

ANALISIS PERBANDINGAN MODEL BERT DAN XLNET UNTUK KLASIFIKASI TWEET BULLY PADA TWITTER

Teuku Radillah^{*1}, Okta Veza², Sarjon Defit³

¹Institut Teknologi Mitra Gama, Kabupaten Bengkalis, ²Universitas Ibnu Sina, Batam,

³Universitas Putra Indonesia YPTK Padang, Padang

Email: ¹t.radillah@gmail.com, ²okta@uis.ac.id, ³sarjon_defit@upiypk.ac.id

*Penulis Korespondensi

(Naskah masuk 26 Agustus 2024, diterima untuk diterbitkan: 26 November 2024)

Abstrak

Fenomena *bullying* di media sosial, khususnya di *Twitter*, telah menjadi isu yang semakin memprihatinkan dengan dampak signifikan terhadap kesehatan mental pengguna. Dalam rangka mengatasi masalah ini, deteksi otomatis *tweet* yang mengandung konten *bullying* menjadi sangat penting. Penelitian ini bertujuan untuk membandingkan performa dua model pemrosesan bahasa alami terbaru, yaitu *BERT* (*Bidirectional Encoder Representations from Transformers*) dan *XLNet*, dalam klasifikasi *tweet* yang mengandung *bullying*. Metodologi penelitian ini melibatkan pengumpulan dataset *tweet* yang telah dilabeli sebagai *bullying* atau *non-bullying*. Proses preprocessing teks dilakukan untuk membersihkan dan menyiapkan data sebelum digunakan dalam pelatihan model. Kedua model, *BERT* dan *XLNet*, dilatih dan diuji menggunakan dataset yang sama. Evaluasi performa dilakukan dengan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa kedua model memiliki kemampuan yang baik dalam mengidentifikasi *tweet bullying*, akan tetapi *XLNet* menunjukkan performa yang lebih unggul dibandingkan *BERT* dengan tingkat akurasi sebesar 95%. Dengan nilai presisi = 100%, *recall* = 0,87%, dan *F1-score* = 0,88%. *XLNet* mampu menangkap konteks dan nuansa bahasa yang lebih kompleks dalam *tweet*, yang berkontribusi pada akurasi klasifikasi yang lebih tinggi. Penelitian ini memberikan kontribusi penting dalam bidang deteksi *bullying* di media sosial dengan menunjukkan bahwa penggunaan model *XLNet* lebih efektif dibandingkan *BERT*. Temuan ini dapat membantu platform seperti *Twitter* dalam mengidentifikasi dan mencegah konten *bullying*, sehingga menciptakan lingkungan *online* yang lebih aman bagi pengguna, serta dapat digunakan sebagai dasar untuk pengembangan sistem deteksi *bullying* yang lebih canggih dan efisien di masa depan.

Kata kunci: Klasifikasi *Tweet Bully* pada *Twitter*, *Text Mining*, Model *Bert* dan *XLNet*

COMPARATIVE ANALYSIS OF BERT AND XLNET MODELS FOR CLASSIFYING BULLYING TWEETS ON TWITTER

Abstract

The phenomenon of bullying on social media, particularly on Twitter, has become an increasingly concerning issue with significant impacts on users' mental health. In order to address this issue, automatic detection of tweets containing bullying content is crucial. This study aims to compare the performance of two recent natural language processing models, namely BERT (Bidirectional Encoder Representations from Transformers) and XLNet, in the classification of tweets containing bullying. The research methodology involves collecting a dataset of tweets that have been labelled as bullying or non-bullying. Text preprocessing is done to clean and prepare the data before it is used in model training. Both models, BERT and XLNet, were trained and tested using the same dataset. Performance evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results show that both models have a good ability to identify bullying tweets, but XLNet shows superior performance compared to BERT with an accuracy rate of 95%. With precision = 100%, recall = 0.87%, and F1-score = 0.88%. XLNet is able to capture more complex context and language nuances in tweets, which contributes to higher classification accuracy. This research makes an important contribution to the field of bullying detection on social media by showing that the use of the XLNet model is more effective than BERT. These findings can help platforms like Twitter identify and prevent bullying content, thereby creating a safer online environment for users, and can be used as a basis for the development of more sophisticated and efficient bullying detection systems in the future.

