

PENGARUH TAHAPAN PREPROCESSING TERHADAP MODEL INDOBERT DAN INDOBERTWEET UNTUK MENDETEKSI EMOSI PADA KOMENTAR AKUN BERITA INSTAGRAM

Ulfia Khairani¹, Viska Mutiawani^{*2}, Hendri Ahmadian³

^{1,2}Universitas Syiah Kuala, Banda Aceh

³Universitas Islam Negeri Ar-Raniry, Banda Aceh

Email: ¹ulfia199@mhs.usk.ac.id, ²viska.mw@usk.ac.id, ³hendri@ar-raniry.ac.id

*Penulis Korespondensi

(Naskah masuk: 18 Desember 2023, diterima untuk diterbitkan: 12 Agustus 2024)

Abstrak

Platform media sosial seperti Instagram telah membentuk ruang di mana berita dapat dengan mudah ditemukan dan menarik perhatian individu. Pada Instagram, dapat memberikan komentar-komentar terhadap berita yang telah dibaca. Pemahaman terhadap emosi yang mengiringi komentar-komentar yang telah diberikan pengguna pada postingan berita dapat membantu memahami bagaimana berita tersebut diserap, diinterpretasi, dan direspons oleh publik. Penelitian ini mengkategorikan empat emosi yaitu marah, senang, takut, dan sedih dengan menggunakan model terlatih IndoBERT dan IndoBERTweet. Penelitian ini bertujuan untuk membandingkan model IndoBERT dan IndoBERTweet dalam mendeteksi emosi pada komentar akun berita Instagram dan mengeksplorasi dampak penggunaan tahapan *preprocessing* khususnya *remove stopwords* dan *stemming* pada kedua model. Hasil penelitian menunjukkan bahwa model yang tidak melalui tahapan *remove stopwords* dan *stemming* menghasilkan kinerja yang lebih baik dibandingkan model yang melalui tahapan *remove stopwords* dan *stemming*, dengan perolehan akurasi sebesar 92,54% untuk model IndoBERTweet dan 88,81% untuk model IndoBERT.

Kata kunci: deteksi emosi, tahapan preprocessing, indoBERT, indoBERTweet

THE INFLUENCE OF PREPROCESSING STAGES ON INDOBERT AND INDOBERTWEET MODELS FOR EMOTION DETECTION IN INSTAGRAM NEWS ACCOUNTS COMMENTS

Abstract

Social media platforms such as Instagram have created a space where news can be easily discovered and attract the attention of individuals. On Instagram, people can provide comments on the news they have read. Understanding the emotions that accompany the comments that users have given on news posts can help understand how the news is absorbed, interpreted and responded to by the public. This research categorizes four emotions, anger, happiness, fear and sadness, using pre-trained models IndoBERT and IndoBERTweet. This research aims to compare the IndoBERT and IndoBERTweet models in detecting emotions in Instagram news account comments and explore the impact of preprocessing stages, especially removing stopwords and stemming on both models. The research results showed that the model that did not go through the remove stopwords and stemming stages produced better performance than the model that went through the remove stopwords and stemming stages, with an accuracy of 92.54% for the IndoBERTweet model and 88.81% for the IndoBERT model.

Keywords: emotion detection, preprocessing stages, IndoBERT, IndoBERTweet

1. PENDAHULUAN

Saat ini, lingkungan jaringan media sosial membentuk wadah dimana berita dapat lebih mudah ditemukan dan menarik perhatian individu (Realita and Setiadi, 2022). Tidak hanya berita besar yang mendapat sorotan, tetapi juga berita lokal dan topik khusus yang dapat dengan mudah diakses dan disebarluaskan. Kini, masyarakat tidak hanya

mengandalkan media tradisional seperti televisi dan surat kabar lagi, tetapi juga berpaling kepada platform-platform media sosial untuk mendapatkan informasi terkini seperti Instagram.

Pada media sosial khususnya instagram, masyarakat tidak hanya dapat membaca berita saja namun juga dapat memberikan komentar dan opini terhadap berita yang telah dibaca. Komentar-

komentar tersebut mencerminkan keragaman sudut pandang dan tanggapan publik terhadap berbagai isu aktual. Komentar-komentar yang muncul tidak hanya berisi kata-kata biasa tetapi juga merupakan cara bagi masyarakat mengungkapkan pikiran dan perasaannya tentang berita-berita tersebut. Komentar-komentar yang diberikan memiliki potensi untuk mencerminkan reaksi dan emosi mereka terhadap berita tersebut.

Emosi memiliki pengaruh yang kuat terhadap bagaimana manusia membuat keputusan dan juga berfungsi sebagai alat bantu yang mendorong komunikasi efektif dengan dunia. Deteksi emosi menjadi sebuah proses yang melibatkan pengenalan dan identifikasi beragam jenis perasaan atau emosi yang dirasakan oleh seseorang, contohnya kebahagiaan, kesedihan, atau kemarahan (Nandwani and Verma, 2021). Kemampuan mendeteksi emosi memiliki dampak dalam mendukung pengambilan keputusan di beragam ranah, seperti manajemen dan pemasaran, interaksi pengguna, aspek keuangan, politik, kesehatan, serta pendidikan (Nisa, Amriza and Supriyadi, 2021). Dalam konteks berita, pemahaman terhadap emosi yang mengiringi komentar-komentar tersebut dapat membantu memahami bagaimana informasi diserap, diinterpretasi, dan direspons oleh publik. Oleh karena itu, analisis emosi yang terkandung dalam komentar-komentar menjadi penting dalam mengurai dinamika respons publik terhadap berita.

