The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 4, August 2024
https://doi.org/10.33022/ijcs.v13i4.4295

Penerapan *Oversampling* Pada Klasifikasi Ujaran Kebencian Menggunakan *Bidirectional Encoder Representations from Transformers*

Risal Syahwaluddin[1], Debby Alita[2]
risalsyah13@gmail.com[1], debbyalita@teknokrat.ac.id[2]
[1,2] Universitas Teknokrat Indonesiaa

Informasi Artikel

Diterima : 11 Jul 2024
Direvisi : 8 Agu 2024
Disetujui : 22 Agu 2024

Kata Kunci

Klasifikasi, SMOTE, BERT, Ketidakseimbangan Kelas, Oversampling

Abstrak

Masalah ketidakseimbangan kelas adalah tantangan umum dalam pembangunan model klasifikasi, terutama dalam konteks ujaran kebencian. Penelitian ini mengevaluasi efektivitas teknik oversampling *SMOTE* dalam meningkatkan kinerja model klasifikasi ujaran kebencian yang menggunakan *BERT*. Dataset yang digunakan memiliki ketidakseimbangan kelas yang signifikan, dengan jumlah sampel terbanyak pada kelas *Hate*, diikuti oleh *Offensive*, dan *Neither*. Dua eksperimen dilakukan: satu tanpa menggunakan *SMOTE* dan satu lagi dengan penerapan SMOTE. Hasil menunjukkan bahwa penerapan *SMOTE* meningkatkan akurasi model keseluruhan dari 85% menjadi 88%. *Precision* untuk kelas minoritas *Offensive* meningkat dari 0.33 menjadi 0.45, meskipun *recall* menurun dari 0.45 menjadi 0.28. Pada kelas *Neither*, F1-*score* mengalami peningkatan, menunjukkan perbaikan dalam keseimbangan antara *precision* dan *recall*. Kinerja pada kelas mayoritas *Hate* tetap stabil, menunjukkan bahwa *SMOTE* tidak mengganggu kinerja model pada kelas yang sudah dominan. Secara keseluruhan, penerapan *SMOTE* memberikan manfaat yang signifikan dalam menangani ketidakseimbangan kelas, terutama dalam meningkatkan precision untuk kelas minoritas, sehingga menghasilkan model klasifikasi yang lebih akurat.

Keywords

Classification, SMOTE, BERT, Class Imbalance, Oversampling

Abstract

The problem of class imbalance is a common challenge in classification model building, especially in the context of hate speech. This study evaluates the effectiveness of the SMOTE oversampling technique in improving the performance of hate speech classification models using BERT. The dataset used has significant class imbalance, with the largest number of samples in the Hate class, followed by Offensive, and Neither. Two experiments were conducted: one without using SMOTE and one with SMOTE applied. Results showed that the application of SMOTE improved the overall model accuracy from 85% to 88%. Precision for the Offensive minority class increased from 0.33 to 0.45, although recall decreased from 0.45 to 0.28. In the Neither class, the F1-score increased, indicating an improvement in the balance between precision and recall. Performance on the majority Hate class remained stable, indicating that SMOTE did not interfere with the model's performance on the already dominant class. Overall, the application of SMOTE provides significant benefits in handling class imbalance, especially in improving precision for minority classes, resulting in a more accurate classification model.

A. Pendahuluan

Kemunculan media sosial telah mengubah lanskap komunikasi manusia secara mendalam. Fenomena ini menciptakan sebuah platform yang memfasilitasi individu untuk berinteraksi, berkolaborasi, dan mengungkapkan opini mereka secara luas dan seketika (Suparno et al., 2023). Namun, di tengah dinamika tersebut, muncul tantangan yang cukup besar dalam bentuk ujaran kebencian yang menyebar di semua platform tersebut. Ujaran kebencian, yang terkadang dimanifestasikan dalam bentuk diskriminasi, komentar yang menghina, atau bahkan ancaman terhadap kelompok tertentu, telah menjadi masalah yang signifikan dalam masyarakat digital (Mondal et al., 2017). Dampaknya bisa sangat merugikan, memperparah perpecahan antarkelompok, menyebarkan ketakutan, dan menghambat dialog yang produktif. Oleh karena itu, mendeteksi dan memitigasi ujaran kebencian di platform media sosial menjadi sangat penting sebagai langkah untuk mempromosikan lingkungan *online* yang aman dan inklusif.