Keywords: Classifying Bully Tweets on Twitter, Text Mining, Bert and XLNET Models

1. PENDAHULUAN

Media sosial merupakan suatu platform *real-time* yang digunakan oleh pengguna dalam memberikan atau bertukar informasi melalui interaksi pengguna berupa percakapan, pengiriman foto dan video singkat (Anggrainingsih, Hassan and Datta, 2023). Adapun media sosial yang sering digunakan adalah twitter (saat ini berubah nama menjadi X) (Asgari-Chenaghlu *et al.*, 2021). Twitter adalah salah satu media sosial untuk berinteraksi bersama komunitas, teman, kenalan untuk menyampaikan gagasan atau ide, perasaan atau mencari hiburan (Ricko and Sasongko, 2021). Twitter memberikan kebebasan kepada penggunanya untuk menulis apapun disana, sehingga twitter sering digunakan dalam penyebaran informasi yang berkaitan dengan berita atau informasi *ter-update* (Nabila and Setiawan, 2024). Akan tetapi tidak semua berita atau informasi yang beredar di *twitter* semuanya benar, karena adanya beberapa oknum pengguna yang memanfaatkan twitter untuk tweet *bully* yang ditujukan pada publik figur atau individu lainnya dan perbuatan ini sering disebut *Cyberbullying*.

Cyberbullying adalah suatu tindakan yang dilakukan secara berkelompok atau secara individu terhadap seseorang yang mengarah pada merendahkan dan pelecehan perilaku melalui teks, foto, dan video (Yaqin *et al.*, 2023). *Cyberbullying* biasanya dilakukan melalui media sosial perantara dan memberikan dampak psikologis yang serius bagi korban, termasuk stres, depresi, dan kecemasan karena merasa tertekan aib serta privasinya dapat diakses oleh banyak orang (Slamet *et al.*, 2020). Deteksi otomatis tweet yang terindikasi konten *bullying* menjadi suatu kebutuhan prioritas untuk melindungi pengguna dan menciptakan lingkungan *online* yang lebih aman dan nyaman. Dalam beberapa tahun terakhir, kemajuan dalam bidang pemrosesan bahasa alami (NLP) telah memungkinkan pengembangan model-model yang dapat secara efektif mengklasifikasikan teks, termasuk mendeteksi konten *bullying*. Dua model NLP yang menjadi kajian utama dalam penelitian ini adalah *BERT* (*Bidirectional Encoder Representations from Transformers*) dan *XLNet*. *BERT* adalah model yang kuat dan fleksibel yang telah membawa perubahan besar dalam bidang NLP. Pendekatannya yang *bidirectional* dan penggunaan arsitektur transformer memungkinkan model ini untuk memahami konteks bahasa dengan cara yang jauh lebih mendalam dibandingkan model-model sebelumnya (Feng *et al.*, 2022). *BERT* adalah model pemrosesan bahasa alami (NLP) yang dikembangkan oleh *Google* dan diperkenalkan pada tahun 2018 (Hu *et al.*, 2022). *BERT* adalah suatu model representasi bahasa yang didesain untuk merepresentasikan teks yang belum memiliki label dengan memperhatikan konteks yang ada di sekitarnya (Al Farisi, Perdana and Adikara, 2023). *BERT* memperkenalkan beberapa inovasi penting dalam NLP yang telah merevolusi

