi keseluruhan sebesar 59,7% menunjukkan bahwa Metode IndoBert berhasil melakukan perhitungan 59,7% dari data dengan benar sesuai kata yang seharusnya.

Untuk menentukan nilai precision, pertama-tama perlu ditetapkan nilai True Positive (TP), yaitu jumlah kemiripan mirip yang berhasil dideteksi oleh Metode IndoBERT dan dianggap mirip juga oleh Expert Judgment, yakni sebanyak 25. Sedangkan False Positive (FP) merupakan komen yang tidak mirip menurut Expert Judgment dan dihitung mirip oleh IndoBERT, yaitu sebesar 21. Untuk menghitung precision dapat dilihat pada persamaan di bawah ini :

$$Precision = \frac{Tp}{Tp + Fp} = \frac{25}{25 + 21} = \frac{25}{46} = 0{,}543\ x\ 100\% = 54{,}3\%$$

Hasil precision sebesar 54,3% ini menunjukkan bahwa dari semua prediksi yang dihitung benar oleh metode ternyata memang benar mirip menurut Expert Judgment juga.

Recall dihitung untuk mengukur seberapa baik Metode IndoBERT dalam mengidentifikasi semua data yang sebenarnya positif atau mirip. Nilai recall didapatkan dengan rumus di mana TP adalah jumlah *True Positives* dan FN adalah jumlah *False Negatives* . Berdasarkan data yang ada, diperoleh nilai TP sebesar 25 dan FN sebesar 16.

$$Recall = \frac{Tp}{Tp + Fn} = \frac{25}{25 + 16} = \frac{25}{41} = 0{,}609\ x\ 100\% = 60{,}9\%$$

Hasil recall sebesar 60,9% ini menunjukkan bahwa metode IndoBERT mampu mendeteksi sebagian besar data yang benar-benar mirip dengan cukup baik. Dengan recall yang cukup baik, model ini efektif dalam mengurangi jumlah data positif atau mirip yang luput dari perhitungan, sehingga memberikan jaminan bahwa mayoritas data positif mirip dapat teridentifikasi secara akurat oleh sistem.

Dan F1 Score digunakan untuk mengukur seberapa baik keseimbangan antara precision dan recall dari hasil deteksi sistem. Precision menunjukkan seberapa banyak hasil yang dinyatakan mirip oleh sistem ternyata memang benar menurut Expert Judgment, sedangkan recall menunjukkan seberapa banyak data yang benar-benar mirip berhasil dikenali oleh sistem.

$$F1\ Score = 2x\frac{Recall\ x\ Precision}{Recall + Precision} = 2x\frac{0{,}609\ x\ 0{,}543}{0{,}609 + 0{,}543} = \frac{0{,}661}{1{,}152} = 0{,}573\ x\ 100\% = 57{,}3\%$$

Pada perhitungan ini, precision-nya adalah 54,3% dan recall-nya 60,9%. Kedua nilai tersebut dimasukkan ke dalam rumus F1 Score, yaitu 2 dikali hasil perkalian precision dan recall, lalu dibagi dengan jumlah keduanya. Hasil akhirnya adalah 57,3%, yang berarti sistem memiliki kinerja yang cukup baik dalam mendeteksi kemiripan sesuai dengan penilaian dari Expert Judgment.

XLnet

Akurasi keseluruhan dihitung untuk menilai seberapa baik Metode XLnet dalam melakukan perhitungan kemiripan semantik antar kata. Akurasi keseluruhan ini dinyatakan sebagai persentase dari total prediksi yang benar dibandingkan dengan jumlah total data. Dengan True Positives (TP) untuk positif = 32, True Negatif (TN) = 38, False Positives (FP) = 14, dan False Negatif (FN) = 8 . Untuk menghitung accuracy dapat dilihat pada persamaan di bawah ini :

$$Accuracy\ = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} = \frac{32 + 38}{32 + 38 + 14 + 8} = \frac{70}{92} = 0{,}760\ x\ 100\% = 76\%$$

Hasil akurasi keseluruhan sebesar 76% menunjukkan bahwa Metode XLnet berhasil melakukan perhitungan 76% dari data dengan benar sesuai kata yang seharusnya.