Penelitian tentang deteksi emosi telah dilakukan dengan menggunakan berbagai model, seperti pada penelitian yang dilakukan oleh Haryadi dan Kusuma (2019) dimana deteksi emosi dilakukan dengan menggunakan LSTM dan Nested-LSTM serta membandingkannya dengan algoritma SVM dimana hasil dari penelitian menunjukkan bahwa Nested LSTM dan LSTM memberikan kinerja yang lebih baik daripada SVM (Haryadi and Kusuma, 2019). Penelitian serupa juga dilakukan oleh Adoma et al. (2020) yang membandingkan *pre-trained* model BERT, RoBERTa, DistilBERT dan XLNET dalam tugas mengklasifikasikan emosi. Hasil penelitian tersebut menunjukkan bahwa model RoBERTa memberikan hasil terbaik dibandingkan dengan model lain yaitu sebesar 0,7431 (Adoma, Henry and Chen, 2020). Penelitian menggunakan BERT dalam mendeteksi emosi dan analisis sentiment juga dilakukan oleh Chiorrini et al. (2021) dimana model mencapai akurasi sebesar 0,92 untuk analisis sentimen dan 0,90 untuk pengenalan emosi (Chiorrini et al., 2021).

Proses *preprocessing* juga menjadi hal yang sangat penting dalam upaya melakukan tugas klasifikasi teks termasuk dalam mendeteksi emosi. Penelitian yang terkait dengan *preprocessing* pada teks telah dilakukan pada penelitian oleh İşik dan Dağ (2020) yang menganalisis teknik *preprocessing* dan pengaruhnya dalam memprediksi rating ulasan menggunakan beberapa algoritma *machine learning*.

Hasil penelitian menunjukkan bahwa penghapusan *stopwords*, penggunaan huruf kecil, penghapusan kata kata umum dan kombinasi 1-to-3 *n-grams* memiliki performa yang lebih baik daripada teknik *preprocessing* lain yang telah diuji (İşik and Dağ, 2020). Selain itu, penelitian yang dilakukan oleh Khomsah et al. (2020) menerapkan berbagai model *preprocessing* pada analisis sentiment dan kemudian melihat pengaruhnya pada akurasi model *classifier*. Hasil pada penelitian tersebut menunjukkan bahwa *cleaning data*, *stopwords*, dan konversi kosakata *slang* menjadi kosakata standar KBBI meningkatkan akurasi sekitar 3,5 % pada kata *unigram* (Khomsah et al., 2020).

Di Indonesia, model bahasa seperti IndoBERT dan IndoBERTweet dapat dimanfaatkan untuk melakukan tugas klasifikasi teks termasuk mendeteksi emosi pada teks. IndoBERT merupakan model yang telah dilatih sebelumnya (*pre-trained*) menggunakan algoritma BERT untuk bahasa Indonesia (Wilie et al., 2020). Sementara IndoBERTweet adalah *pre-trained* model yang dilatih dengan memperluas model BERT berbahasa Indonesia (Koto, Lau and Baldwin, 2021).

Dalam memproses data teks terutama komentar pada media sosial, *preprocessing* memberikan peran yang sangat penting. *Preprocessing* digunakan untuk mengurangi *noise* dan membuat data lebih terstruktur serta efisien. Salah satu tahapan *preprocessing* yaitu *remove stopwords* dan *stemming*. *Remove stopwords* merupakan tahapan untuk menghilangkan kata-kata umum yang tidak memberikan banyak informasi penting dalam analisis teks (Miyajiwala et al., 2022). Sedangkan proses *stemming* adalah tahapan untuk menyederhanakan kata-kata menjadi bentuk dasarnya (Rosid et al., 2020). Namun dalam konteks penggunaan IndoBERT dan IndoBERTweet penting untuk melihat bagaimana *remove stopwords* dan *stemming* mempengaruhi efektivitas model karena proses tersebut dapat mempengaruhi kemampuan model dalam mengenali makna kata yang lebih rumit dan kontekstual dalam bahasa Indonesia.

Berdasarkan penjelasan yang telah diuraikan, maka dilakukan pengujian untuk mengetahui bagaimana pengaruh tahapan *preprocessing* terutama *remove stopwords* dan *stemming* terhadap model IndoBERT dan IndoBERTweet. Tujuan dari penelitian ini adalah menerapkan model IndoBERT dan IndoBERTweet untuk mendeteksi emosi pada komentar akun berita Instagram serta mengeksplorasi dampak penggunaan tahapan *remove stopwords* dan *stemming* pada kedua model tersebut. Harapannya, temuan penelitian ini juga dapat digunakan sebagai referensi untuk menghadirkan konten yang lebih relevan secara emosional, serta dapat membantu memahami bagaimana cara masyarakat merespons berita tertentu.