Di tengah upaya untuk mengatasi tantangan-tantangan ini, teknologi pemrosesan bahasa alami (*natural language processing*/NLP) telah menjadi alat yang sangat penting. Metode seperti klasifikasi teks telah banyak digunakan untuk mengidentifikasi konten berbahaya seperti ujaran kebencian (Themeli et al., 2021). Namun, salah satu tantangan yang dihadapi dalam pemrosesan ujaran kebencian adalah ketidakseimbangan kelas, di mana jumlah contoh positif (ujaran kebencian) secara signifikan lebih kecil daripada jumlah contoh negatif (konten netral atau positif) (Mifsud et al., 2023).

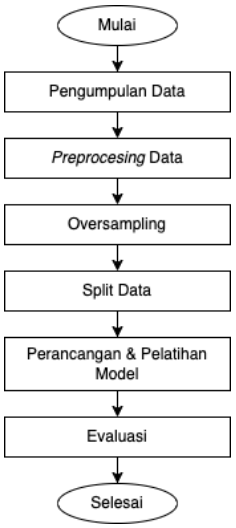
Ketidakseimbangan kelas ini menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga performa klasifikasi pada kelas minoritas (ujaran kebencian) menjadi kurang optimal. Untuk menangani masalah ketidakseimbangan kelas ini, pendekatan oversampling menjadi fokus utama dalam penelitian. Oversampling adalah teknik yang digunakan untuk mengatasi ketidakseimbangan kelas dengan cara meningkatkan jumlah sampel dari kelas minoritas (Zhu et al., 2023). Salah satu metode oversampling yang sering digunakan dalam klasifikasi teks adalah *Synthetic Minority Over-sampling Technique* atau SMOTE. SMOTE menghasilkan sampel sintetis baru dari kelas minoritas dengan menginterpolasi antara sampel minoritas yang ada. Teknik ini lebih canggih dibandingkan random oversampling karena membantu mengurangi risiko overfitting dengan memperkenalkan variasi dalam data sintetis yang dihasilkan (Nguyen et al., 2023).

Untuk menangani masalah ketidakseimbangan kelas ini, pendekatan *oversampling* menjadi fokus utama dalam penelitian. Dalam konteks klasifikasi teks, *oversampling* melibatkan peningkatan jumlah sampel ujaran kebencian dalam dataset pelatihan untuk meningkatkan pemahaman model terhadap pola-pola yang terdapat dalam konten tersebut (Wei et al., 2022). Dalam penelitian ini, penggunaan *oversampling* dalam klasifikasi ujaran kebencian menggunakan teknik representasi bahasa terbaru yang dikenal sebagai *Bidirectional Encoder Representations from Transformers* (BERT). BERT, sebuah model bahasa yang dilatih secara ekstensif pada beberapa tugas pemrosesan bahasa alami, telah terbukti efektif dalam memahami konteks tekstual yang kompleks (Ali et al., 2023). Kami akan memanfaatkan kemampuan BERT dalam menangkap informasi kontekstual dari teks untuk meningkatkan kinerja klasifikasi ujaran kebencian. Maka dari itu, peneliti

ini bertujuan untuk meningkatkan kinerja klasifikasi ujaran kebencian dengan mengintegrasikan *oversampling* dengan model BERT.