Recall dihitung untuk mengukur seberapa baik Metode XLnet dalam mengidentifikasi semua data yang sebenarnya positif atau mirip. Nilai recall didapatkan dengan rumus di mana TP adalah jumlah *True Positives* dan FN adalah jumlah *False Negatives* . Berdasarkan data yang ada, diperoleh nilai TP sebesar 32 dan FN sebesar 8.

$$Recall\ = \frac{Tp}{Tp + Fn} = \frac{32}{32 + 8} = \frac{32}{40} = 0{,}8\ x\ 100\% = 80\%$$

Hasil recall sebesar 80% ini menunjukkan bahwa metode XLnet mampu mendeteksi sebagian besar data yang benar-benar mirip dengan cukup baik. Dengan recall yang cukup baik, model ini efektif dalam mengurangi jumlah data positif atau mirip yang luput dari perhitungan, sehingga memberikan jaminan bahwa mayoritas data positif mirip dapat teridentifikasi secara akurat oleh sistem.

Untuk menentukan nilai precision, pertama-tama perlu ditetapkan nilai True Positive (TP), yaitu jumlah kemiripan mirip yang berhasil dideteksi oleh Metode XLnet dan dianggap mirip juga oleh Expert Judgment, yakni sebanyak 32. Sedangkan False Positive (FP) merupakan komen yang tidak mirip menurut Expert Judgment dan dihitung mirip oleh XLnet, yaitu sebesar 14. Untuk menghitung precision dapat dilihat pada persamaan di bawah ini :

$$Precision\ = \frac{Tp}{Tp + Fp} = \frac{32}{32 + 14} = \frac{32}{46} = 0{,}695\ x\ 100\% = 69{,}5\%$$

Hasil precision sebesar 69,5% ini menunjukkan bahwa dari semua prediksi yang dihitung benar oleh metode ternyata memang benar mirip menurut Expert Judgment juga.

Dan F1 Score digunakan untuk mengukur seberapa baik keseimbangan antara precision dan recall dari hasil deteksi sistem. Precision menunjukkan seberapa banyak hasil yang dinyatakan mirip oleh sistem ternyata memang benar menurut Expert Judgment, sedangkan recall menunjukkan seberapa banyak data yang benar-benar mirip berhasil dikenali oleh sistem.

$$F1\ Score = 2x\frac{Recall\ x\ Precision}{Recall + Precision} = 2x\frac{0{,}80\ x\ 0{,}695}{0{,}80 + 0{,}695} = \frac{1{,}112}{1{,}495} = 0{,}743\ x\ 100\% = 74{,}3\%$$

Pada perhitungan ini, precision-nya adalah 69,5% dan recall-nya 80%. Kedua nilai tersebut dimasukkan ke dalam rumus F1 Score, yaitu 2 dikali hasil perkalian precision dan recall, lalu dibagi dengan jumlah keduanya. Hasil akhirnya adalah 74,3%, yang berarti sistem memiliki kinerja yang cukup baik dalam mendeteksi kemiripan sesuai dengan penilaian dari Expert Judgment.

Sebelum dilakukan evaluasi kuantitatif terhadap hasil keluaran model, tahap awal dilakukan pengukuran manual menggunakan metode Expert Judgment. Pada tahap ini, sejumlah pakar yang

memahami konteks sosial, politik, dan bahasa dalam isu Pemilu 2024 diminta untuk menilai tingkat kemiripan makna antara komentar publik di media sosial dengan indikator pada Indeks Kerawanan Pemilu (IKP). Penilaian dilakukan secara independen dengan memberikan label "mirip" atau "tidak mirip" terhadap setiap pasangan teks berdasarkan pemahaman kontekstual, relevansi isu, serta potensi makna implisit dalam bahasa publik. Hasil penilaian manual ini kemudian dijadikan sebagai ground truth yang menjadi acuan pembanding dalam evaluasi kinerja model IndoBERT dan XLNet.