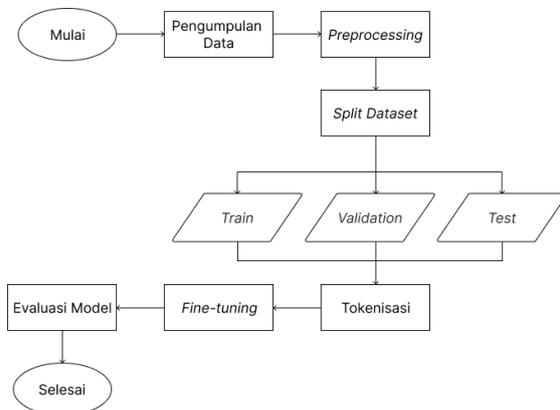
2. METODE PENELITIAN

Penelitian dilakukan dengan melalui beberapa tahapan sebagaimana ditunjukkan pada Gambar 1.

2.1 Pengumpulan Data

Penelitian ini memanfaatkan data yang diperoleh dari interaksi pengguna dalam bentuk komentar pada akun berita di platform Instagram. Komentar-komentar dikumpulkan dari postingan dua akun berita yaitu @kumparancom (<https://www.instagram.com/kumparancom/>) dan @detikcom (<https://www.instagram.com/detikcom/>) dari bulan April 2023 hingga bulan Agustus 2023. Proses untuk mengumpulkan data komentar dilakukan dengan menggunakan teknik *scraping*, dimana informasi diambil secara otomatis dari halaman website. *Scraping* dilakukan dengan menggunakan Selenium dan Chromedriver. Selenium digunakan untuk mengoperasikan *browser* secara otomatis sebagai respons terhadap perintah pengguna. Selenium dapat melakukan berbagai tugas yang dilakukan oleh pengguna saat menjelajahi seperti mengklik tombol, membuka tab baru, dan mengakses halaman web (Levi *et al.*, 2020). Sedangkan Chrome Driver digunakan untuk membuka Google Chrome dan mengontrolnya secara otomatis (Hudha *et al.*, 2022).

Data yang berhasil dikumpulkan yaitu sebanyak 5.002 komentar.



Gambar 1. Metode Penelitian

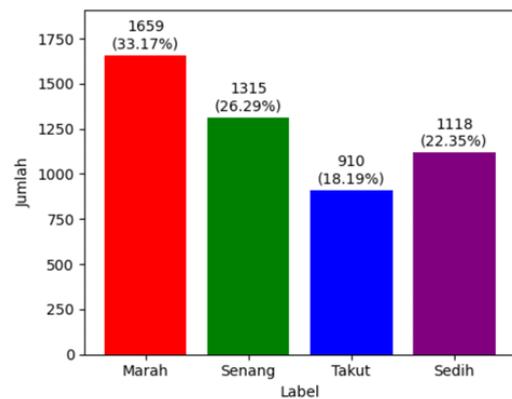
2.2 Pelabelan Data

Proses pelabelan data dilakukan secara manual dengan melakukan analisis mendalam terhadap setiap komentar untuk mengenali emosi yang paling dominan. Dalam penelitian ini, terdapat empat label emosi yang digunakan yaitu marah, sedih, senang, dan takut. Tujuan dari pelabelan data adalah untuk mendapatkan kualitas data yang baik.

Data yang sudah dikumpulkan pada tahap sebelumnya diberi label oleh penulis dan satu orang sarjana psikologi. Pemberian label pada setiap data yang telah dikumpulkan dilakukan secara pribadi. Hasil pelabelan kedua belah pihak diperiksa dan dibandingkan setelah pelabelan selesai. Jika terdapat perbedaan label, maka dilakukan diskusi untuk

mencapai kesepakatan. Proses ini bertujuan untuk memastikan bahwa penafsiran emosi dari setiap komentar tidak bias dan sudah sesuai dengan teori psikologi yang berlaku.

Distribusi emosi pada data yang sudah diberi label seperti pada Gambar 2. Emosi marah merupakan emosi terbanyak pada dataset dengan persentase 33,2% yang mengindikasikan bahwa postingan berita cenderung menciptakan emosi marah pada pengguna. Selanjutnya, emosi senang mencapai 26,3%, emosi sedih 22,4% dan emosi takut 18,2%.



Gambar 2. Distribusi Emosi pada Data

2.3 Preprocessing

Preprocessing melibatkan serangkaian tindakan untuk menyiapkan dan menyesuaikan data teks sebelum dilakukan analisis lebih lanjut. *Preprocessing* data dilakukan untuk membersihkan data agar data tersebut menjadi lebih bersih dan terstruktur sehingga lebih efisien dan efektif untuk dilakukan analisis lebih lanjut. Proses yang dilakukan pada tahap *preprocessing* adalah *case folding*, *data cleaning*, normalisasi, *remove stopwords* dan *stemming*.

2.4 Split Dataset

Pada proses *split dataset*, data yang sudah diberi label akan dipartisi menjadi tiga bagian, yaitu data latih (*train*), data uji (*test*), dan data validasi (*validation*). Data latih digunakan untuk mengembangkan model, data uji digunakan untuk menguji dan mengevaluasi akurasi model, sementara data validasi digunakan untuk memvalidasi kinerja model dan mengurangi risiko *overfitting* (Imaduddin, A'la and Nugroho, 2023). Proporsi split dataset pada penelitian ini yaitu 70% data *train*, 15% data *validation*, dan 15% data *test*. Distribusi label setelah dilakukan *split dataset* dapat dilihat pada Tabel 1.