B. Metode Penelitian

Penelitian ini dilakukan dengan mengikuti kerangka kerja untuk memastikan bahwa penelitian ini terstruktur dan terarah, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Kerangka Kerja

Pengumpulan Data

Tahap pertama dari penelitian ini melibatkan pengumpulan dataset yang sesuai untuk klasifikasi ujaran kebencian. Dataset yang digunakan adalah "*Hate Speech* and *Offensive Language Dataset*" yang tersedia di Kaggle dan dikembangkan oleh Andrii Samoshyn. Deskripsi dataset menunjukkan bahwa teks dalam dataset ini diklasifikasikan sebagai ujaran kebencian (hate-speech), bahasa yang menyinggung (*offensive language*), atau bukan keduanya yang terdiri dari 15056 data. Penting untuk dicatat bahwa dataset ini mengandung teks yang dapat dianggap rasialis, seksis, homofobik, atau secara umum mengandung konten yang menyinggung. Contoh dari dataset yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Sampel Dataset Ujaran Kebencian

No	*Class*	*Tweet*
1	Hate	!!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. &amp
2	Offensive	!!!!! RT @mleew17: boy dats cold...tyga dwn bad for cuffin dat hoe in the 1st place!!
3	Offensive	!!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby4life: You ever fuck a bitch and she start to cry? You be confused as shit
4	Offensive	!!!!!!!!! RT @C_G_Anderson: @viva_based she look like a tranny
5	Offensive	!!!!!!!!!!!! RT @ShenikaRoberts: The shit you hear about me might be true or it might be faker than the bitch who told it to ya &#57361
…	…	…
15056	Neither	you're such a retard i hope you get type 2 diabetes and die from a sugar rush you fucking faggot @Dare_ILK

Preprocessing Data

Data yang terkumpul kemudian akan menjalani *preprocessing* untuk membersihkan, menormalkan, dan memberi token pada teks. Langkah pertama dalam fungsi ini adalah menghapus angka dan karakter khusus menggunakan ekspresi reguler, sehingga hanya huruf dan spasi yang tersisa dalam teks (Arief et al., 2022). Setelah itu, langkah kedua membantu menghilangkan spasi tambahan yang mungkin muncul setelah proses penghapusan karakter khusus sebelumnya (Adityani et al., 2022). Kemudian, langkah ketiga menghilangkan URL yang ada dalam teks, yang dapat mengganggu analisis atau pemodelan teks (Wolf et al., 2022). Langkah selanjutnya adalah normalisasi huruf, di mana semua huruf diubah menjadi huruf kecil untuk menghindari perbedaan antara huruf besar dan kecil saat pemrosesan (Zhou et al., 2021). Normalisasi kontraksi adalah langkah berikutnya, di mana kontraksi seperti '*didn't*' dikembalikan ke bentuk lengkapnya seperti *'did not*' (Bainbridge & Baker, 2020). Selanjutnya, teks dijalani proses lemmatisasi, yang mengubah kata-kata ke dalam bentuk dasarnya untuk mempermudah analisis dan pemodelan lebih lanjut (Shaukat et al., 2023). Langkah terakhir adalah penanganan *typo* dengan menghapus kata 'rt', yang mungkin merupakan langkah khusus tergantung pada sifat dataset (Shah et al., 2021). Hasil dari *preprocessing* dapat dilihat pada Tabel 2.

Tabel 2. Sampel Hasil *Pre-proccessing* Data

No	Teks Asli	Hasil	*Pre-proccessing*
1	!!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. &amp	As a woman you shouldnt complain about cleaning up your house	• Hapus pola seperti '!!!!! RT @username:' • Hapus tanda seru berlebihan di awal • Hapus '&amp'
2	RT @mleew17: boy dats cold...tyga dwn bad for cuffin dat hoe in the 1st place!!	boy dats coldtyga dwn bad for cuffin dat hoe in the 1st place	• Hapus pola seperti '!!!!! RT @username:' • Hapus tanda seru berlebihan di awal
3	!!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby4life: You ever fuck a bitch and she start to cry? You be confused as shit	Dawg You ever fuck a bitch and she start to cry You be confused as shit	• Hapus pola seperti '!!!!! RT @username:' • Hapus pola 'RT @username:' • Hapus tanda seru berlebihan di awal
4	!!!!!!!!! RT @C_G_Anderson: @viva_based she look like a tranny	she look like a tranny	• Hapus pola seperti '!!!!! RT @username:' • Hapus pola '@username' • Hapus tanda seru berlebihan di awal
5	!!!!!!!!!!!! RT @ShenikaRoberts: The shit you hear about me might be true or it might be faker than the bitch who told it to ya &#57361	The shit you hear about me might be true or it might be faker than the bitch who told it to ya	• Hapus pola seperti '!!!!! RT @username:' • Hapus '&amp' • Hapus tanda seru berlebihan di awal
6	you're such a retard i hope you get type 2 diabetes and die from a sugar rush you fucking faggot @Dare_ILK	youre such a retard i hope you get type diabetes and die from a sugar rush you fucking faggot	• Hapus pola '@username'