Pendekatan Expert Judgment memiliki peran penting karena pakar dapat mempertimbangkan nuansa semantik yang sulit ditangkap oleh sistem berbasis kecerdasan buatan. Misalnya, suatu komentar dengan nilai kemiripan semantik rendah secara numerik mungkin tetap dianggap relevan oleh pakar karena memiliki implikasi politik atau sosial yang signifikan. Dengan demikian, hasil evaluasi manual memberikan dasar yang lebih kontekstual dan realistis untuk menilai kemampuan model dalam memahami makna bahasa alami masyarakat Indonesia di media sosial.

Berdasarkan hasil perhitungan yang dibandingkan dengan Expert Judgment, model XLNet menunjukkan performa yang lebih unggul dibandingkan IndoBERT dalam seluruh metrik evaluasi. IndoBERT mencapai akurasi keseluruhan sebesar 59,7%, sedangkan XLNet mencapai akurasi yang jauh lebih tinggi yaitu 76%. Dalam hal precision, recall, dan F1-score, keduanya menunjukkan hasil yang cukup baik, namun XLNet secara konsisten mencatatkan nilai yang lebih tinggi.

Nilai precision untuk IndoBERT sebesar 54,3%, sementara XLNet mencapai 69,5%, menunjukkan bahwa XLNet lebih akurat dalam mengidentifikasi pasangan teks yang benar-benar mirip. Pada metrik recall, IndoBERT memperoleh 60,9% sedangkan XLNet mencapai 80%, menandakan bahwa XLNet lebih baik dalam mendeteksi semua pasangan teks yang relevan. Adapun F1-score menunjukkan perbedaan yang signifikan antara kedua model, di mana IndoBERT memperoleh 57,3%, sedangkan XLNet mencapai 74,3%.

Secara keseluruhan, hasil ini menunjukkan bahwa XLNet lebih unggul dibandingkan IndoBERT dalam melakukan perhitungan kemiripan semantik terhadap data komentar publik dengan indikator kerawanan Pemilu. Performa XLNet yang lebih baik dapat dijelaskan oleh kemampuannya dalam memahami konteks bahasa yang kompleks melalui pendekatan permutation language modeling, yang memungkinkan model mempertimbangkan semua kemungkinan urutan kata dalam kalimat. Dengan demikian, dalam penelitian ini XLNet terbukti lebih efektif dan akurat dalam menangkap makna semantik komentar publik terkait kata kunci "Pemilu 2024" dibandingkan dengan IndoBERT.

4. Kesimpulan

Berdasarkan hasil dari penelitian ini, maka telah didapatkan kesimpulan sebagai berikut:

a. Perhitungan kemiripan data Tweet dengan kata kunci "pemilu 2024" dimulai dari processing data, dan setelah itu di lihat tidak miripnya oleh Expert Judgment di bandingkan dengan data utama yaitu tweet dengan Indikator IKP tahun 2024 dan dilakukan perhitungan menggunakan confusion matrix untuk melihat akurasinya dan score yang di dapatkan dari adalah IndoBERT Accuracy 59,7%, Precision 54,3%, Recall 60,9% dan F1 Scorenya 57,3% Sedangkan XLnet adalah Accuracy 76%, Precision 69,5%, Recall 80%, dan F1 Scorenya 74,3%.

b. Dengan menggunakan confusssion matrix dan juga validasi dari ahli bahasa, hasil akurasi score yang didapatkan berdasarkan perhitungan menggunakan XLnet adalah 76%. Sedangkan hasil akurasi score berdasarkan perhitungan menggunakan IndoBERT yaitu 57,3%. Maka dapat disimpulkan bahwa nilai akurasi menggunakan metode XLnet lebih besar daripada menggunakan metode IndoBERT.