Tabel 1. Distribusi label setelah *split dataset*

Label	Train	Validation	Test
Marah	1.161	249	249
Senang	920	197	198
Takut	637	137	136
Sedih	783	167	168
Total	3.501	750	751

2.5 Tokenisasi

Dataset yang digunakan harus disesuaikan dengan format *input* yang diterima oleh model, baik itu *IndoBERT* ataupun *IndoBERTweet*. Untuk itu, diperlukan penambahan token khusus yaitu token [CLS] pada awal untuk memberi tahu model bahwa proses klasifikasi sedang dilakukan, token [SEP] pada tiap akhir kalimat dan token [PAD] untuk menyesuaikan panjang kalimat menjadi sama dengan panjang maksimum urutan (Li, Zhang and Zhou, 2020). Pada penelitian ini panjang urutan kalimat maksimum diatur menjadi 98 sesuai dengan panjang maksimal teks pada dataset. Jika panjang suatu kalimat melebihi batas maksimum yang ditetapkan, maka kalimat akan dipotong (*truncate*). Sebaliknya, jika panjang kalimat kurang dari batas maksimum yang telah ditentukan, padding akan dilakukan dengan menambahkan token [PAD].

Proses tokenisasi dilakukan dengan menggunakan *tokenizer* milik *IndoBERT* untuk model *IndoBERT* dan *tokenizer* milik *IndoBERTtweet* untuk model *IndoBERTtweet*. *Tokenizer* akan memeriksa apakah setiap kata dalam kalimat terdapat dalam kamus (*vocabulary*). Jika tidak ditemukan, kata tersebut akan dipecah menjadi sub kata (*subword*) dengan menggunakan simbol *##*. Selanjutnya, dilakukan pengkodean (*encoding*) menggunakan *tokenizer* sesuai dengan indeks *vocabulary* dari *pre-trained* *IndoBERT* dan *IndoBERTtweet*.

2.6 Fine-tuning

Fine-tuning adalah suatu metode di mana model yang sudah ada diambil dan diubah sedikit agar sesuai dengan tugas atau dataset tertentu (Gupta, 2021). Sebelum melakukan *fine-tuning* maka *hyperparameter* harus diatur terlebih dahulu.

Dalam mencari nilai *hyperparameter* yang optimal, digunakan teknik pencarian dimana beberapa kombinasi nilai *hyperparameter* diuji untuk mengidentifikasi *hyperparameter* mana yang menghasilkan model terbaik. Berdasarkan penelitian terkait BERT, terdapat tiga parameter utama yang dapat disesuaikan untuk mengoptimalkan kinerja pada tahap *fine-tuning*, yaitu *batch size*, tingkat pembelajaran (*learning rate*), dan jumlah *epoch* (Devlin *et al.*, 2019).

Nilai *hyperparameter* yang diterapkan sebagaimana disebutkan pada Tabel 2. *Hyperparameter* dipilih berdasarkan pada hasil uji coba yang telah dilakukan untuk memastikan bahwa nilai tersebut memberikan hasil yang optimal.

Tabel 2. *Hyperparameter* yang digunakan

<i>Hyperparameter</i>	Nilai
<i>Max sequence length</i>	98
<i>Batch size</i>	16
<i>Dropout</i>	0,5
<i>Optimizer</i>	AdamW
<i>Learning rate</i>	2e-5
<i>Epoch</i>	10

Nilai *learning rate* biasanya berkisar antara 0 dan 1, dengan nilai yang lebih tinggi mempercepat pelatihan tetapi tidak menjamin hasil yang lebih baik. Nilai *learning rate* yang tepat bervariasi dari kasus ke kasus karena nilai yang terlalu tinggi atau terlalu kecil dapat mengarah pada solusi yang tidak optimal. Oleh karena itu, untuk mencapai hasil terbaik, sangat penting untuk memilih nilai *learning rate* dengan hati-hati (Imaduddin, A'la and Nugroho, 2023).

2.7 Evaluasi Model

Evaluasi bertujuan untuk mengukur kinerja model yang dihasilkan dari proses *fine-tuning*. Evaluasi dilakukan dengan menggunakan *confusion matrix*. Metode *confusion matrix* ini 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* (Wang *et al.*, 2021). Akurasi bertujuan untuk mengukur persentase sampel yang diprediksi dengan benar dalam data tertentu. Perhitungan *accuracy* dapat 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 *precision* dapat dilihat pada Persamaan 2.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall adalah tingkat keberhasilan kasus positif dari data sebenarnya diprediksi secara positif dengan benar. Perhitungan *recall* dapat dilihat pada Persamaan 3.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score adalah metrik evaluasi yang menggabungkan nilai *precision* dan *recall* secara bersamaan. Perhitungan *f1-score* dapat dilihat pada Persamaan 4.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Pada penelitian ini terdapat dua skenario. Pada skenario pertama, dataset yang digunakan adalah dataset yang melalui semua proses *preprocessing*. Sedangkan pada skenario kedua, dataset yang digunakan adalah dataset yang hanya melalui proses *case folding*, *data cleaning* dan normalisasi tanpa melibatkan tahap *remove stopwords* dan *stemming*.