cara model memahami dan memproses bahasa. Sedangkan *XLNet* adalah sebuah model jaringan saraf dalam (*deep neural network architecture*) yang mengadopsi arsitektur Transformer, dirancang khusus untuk menangani data berurutan seperti teks (Dhivyaa *et al.*, 2023). *XLNet* digunakan untuk menghitung kemiripan antara teks kueri dan konteks ekspresi matematika dari dokumen ilmiah (Yan *et al.*, 2022). *XLNet* telah mencapai prestasi yang luar biasa hasil implementasinya banyak memberikan kontribusi pada tugas bahasa yaitu penalaran bahasa secara alami seperti memberikan dan menjawab pertanyaan, menyortir dokumen, dan analisis emosi (Habbat, Anoun and Hassouni, 2023). Penelitian ini bertujuan untuk membandingkan performa kedua model tersebut dalam tugas klasifikasi *tweet bullying* di *Twitter*. Analisis perbandingan ini, diharapkan dapat diperoleh pemahaman yang lebih baik mengenai keunggulan dan kelemahan masing-masing model, serta memberikan rekomendasi untuk pengembangan sistem deteksi *bullying* yang lebih efektif.

2. TELAAH PUSTAKA

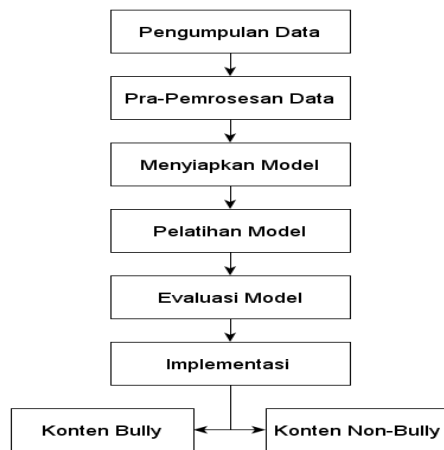
2.1. Penelitian Terdahulu

Penelitian sebelumnya dilakukan (Li *et al.*, 2020) Untuk membandingkan BERT dan XLNet dari sudut pandang karakteristik komputasi. Pada penelitian ini. membandingkan karakteristik komputasi BERT dan XLNet menggunakan *MPRC* (*Microsoft Research Paraphrase Corpus*), sebagai salah satu tolok ukur pemahaman bahasa yang populer. Hasil penelitian ini XLNet mencapai skor benchmark yang lebih baik, dengan biaya operasi aritmatika 1,2 kali lebih banyak dan waktu eksekusi 1,5 kali lebih lama pada CPU modern dibandingkan BERT. Penelitian selanjutnya dilakukan (Adoma, Henry and Chen, 2020), Pada penelitian ini membahas analisis keluaran yang dihasilkan model *BERT*, *RoBERTa*, *DistilBERT*, dan *XLNet pre-trained* dalam mengenali emosi dari teks. Penelitian ini melakukan analisis terhadap perbedaan emosi marah, jijik, kesedihan, ketakutan, rasa malu, kegembiraan dan rasa bersalah. Penelitian serupa juga dilakukan oleh (Arabadzheva-Kalcheva and Kovachev, 2021) yang membandingkan tingkat akurasi model *BERT* dan *XLNet* dalam mengklasifikasi teks. Pada penelitian ini menggunakan 50.000 ulasan dan *XLNet* menempati peringkat atas dengan tingkat akurasi 96%.

Penelitian selanjutnya dilakukan oleh (Gupta, Jinad and Liu, 2023) yang membahas tentang deteksi depresi pada data twitter dengan membandingkan Model NLP yaitu *BERT*, *RoBERTa*, *DistilBERT*, *ALBERT*, *Electra*, dan *XLNet*. Hasil penelitian ini menunjukkan XLNET dan BERT memiliki hasil yang sama yaitu 99% lebih akurat dibandingkan model lainnya.