Daftar Pustaka

[1] Suhariyanto Didik, "Pengabdian Masyarakat Pemetaan Kerawanan Pemilihan Umum (Pemilu) Serentak Tahun 2024 (Pencegahan Kampanye Politisasi Sara, Hoax Dan Ujaran Kebencian)," *Communnity Dev. J.*, vol. 5, no. 1, pp. 2348–2352, 2024, [Online]. Available: https://journal.universitaspahlawan.ac.id/index.php/cdj/article/view/25792

[2] Bawaslu, "Pemutakhiran Indeks Kerawanan Pemilu (IKP) Pilkada 2020," 2020.

[3] Bawaslu RI, "Indeks Kerawanan Pemilu Dan Pemilihan Serentak 2024," *Bawaslu RI*, pp. 1–23, 2023,[Online].Available:https://ppidapp.bawaslu.go.id/api/services/file/public/dip/3319/1718353397618-Indeks Kerawanan Pemilu 2024.pdf

[4] D. Irawan, "Kampung Pengawasan Partisipatif dan Road Map Indeks Kerawanan Pemilu di Kabupaten Indramayu," *J. Adhyasta Pemilu*, vol. 5, no. 1, pp. 19–31, 2022, doi: 10.55108/jap.v5i1.85.

[5] B. S. Utomo and I. Irwansyah, "Peran Media Sosial dalam Gerakan Menolak Penundaan Pemilu di Indonesia," *J. Polit. Indones.*, vol. 8, no. 2, pp. 108–128, 2023, doi: 10.35706/jpi.v8i2.10214.

[6] S. Aeni, “Analisis Opini Publik Terhadap Pemilu 2024 Pada Media Sosial X 1 Tafana Destiana Larassetya, 2 Arfian Suryasuciramdhan , 3 Nuril Ulia Salsa , 4 Ira,” *Sos. dan Hum.*, no. 2, pp. 292–301, 2024, [Online]. Available: https://doi.org/10.47861/tuturan.v2i2.994

[7] Jayus, Sumaiyah, Desy Mairita, and Assyari Abdullah, “Media Sosial sebagai Media Kampanye Politik Menjelang Pemilu 2024,” *J. SIMBOLIKA Res. Learn. Commun. Study*, vol. 10, no. 1, pp. 72–81, 2024, doi: 10.31289/simbolika.v10i1.11468.

[8] A. S. S. Pratama, E. Y. M. S. Danan, G. Angeline, I. Z. Abdillah, J. A. Prayoga, and N. A. M. Jabari, “A roadmap for the successful use of social media in electoral campaigns,” *Bull. Soc. Informatics Theory Appl.*, vol. 5, no. 1, pp. 28–37, 2021, doi: 10.31763/businta.v5i1.430.

[9] A. Danaditya, L. H. X. Ng, and K. M. Carley, “From curious hashtags to polarized effect: profiling coordinated actions in indonesian twitter discourse,” *Soc. Netw. Anal. Min.*, vol. 12, no. 1, pp. 1–24, 2022, doi: 10.1007/s13278-022-00936-2.

[10] I. Tangkawarow, “Analisis Sentimen Pada Sosial Media Twitter Terhadap Pemilu 2024 Dengan Metode Support Vector Machine,” *J. Ilm. Wahana Pendidik.*, vol. 11, no. 5.A, pp. 315–330, 2025, [Online]. Available: http://ai-soc.org/ijcs/index.php/ijcs/article/view/4777/1032

[11] I. Tangkawarow and S. Mawikere, “Perbandingan Akurasi SVM & Naive Bayes Pada Analisis Sentimen Data Berita Online Terhadap COKLIT Pemilu 2024,” vol. 14, no. 2, pp. 3159–3169, 2025.

[12] I. Tangkawarow and I. W. Suardi, “Perbandingan Nilai Akurasi Analisa Sentiment Pada Kata Kunci Pemilu 2024,” vol. 14, no. 2, pp. 3105–3118, 2025.

[13] A. Kenap and N. Abram, “Analisis Kemiripan Menggunakan Metode BERT dan Jaccard Pada Proses Bisnis Bawaslu RI,” *Indones. J. Comput. Sci.*, vol. 14, no. 3, pp. 5088–5098, 2025, doi: 10.33022/ijcs.v14i3.4890.

[14] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.