3.1 Preprocessing Output

Tahapan pertama dalam *preprocessing* yaitu *case folding*. Langkah *case folding* digunakan untuk mengubah seluruh huruf dalam teks menjadi huruf kecil (Sari and Ruldeviyani, 2020). Tujuannya untuk

mencegah adanya perbedaan dalam pengenalan antara huruf besar dan huruf kecil dalam data. Hasil dari tahapan case folding dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Case Folding*

<i>Input</i>	<i>Output</i>
@nadiemmakarim Izin dgn segala hormat, tolong diusust tuntas sistem kebijakan pendidikan kita pak, skrg hanya perkara iuran kas yg udh disetujui udh berani membunuh org !!!,	@nadiemmakarim izin dgn segala hormat, tolong diusust tuntas sistem kebijakan pendidikan kita pak, skrg hanya perkara iuran kas yg udh disetujui udh berani membunuh org !!!,

Tahapan kedua adalah *data cleaning*. Pada proses *data cleaning*, dilakukan pembersihan data untuk menghilangkan unsur-unsur yang tidak terlalu penting untuk penelitian ini. Tahapan *data cleaning* yang dilakukan diantaranya menghapus *tag*, *hashtag*, URL, *mention (username)*, tanda baca, angka, emoji, spasi berlebih, dan menghilangkan pengulangan karakter yang sama. Hasil dari tahapan sebelumnya diproses untuk *data cleaning*, seperti yang ditunjukkan dalam Tabel 4.

Tabel 4. Hasil *Data Cleaning*

<i>Input</i>	<i>Output</i>
@nadiemmakarim izin dgn segala hormat, tolong diusust tuntas sistem kebijakan pendidikan kita pak, skrg hanya perkara iuran kas yg udh disetujui udh berani membunuh org !!!,	izin dgn segala hormat tolong diusust tuntas sistem kebijakan pendidikan kita pak skrg hanya perkara iuran kas yg udh disetujui udh berani membunuh org

Tabel 5. Hasil Normalisasi

<i>Input</i>	<i>Output</i>
izin dgn segala hormat tolong diusust tuntas sistem kebijakan pendidikan kita pak skrg hanya perkara iuran kas yg udh disetujui udh berani membunuh org	izin dengan segala hormat tolong diusust tuntas sistem kebijakan pendidikan kita pak sekarang hanya perkara iuran kas yang sudah disetujui sudah berani membunuh orang

Tahapan ketiga adalah normalisasi. Normalisasi dilakukan dengan mengganti kata-kata *slang* atau tidak formal dalam kalimat menjadi kata-kata yang lebih umum atau sesuai dengan kamus yang tersedia. Tahapan ini bertujuan untuk memastikan konsistensi dalam gaya penulisan kata atau frasa spesifik. Proses normalisasi teks pada penelitian ini menggunakan leksikon colloquial-indonesian-lexicon.csv (Aliyah Salsabila *et al.*, 2018). Kamus yang digunakan berisi 151.167 kata *slang* beserta artinya. Hasil dari tahap sebelumnya diproses untuk normalisasi, seperti yang ditunjukkan dalam Tabel 5.

Tahapan keempat adalah *remove stopwords*. *Remove stopwords* bertujuan untuk menghilangkan kata-kata umum yang tidak memberikan banyak informasi penting dalam analisis teks (Miyajiwala *et al.*, 2022). Tahapan ini menggunakan *library stopwords* yang disediakan oleh *library Sastrawi*. Sastrawi merupakan *library NLP* yang dirancang khusus untuk Bahasa Indonesia (Albab, Karuniawati

P and Fawaiq, 2023). Hasil dari tahapan sebelumnya diproses untuk *remove stopwords*, seperti yang ditunjukkan dalam Tabel 6.

Tabel 6. Hasil *Remove Stopwords*

<i>Input</i>	<i>Output</i>
izin dengan segala hormat tolong diusust tuntas sistem kebijakan pendidikan kita pak sekarang hanya perkara iuran kas yang sudah disetujui sudah berani membunuh orang	izin hormat tolong diusust tuntas sistem kebijakan pendidikan hanya perkara iuran kas disetujui berani membunuh

Tahapan terakhir yaitu *stemming*. Proses *stemming* bertujuan untuk menyederhanakan kata-kata menjadi bentuk dasarnya (Rosid *et al.*, 2020). Pada penelitian ini, *stemming* dilakukan dengan menggunakan *library Sastrawi*. Hasil dari tahapan sebelumnya diproses untuk *stemming*, seperti yang ditunjukkan dalam Tabel 7.

Tabel 7. Hasil *Stemming*

<i>Input</i>	<i>Output</i>
izin hormat tolong diusust tuntas sistem kebijakan pendidikan perkara iuran kas disetujui berani membunuh	izin hormat tolong diusust tuntas sistem bijak didik perkara iur kas tuju berani bunuh

3.2 Visualisasi Data

Visualisasi data pada teks dilakukan untuk membantu memperjelas dan mempermudah pemahaman informasi yang terkandung dalam teks. *Word cloud* merupakan teknik analisis teks yang menampilkan representasi grafis dari frekuensi kata-kata dalam teks asli. Semakin besar ukuran kata dalam visualisasi, semakin sering kata tersebut muncul dalam dokumen.