3. METODE PENELITIAN

Penelitian ini menggunakan dua model yaitu *BERT* dan *XLNet* yang bertujuan untuk membandingkan performa kedua model tersebut dalam tugas klasifikasi *tweet bullying* di *Twitter*. Analisis perbandingan ini, diharapkan dapat memberikan rekomendasi untuk pengembangan sistem deteksi *bullying* yang lebih efektif. Implementasi model tersebut pada klasifikasi *tweet bullying* di *Twitter* menggunakan bahasa pemrograman *python*. Adapun tahapan yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Klasifikasi *tweet bullying* di *Twitter*

Adapun penjelasan yang berkaitan dengan tahapan klasifikasi *tweet bullying* di *Twitter* sebagai berikut :

1. Dataset *cyberbullying_tweets.csv*
Dataset untuk klasifikasi *tweet bullying* berisi kumpulan teks yang disimpan dalam format data *csv*.
2. Pemrosesan Data
Data *tweet bullying* yang sudah dikumpulkan dan diberi pelabelan (*bullying or not bullying*).
3. Menyiapkan Model
Pada model *BERT* persiapan dengan memuat model *pre-trained BERT* yaitu *bert-base-uncased*, selanjutnya mengonversi teks ke dalam format yang dapat diterima oleh *BERT* menggunakan *tokenizer BERT*. Sedangkan model *XLNet* dengan persiapan memuat model *pre-trained XLNet* yaitu, *XLNet-base-cased*. Selanjutnya Mengonversi teks ke dalam format yang dapat diterima oleh *XLNet* menggunakan *tokenizer XLNet*.
4. Pelatihan Model
Pada pelatihan model dengan mempersiapkan data, yaitu mengonversi data teks dan label ke dalam tensor, dan membagi *dataset* menjadi data latih dan data uji. Selanjutnya melakukan *fine-tuning*, yaitu dengan melatih model *BERT* dan *XLNet* pada data latih. Proses ini melibatkan *fine-tuning* model *pre-trained* pada dataset spesifik untuk klasifikasi *bullying*, dan setelah proses *fine-*

tuning selesai dilakukan validasi selama pelatihan untuk menghindari *overfitting*.

5. Evaluasi Model
Evaluasi model dilakukan dengan menggunakan data uji untuk mengevaluasi performa model, dan mengukur metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*.
6. Implementasi
Menggunakan model yang telah dilatih untuk mengklasifikasikan *tweet* baru apakah mengandung *bullying* atau tidak.

3.1. BERT

BERT (Bidirectional Encoder Representations from Transformers) adalah model pemrosesan bahasa alami (NLP) yang dikembangkan oleh *Google* dan diperkenalkan pada tahun 2018 (Hu et al., 2022). *BERT* berbeda dengan model dua arah tradisional karena tidak hanya mempertimbangkan konteks informasi dari sisi kiri dan kanan kalimat, tetapi juga mengintegrasikan informasi konteks dari kedua sisi tersebut di semua lapisan (Li, Wang and Liu, 2021).

Adapun rumus model *BERT* sebagai berikut :

1. Representasi *Input* :

Setiap token dalam teks diubah menjadi vektor *embedding* yang merupakan gabungan dari tiga jenis *embedding* (Santhiya et al., 2024): token *embeddings*, *segment embeddings*, dan *position embeddings*. Representasi *input E* dihitung sebagai :

$$E = E_{(token)} + E_{(segment)} + E_{(position)}$$

2. *Attention Mechanism* :

BERT menggunakan perhatian diri (*self-attention*) untuk memodelkan hubungan antar kata dalam kalimat. Adapun Skor perhatian dihitung sebagai berikut :

$$Attention(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (1)$$

di mana *Q* (*query*), *K* (*key*), dan *V* (*value*) adalah hasil transformasi linier dari representasi *input*, dan d_k adalah dimensi dari *K*

3. *Transformer Layer* :

$$\mathbf{H}^{(l)} = \text{LayerNorm} (\mathbf{H}^{(l-1)} + \text{Attention } \mathbf{H}^{(l-1)})$$

$$\mathbf{H}^{(l)} = \text{LayerNorm} (\mathbf{H}^{(l)} + \text{FFN } \mathbf{H}^{(l)}) \quad (2)$$

Dimana $\mathbf{H}^{(l)}$ = adalah *output* dari lapisan ke-*l*

3.2. XLNet

XLNet adalah metode pra-pelatihan yang menggunakan permutasi tujuan pemodelan bahasa untuk menggabungkan keunggulan Metode Autoregressive (AR) dan Autoencoder (AE) yang

diperkenalkan pada tahun 2019 (Salma, Saptawati and Rusmawati, 2021). Adapun rumus XLNET sebagai berikut:

1. Permutation Language Modeling:

XLNet menggunakan metode yang disebut permutation language modeling untuk mengantisipasi token dalam urutan yang berbeda. Ini memungkinkan model untuk menangkap hubungan dua arah tanpa harus mengandalkan pendekatan tradisional yang hanya mempertimbangkan konteks dari kedua arah seperti BERT. Fungsi probabilitas token ke- t dalam urutan yang dipermutasikan z adalah: $P(X_t | X_z < t; \theta)$ (3) di mana z adalah permutasi dari urutan asli dan θ adalah parameter model.

2. Segment-Level Recurrence:

XLNet memperkenalkan segment-level recurrence untuk mempertahankan konteks dari segmen teks sebelumnya. Representasi hidden state dari token i di segmen j adalah: $h_j^i = \text{Transformer}(h_{j-1}^i, \text{hidden states from segment } j-1)$ (4)

3. Relative Positional Encodings:

XLNet menggunakan posisi relatif untuk encoding, yang memungkinkan model menangkap informasi urutan lebih fleksibel.

$$\text{Attention}(Q, K, V, R) = \text{softmax}\left(\frac{Q(K+R)^T}{\sqrt{d_k}}\right)V \quad (5)$$

di mana R adalah embedding posisi relatif

3.2. Evaluasi BERT dan XLNET

Untuk mengevaluasi klasifikasi *tweet bully* pada *twitter* dalam pengujian analisis perbandingan metode BERT dan XLNet pada penelitian ini menggunakan metrik akurasi, presisi dan recall. Metrik akurasi digunakan untuk total prediksi yang benar dari semua prediksi yang ada, sedangkan recall digunakan untuk proporsi *tweet bully* yang diidentifikasi, dan metrik presisi digunakan untuk mengukur proporsi *tweet bully* yang diprediksi dari total text yang dilaksifikasi terindikasi *tweet bully*.

$$\text{Accuracy} = \frac{TP + TN}{FP + TP + FN + TN} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

4. HASIL DAN PEMBAHASAN

Pada bagian ini membahas penerapan metode BERT dan XLNet dalam melakukan kalsifikasi *tweet*