Gambar 3. *Word Cloud* Emosi Marah

Pada Gambar 3 menunjukkan kata-kata yang sering muncul pada data dengan label emosi marah seperti kata “anak”, “rakyat”, “hukum”, “jabat” dan “negara” yang mencerminkan isu-isu sosial, hukum, dan politik memicu emosi marah pada pengguna. Selain itu terdapat juga kata “bunuh” yang terkait erat dengan pemicu kemarahan pada pengguna.

Pada Gambar 4 menunjukkan kata-kata yang sering muncul pada data dengan label emosi sedih seperti kata “anak”, “korban”, “guru”, dan “keluarga” yang mencerminkan situasi yang melibatkan kesedihan. Selain itu juga terdapat kata “kasihan”, “miris”, “moga” yang mengekspresikan kepedulian dan simpati terhadap sesuatu hal yang menyedihkan.



Gambar 4. Word Cloud Emosi Sedih



Gambar 5. Word Cloud Emosi Senang

Pada Gambar 5 menunjukkan kata kata yang sering muncul pada data dengan label emosi senang seperti kata “keren”, “bagus”, “hebat”, dan “mantap” yang menggambarkan kekaguman atau apresiasi terhadap sesuatu yang dianggap baik. Selain itu juga terdapat kata “alhamdulillah” dan “moga” yang menunjukkan ungkapan syukur dan harapan terhadap berbagai situasi yang memicu perasaan senang.



Gambar 6. Word Cloud Emosi Takut

Pada Gambar 6 menunjukkan kata kata yang sering muncul pada data dengan label takut seperti kata “bahaya”, “takut”, “ngeri”, “seram”, “bunuh”, dan “nyawa” yang menggambarkan kecemasan, ketakutan, atau kekhawatiran yang muncul dalam situasi yang berpotensi membahayakan.

3.3 Evaluasi Model IndoBERT

Hasil deteksi emosi ditampilkan menggunakan *confusion matrix* untuk menentukan kinerja model IndoBERT dalam setiap skenario yang ada. Tabel 8 menampilkan *confusion matrix* pada model IndoBERT skenario pertama yang menunjukkan bahwa model mampu memprediksi emosi marah dengan baik yaitu sebesar 218 dari 249 data, namun model sedikit sulit memisahkan antara emosi sedih dan marah dimana 20 data sedih salah diprediksi sebagai marah dan 18 data marah salah diprediksi sebagai sedih. Model juga salah memprediksi senang

sebagai marah dimana terdapat 23 data yang sebenarnya masuk ke dalam emosi senang namun diprediksi sebagai marah.

Tabel 8. *Confusion Matrix* IndoBERT Skenario Pertama

Aktual	Prediksi			
	Marah	Senang	Takut	Sedih
Marah	218	6	7	18
Senang	23	166	3	6
Takut	8	1	117	10
Sedih	20	7	1	140

Tabel 9 menampilkan *confusion matrix* pada model IndoBERT skenario kedua. *Confusion matrix* pada model IndoBERT skenario kedua menunjukkan bahwa terdapat peningkatan dalam memprediksi emosi senang dibandingkan skenario pertama dimana terdapat 190 data yang diprediksi dengan benar, sementara pada skenario pertama hanya 166 data dari 198 data yang diprediksi dengan benar. Namun, masih terdapat kesulitan yang sama dalam memisahkan antara emosi sedih dan marah meskipun terdapat sedikit peningkatan dalam memprediksi secara tepat pada skenario kedua.

Tabel 9. *Confusion Matrix* IndoBERT Skenario Kedua

Aktual	Prediksi			
	Marah	Senang	Takut	Sedih
Marah	219	11	4	15
Senang	5	190	1	2
Takut	13	3	115	5
Sedih	14	10	1	143

3.4 Evaluasi Model IndoBERTweet

Hasil deteksi emosi ditampilkan menggunakan *confusion matrix* untuk menentukan kinerja model IndoBERTweet dalam setiap skenario yang ada. Tabel 10 menampilkan *confusion matrix* pada model IndoBERTweet skenario pertama yang menunjukkan bahwa model cenderung lebih baik dalam mengidentifikasi label emosi senang dengan 178 prediksi yang benar dari 198 data . Model juga baik dalam memprediksi label emosi takut dengan 119 prediksi yang benar dari 136 data. Sementara itu, terdapat 25 data dengan label sedih yang salah diklasifikasikan sebagai emosi marah dan 12 data dengan label marah yang salah diklasifikasikan sebagai emosi sedih.

Tabel 10. *Confusion Matrix* IndoBERTweet Skenario Pertama

Aktual	Prediksi			
	Marah	Senang	Takut	Sedih
Marah	228	6	3	12
Senang	14	178	1	5
Takut	7	2	119	8
Sedih	25	5	0	138

Tabel 11 menampilkan *confusion matrix* pada model IndoBERTweet skenario kedua. *Confusion matrix* pada model IndoBERTweet skenario kedua menunjukkan bahwa terjadi peningkatan dalam memprediksi label marah, senang, takut dan sedih dibandingkan dengan skenario pertama.