bullying, dan *not bullying*. Adapun Pseudocode komparasi BERT dan XLNet sebagai berikut:

```

predictions_bert =
predictUsingBERTModel(texts);

predictions_XLNet =
predictUsingXLNETModel(texts);

% 6. Hitung confusion matrix dan evaluasi
performa untuk model BERT
C_bert = confusionmat(labels,
predictions_bert);
accuracy_bert = sum(diag(C_bert)) /
sum(C_bert(:));
precision_bert = C_bert(2,2) /
sum(C_bert(:,2));
recall_bert = C_bert(2,2) /
sum(C_bert(2,:));
F1_score_bert = 2 * (precision_bert *
recall_bert) / (precision_bert +
recall_bert);

% 7. Hitung confusion matrix dan evaluasi
performa untuk model XLNET
C_XLNet = confusionmat(labels,
predictions_XLNet);
accuracy_XLNet = sum(diag(C_XLNet)) /
sum(C_XLNet(:));
precision_XLNet = C_XLNet(2,2) /
sum(C_XLNet(:,2));
recall_XLNet = C_XLNet(2,2) /
sum(C_XLNet(2,:));
F1_score_XLNet = 2 * (precision_XLNet *
recall_XLNet) / (precision_XLNet +
recall_XLNet);

% 8. Tampilkan hasil
fprintf('Results for BERT Model:\n');
fprintf('Confusion Matrix:\n');
disp(C_bert);
fprintf('Accuracy: %.2f\n', accuracy_bert);
fprintf('Precision: %.2f\n',
precision_bert);
fprintf('Recall: %.2f\n', recall_bert);
fprintf('F1 Score: %.2f\n\n',
F1_score_bert);

fprintf('Results for XLNET Model:\n');
fprintf('Confusion Matrix:\n');
disp(C_XLNet);
fprintf('Accuracy: %.2f\n', accuracy_XLNet);
fprintf('Precision: %.2f\n',
precision_XLNet);
fprintf('Recall: %.2f\n', recall_XLNet);
fprintf('F1 Score: %.2f\n', F1_score_XLNet);

```

3.3. Dataset Tweet Bully

Untuk melakukan klasifikasi *tweet bully* menggunakan model BERT dan XLNET dalam melakukan kalsifikasi *tweet bullying*, dan *not bullying*, dibutuhkan dataset untuk dijadikan data latih dan data uji, dengan pelabelan untuk jenis *tweet*, yaitu *not bullying* untuk *tweet* normal, dan *bullying* untuk konten yang mengandung *cyberbullying*. Untuk memperoleh dataset *tweet cyberbullying* tersebut dapat didownload di www.kaggle.com, dengan kata kunci *detecting bullying tweet*, dan untuk penelitian ini sumber dataset diperoleh dari <https://www.kaggle.com/code/ludovicocuoghi/detecting-bullying-tweets-pytorch-lstm-bert/input> dengan

jumlah data sebanyak 47.692 *tweet* dalam format csv. Adapun dari jumlah *tweet* tersebut berisi konten *not bullying* sebanyak 7.945 dan konten *bullying* sebanyak 39.747

3.2. Pengujian Model BERT dan XLNet

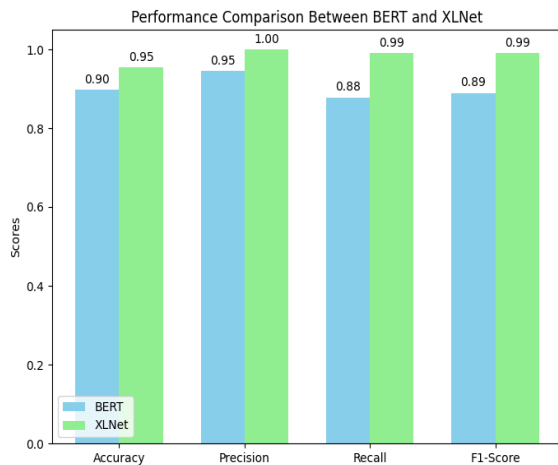
Hasil pengujian dengan menerapkan metode *BERT* dapat dilihat pada Tabel 1. Hasil pengujian tabel metrik menggunakan *BERT* dan *XLNET*

Table 1. Metrik Pengujian *BERT* dan *XLNET*

Models	Accuracy	Precision	Recall	F1-Score
<i>BERT</i>	0.8968	0.946	0.8781	0.8894
<i>XLNet</i>	0.9543	1.00	0.99	0.99

Dari hasil pengujian tersebut diperoleh tingkat akurasi model *XLNet* lebih baik **95%** dibandingkan model *BERT* dengan tingkat akurasi **89%**

Secara keseluruhan dari hasil pengujian metrik dapat dianalisis bahwa model *XLNet* jelas lebih unggul dibandingkan *BERT* dalam klasifikasi *tweet bullying* pada dataset penelitian saat ini. *XLNet* menunjukkan performa yang lebih baik pada semua metrik evaluasi (*accuracy, precision, recall, dan F1-Score*), dengan hasil yang hampir sempurna, sedangkan analisis yang dilakukan dari kapabilitas model *XLNet* memiliki keunggulan dalam menangkap informasi kontekstual secara lebih menyeluruh karena sifatnya yang *autoregressive*, dibandingkan model *BERT* yang lebih bergantung pada pemahaman kontekstual *bidirectional* yang menyebabkan keterbatasan dalam menangani jenis data yang panjang atau beragam seperti *tweet*.