Tabel 11 *Confusion Matrix* IndoBERTweet Skenario Kedua

Aktual	Prediksi			
	Marah	Senang	Takut	Sedih
Marah	229	2	3	15
Senang	1	194	2	1
Takut	5	0	122	9
Sedih	16	2	0	150

3.5 Perbandingan Hasil Evaluasi

Confusion matrix selanjutnya digunakan untuk menganalisis hasil kerja model dengan menghitung nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Perhitungan nilai akurasi, presisi, recall, dan skor F1 model IndoBERT dan IndoBERTweet pada skenario pertama dan kedua ditampilkan dalam Tabel 12 dan Tabel 13.

Tabel 12. Perbandingan Hasil Evaluasi Model Skenario Pertama

Matriks	IndoBERT	IndoBERTweet
Accuracy	0,8535	0,8828
Precision	0,8628	0,8895
Recall	0,8519	0,8945
F-1 score	0,8563	0,8850

Secara keseluruhan model yang dilatih menggunakan dataset skenario kedua lebih baik daripada dataset skenario pertama. Selain itu, model IndoBERTweet menunjukkan kinerja yang lebih unggul dibandingkan dengan model IndoBERT untuk kedua skenario yang telah diuji karena model IndoBERTweet dilatih menggunakan data dari twitter yang memiliki struktur bahasa yang lebih mirip dengan data komentar pada instagram.

Tabel 13. Perbandingan Hasil Evaluasi Model Skenario Kedua

Matriks	IndoBERT	IndoBERTweet
Accuracy	0,8881	0,9254
Precision	0,8944	0,9275
Recall	0,8840	0,9223
F-1 score	0,8880	0,9245

3.6 Analisa Hasil Prediksi

Kami menganalisis beberapa hasil label emosi yang dihasilkan oleh model yang digunakan dalam penelitian kami. Tabel 14 dan 15 menunjukkan contoh komentar yang memanfaatkan penggunaan IndoBERT dan IndoBERTweet pada skenario pertama dan kedua.

Proses *remove stopwords* dapat mempengaruhi performa model baik itu model IndoBERT maupun IndoBERTweet. *Stopwords* yang ada di dalam kalimat dapat memberikan informasi penting untuk membantu mengidentifikasi emosi dalam teks. Contohnya kata "tidak" dalam kalimat "sekumpulan orang yang tidak bisa bersyukur" mempunyai arti yang negatif. Jika kata *stopwords* pada kalimat tersebut dihilangkan, maka dapat menyebabkan makna penting pada kalimat tersebut hilang yang berdampak pada cara model memahami teks. Contoh lainnya yaitu kata-kata "tetapi", "namun", "belum", "kecuali", "tanpa", "melainkan" yang dianggap sebagai *stopwords* namun sebenarnya dapat memberikan pengaruh dalam memperjelas emosi pada teks.

Tabel 14. Contoh Hasil Prediksi IndoBERT

Skenario	Komentar	Aktual	Prediksi
I	kumpul syukur	Marah	Senang
	naro bawang sehat sehat iya mas	Sedih	Senang
II	sekumpulan orang yang tidak bisa bersyukur	Marah	Marah
	siapa yang naro bawang disini sehat sehat iya mas	Sedih	Senang

Tabel 15. Contoh Hasil Prediksi IndoBERTweet

Skenario	Komentar	Aktual	Prediksi
I	kumpul syukur	Marah	Senang
	naro bawang sehat sehat iya mas	Sedih	Sedih
II	sekumpulan orang yang tidak bisa bersyukur	Marah	Marah
	siapa yang naro bawang disini sehat sehat iya mas	Sedih	Sedih

Selain proses *remove stopwords*, model IndoBERT dan IndoBERTweet juga dipengaruhi oleh tahapan *stemming*. Model IndoBERT dan IndoBERTweet dilatih dengan kumpulan data teks yang besar dan memuat banyak kata-kata. Tidak melakukan proses *stemming* dapat membantu model mendapatkan hasil yang lebih baik dikarenakan *input* yang diberikan kepada model lebih sesuai dengan data yang digunakan ketika model dilatih.

Model IndoBERTweet memiliki akurasi yang lebih baik dibandingkan model IndoBERT dalam memprediksi data yang berasal dari media sosial. Performa model IndoBERTweet yang lebih baik dibandingkan IndoBERT dikarenakan model IndoBERTweet telah dilatih khusus dengan data yang berasal dari media sosial. Model IndoBERTweet telah dioptimalkan untuk memahami teks informal, gaya bahasa, ataupun singkatan-singkatan yang umumnya ditemukan di platform media sosial seperti instagram. Hal ini terlihat pada contoh komentar yang berhasil diprediksi dengan benar oleh model IndoBERTweet namun tidak berhasil diprediksi dengan benar menggunakan model IndoBERT. Sebagai contoh, komentar yang mengandung kata "bawang" berhasil diprediksi sebagai emosi sedih pada IndoBERTweet. Kata "bawang" sering menjadi kiasan pada pengguna media sosial untuk menyampaikan kesedihan.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian, dataset yang melalui tahapan *remove stopwords* dan *stemming* menghasilkan akurasi sebesar 88,28% untuk model IndoBERTweet dan 85,35% untuk model IndoBERT. Sementara itu, dataset yang tidak melalui tahapan *remove stopwords* dan *stemming* menghasilkan akurasi sebesar 92,54% untuk model IndoBERTweet dan 88,51% untuk model IndoBERT. Dari hasil akurasi yang diperoleh, model yang tidak melalui proses *remove stopwords* dan *stemming* lebih baik