Gambar 2. Grafik Pengujian Hasil Metrik menggunakan *BERT* dan *XLNET*

Jumlah *tweet bullying* yang jauh lebih besar (39.747) dibandingkan dengan *tweet not bullying* (7.945) juga mempengaruhi hasil, yaitu dataset tidak seimbang, sehingga kemampuan model dalam mengatasi ketidakseimbangan kelas pada model *XLNet* lebih baik, dibandingkan model *BERT*, dan hasil ini terlihat dari nilai *recall* yang lebih rendah.

Hasil perbandingan tersebut dapat dilihat dari grafik hasil pengujian metrik pada Gambar 2. Grafik pengujian hasil metrik menggunakan *BERT* dan *XLNET*

Hasil dari gambar grafik pengujian pada metrik (*accuracy, precision, recall, dan F1-Score*) terlihat jelas, bahwa model *XINET* lebih unggul dibandingkan model *BERT* dalam klasifikasi *tweet bully* pada *twitter*.

5. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa analisis perbandingan model *BERT* dan *XLNet* untuk klasifikasi *tweet bullying* pada *twitter* memerlukan proses pelatihan pada dataset yang telah mendapatkan pelabelan untuk konten yang terindikasi *bullying*, diberi label = 1, sedangkan *not bullying* diberi label = 0. Adapun dataset pelatihan dan pengujian untuk mengukur kinerja kedua model. Evaluasi kinerja model dilakukan dengan menggunakan berbagai metrik, seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model *XLNet* mampu memberikan klasifikasi yang lebih baik dibandingkan model *BERT* dalam membedakan konten *bully* dan *not bully* dengan tingkat akurasi sebesar 95%. Dengan nilai presisi = 100%, *recall* = 0,87%, dan *F1-score* = 0,88%.

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. doi: 10.1109/ICCWAMTIP51612.2020.9317379.
- ANGGRAINGSIH, R., HASSAN, G. M. AND DATTA, A. 2023. CE-BERT: Concise and Efficient BERT-Based Model for Detecting Rumors on Twitter. *IEEE Access*, 11(July), pp. 80207–80217. doi: 10.1109/ACCESS.2023.3299858.
- ARABADZHIEVA-KALCHEVA, N. AND KOVACHEV, I. 2021. Comparison of BERT and XLNet accuracy with classical methods and algorithms in text classification. *Proceedings of the International Conference on Biomedical Innovations and Applications, BIA 2021*. IEEE, 1(8), pp. 74–76. doi: 10.1109/BIA52594.2022.9831281.
- ASGARI-CHENAGHLU, M. et al. 2021. Topic Detection and Tracking Techniques on Twitter: A Systematic Review’, *Complexity*, 2021. doi: 10.1155/2021/8833084.

- DHIVYAA, C. R. *et al.* 2023. XLNet Transfer Learning Model for Sentimental Analysis. *International Conference on Sustainable Computing and Smart Systems, ICSCSS 2023 - Proceedings*, (Icscss), pp. 76–84. doi: 10.1109/ICSCSS57650.2023.10169445.
- AL FARISI, F. A., PERDANA, R. S. AND ADIKARA, P. P. 2023. Klasifikasi Intensi dengan Metode Ling Short-Term Memory pada Chatbot Bahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(7), pp. 1511–1518. doi: 10.25126/jtiik.1078000.
- FENG, C. *et al.* 2022. BERT-PPII: The Polyproline Type II Helix Structure Prediction Model Based on BERT and Multichannel CNN. *BioMed Research International*, 2022. doi: 10.1155/2022/9015123.
- GUPTA, K., JINAD, R. AND LIU, Q. 2023. Comparative Analysis of NLP Models for Detecting Depression on Twitter. *Proceedings - 2023 International Conference on Communications, Computing and Artificial Intelligence, CCCAI 2023*. IEEE, pp. 23–28. doi: 10.1109/CCCAI59026.2023.00013.
- HABBAT, N., ANOUN, H. AND HASSOUNI, L. 2023. Combination of GRU and CNN Deep Learning Models for Sentiment Analysis on French Customer Reviews Using XLNet Model. *IEEE Engineering Management Review*, 51(1), pp. 41–51. doi: 10.1109/EMR.2022.3208818.
- HU, Y. *et al.* 2022. Short-Text Classification Detector: A Bert-Based Mental Approach. *Computational Intelligence and Neuroscience*, 2022. doi: 10.1155/2022/8660828.
- LI, B., WANG, J. AND LIU, X. 2021. Parallel Cleaning Algorithm for Similar Duplicate Chinese Data Based on BERT. *Scientific Programming*, 2021(i). doi: 10.1155/2021/5916748.
- LI, H. *et al.* 2020. Comparing BERT and XLNet from the Perspective of Computational Characteristics. *2020 International Conference on Electronics, Information, and Communication, ICEIC 2020*. IEEE, pp. 1–4. doi: 10.1109/ICEIC49074.2020.9051081.
- NABILA, P. AND SETIAWAN, E. B. 2024. Adam and AdamW Optimization Algorithm Application on BERT Model for Hate Speech Detection on Twitter. *2024 International Conference on Data Science and Its Applications (ICoDSA)*. IEEE, pp. 346–351. doi: 10.1109/icodsa62899.2024.10651619.
- RICKO AND SASONGKO, P. S. 2021. Classification Bullying Tweet Using Convolutional Neural Network with Word2vec. *Proceedings - International Conference on Informatics and Computational Sciences*. IEEE, 2021-Novem, pp. 58–63. doi: 10.1109/ICICoS53627.2021.9651842.
- SALMA, T. D., SAPTAWATI, G. A. P. AND RUSMAWATI, Y. 2021. Text Classification Using XLNet with Infomap Automatic Labeling Process. *Proceedings - 2021 8th International Conference on Advanced Informatics: Concepts, Theory, and Application, ICAICTA 2021*. IEEE, pp. 1–6. doi: 10.1109/ICAICTA53211.2021.9640255.
- SANTHIYA, S. *et al.* 2024. A Comparative Exploration in Text Classification for Hate Speech and Offensive Language Detection Using BERT-Based and GloVe Embeddings. *2024 2nd International Conference on Disruptive Technologies, ICDT 2024*. IEEE, pp. 1506–1509. doi: 10.1109/ICDT61202.2024.10489019.
- SLAMET, C. *et al.* 2020. Deep learning approach for bullying classification on twitter social media with Indonesian language. *Proceedings - 2020 6th International Conference on Wireless and Telematics, ICWT 2020*. doi: 10.1109/ICWT50448.2020.9243653.
- YAN, M. *et al.* 2022. A Multimodal Retrieval and Ranking Method for Scientific Documents Based on HFS and XLNet. *Scientific Programming*, 2022. doi: 10.1155/2022/5373531.
- YAQIN, A. *et al.* 2023. Classification of Indonesian Tweet Bullying on Twitter Using K-Nearest Neighbor. *2023 International Conference on Informatics, Multimedia, Cyber and Information Systems, ICIMCIS 2023*. IEEE, pp. 330–334. doi: 10.1109/ICIMCIS60089.2023.10348992.