dibandingkan dengan model yang melalui tahapan *remove stopwords* dan *stemming*. Selain itu, model IndoBERTweet menunjukkan performa yang lebih baik dibandingkan dengan model IndoBERT.

Pada penelitian selanjutnya disarankan untuk menggunakan kumpulan data yang lebih besar serta menambah cakupan emosi yang dapat diidentifikasi. Selain itu, penelitian selanjutnya dapat mengeksplorasi berbagai kombinasi tahapan preprocessing lainnya untuk melihat pengaruhnya pada *pre-trained model* IndoBERT dan IndoBERTweet.

DAFTAR PUSTAKA

- ADOMA, A.F., HENRY, N.M. and CHEN, W., 2020. Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition. *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020*, pp. 117–121.
- ALBAB, M.U., and FAWAIQ, M. N., 2023. Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic. *Jurnal TRANSFORMATIKA*, 20(2), pp. 1–10.
- SALSABILA, N.A., WINATMOKO, Y.A., SEPTIANDRI, A.A., and JAMAL, A., 2018. Colloquial Indonesian Lexicon. *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*, pp. 226–229.
- CHIORRINI, A. DIAMANTINI, C., MIRCOLI, A., & POTENA, D., 2021. Emotion and sentiment analysis of tweets using BERT. In *EDBT/ICDT Workshops (Vol. 3)*.
- GUPTA, N., 2021. A Pre-Trained Vs Fine-Tuning Methodology in Transfer Learning. *Journal of Physics: Conference Series*, 1947(1).
- HARYADI, D. and KUSUMA, G.P., 2019. Emotion detection in text using nested Long Short-Term Memory. *International Journal of Advanced Computer Science and Applications*, 10(6), pp. 351–357.
- HUDHA, M., SUPRIYATI, E., and Listyorini, T., 2022. Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier. *JIKO (Jurnal Informatika dan Komputer)*, 5(1), pp. 1–6.
- IMADUDDIN, H., A'LA, F.Y. and NUGROHO, Y.S., 2023. Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach. *International Journal of Advanced Computer Science and Applications*, 14(8), pp. 113–117.
- IŞIK, M. and DAĞ, H., 2020. The impact of text preprocessing on the prediction of review ratings. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(3), pp. 1405–1421.
- KHOMSAH, S., and ARIBOWO, A. S., 2020. Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), pp. 648–654.
- KOTO, F., LAU, J.H. and BALDWIN, T., 2021. INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization', *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, pp. 10660–10668.
- LEVI, M., PALIT, H. N., and ROSTIANINGSIH, S., 2020. Perbandingan Performa Tools Web Scraping pada Website dengan Data Statis dan Dinamis. *Jurnal Infra*, 8, pp. 1–7.
- LI, X., ZHANG, H. and ZHOU, X.H., 2020. Chinese clinical named entity recognition with variant neural structures based on BERT methods', *Journal of Biomedical Informatics*, 107(April), p. 103422.
- MIYAJIWALA, A., LADKAT, A., JAGADALE, S., and JOSHI, R., 2022. On Sensitivity of Deep Learning Based Text Classification Algorithms to Practical Input Perturbations. *Lecture Notes in Networks and Systems*, 507 LNNS, pp. 613–626.
- NANDWANI, P. and VERMA, R., 2021. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), pp. 1–19. Nisa, R., Amriza, S. and Supriyadi, D. (2021) 'Komparasi Metode', 13(2), pp. 130–139.
- REALITA, E. and SETIADI, U., 2022. Konsumsi Berita Insidental di Media Sosial pada Generasi Dewasa. *Jurnal Riset Komunikasi*, 5(1), pp. 99–112.
- ROSID, M. A., FITRANI, A. S., ASTUTIK, I. R. I., MULLOH, N. I., and GOZALI, H. A., 2020. Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1).
- SARI, I.C. AND RULDEVIYANI, Y., 2020. Sentiment Analysis of the Covid-19 Virus Infection in Indonesian Public Transportation on Twitter Data: A Case Study of Commuter Line Passengers. In *2020 International Workshop on Big Data and Information Security (IWBI)*. IEEE, pp. 23–28.
- WANG, H., ZHANG, L., YIN, K., LUO, H., and LI, J., 2021. Landslide identification using machine learning. *Geoscience Frontiers*, 12(1), pp. 351–364.
- WILIE, B., VINCENTIO, K., WINATA, G. I., CAHYAWIJAYA, S., LI, X., LIM, Z. Y., ... and PURWARIANTI, A., 2020, IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. pp. 843–857.

Halaman ini sengaja dikosongkan.