Oversampling

Setelah *preprocessing*, dataset akan dianalisis untuk menentukan ketidakseimbangan kelas antara ujaran kebencian dan konten netral/positif. Teknik *oversampling* akan diimplementasikan untuk meningkatkan jumlah sampel ujaran kebencian, sehingga dapat menyeimbangkan dataset. Metode *oversampling* yang digunakan SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menangani ketidakseimbangan kelas. SMOTE bekerja dengan cara membuat sampel sintetis baru di antara titik-titik minoritas yang ada dalam ruang fitur (Wang & Liu, 2023). Langkah pertama dalam proses *oversampling* menggunakan SMOTE adalah menentukan jumlah tetangga terdekat yang akan digunakan untuk membuat sampel sintetis baru. Setelah itu, SMOTE memilih secara acak titik minoritas dari dataset yang ada dan menentukan tetangga terdekat untuk titik tersebut. Selanjutnya, SMOTE membuat sampel sintetis baru di antara titik tersebut dengan menggabungkan atribut dari titik minoritas yang dipilih dengan atribut dari beberapa tetangga terdekat (Vitianingsih et al., 2022).

Pembagian Dataset

Dataset yang diproses akan dibagi menjadi subset pelatihan, validasi, dan pengujian. Biasanya, proporsi dataset yang umum digunakan adalah 80% untuk pelatihan dan 20% untuk pengujian. Pembagian ini akan memastikan bahwa model dapat dilatih secara efektif, dievaluasi, dan diuji secara *independent* (Hou et al., 2023).

Perancangan dan Pelatihan Model

Model BERT akan digunakan sebagai dasar untuk klasifikasi ujaran kebencian. BERT akan diinisialisasi menggunakan bobot yang telah dilatih sebelumnya pada tugas pemrosesan bahasa alami berskala besar. Selanjutnya, lapisan klasifikasi akan ditambahkan di atas model BERT untuk menghasilkan prediksi kelas. Model BERT yang diperluas akan dilatih menggunakan subset dari kumpulan data pelatihan (Su et al., 2020). Proses pelatihan akan melibatkan penyesuaian bobot model secara berulang-ulang selama beberapa *epoch*, menggunakan fungsi kerugian yang sesuai seperti kerugian *cross-entropy.* Konfigurasi BERT dapat dilihat pada Tabel 3.

Tabel 3. Konfigurasi Bert

Komponen	Nilai
Arsitektur BERT	bert-base-uncased
Jumlah Kelas Output	3
Optimizer	Adam (learning_rate=2e-5)
Fungsi Loss	Categorical Crossentropy
Fungsi Aktivasi	Softmax
Metrik	Accuracy
Patience	5
Restore Best Weights	True
Batch Size	32
Epochs	100

Berdasarkan Tabel 3, model BERT yang digunakan dalam penelitian ini dikonfigurasi dengan beberapa parameter penting untuk klasifikasi ujaran kebencian. Arsitektur dasar yang dipilih adalah bert-base-uncased, sebuah varian BERT yang terdiri dari 12 lapisan, 768 unit tersembunyi, dan 12 attention head (Haque et al., 2022). Model ini bekerja pada teks yang dikonversi menjadi huruf kecil, yang berarti teks input akan diubah menjadi huruf kecil sebelum diproses. Model ini dirancang untuk mengklasifikasikan teks ke dalam tiga kategori: Hate, Offensive, dan Neither. Oleh karena itu, lapisan output memiliki tiga neuron untuk mengakomodasi tiga kelas tersebut. Optimizer yang digunakan adalah Adam (Adaptive Moment Estimation) dengan laju pembelajaran sebesar 2e-5, yang membantu dalam menyesuaikan bobot model secara efektif selama pelatihan (Reyad et al., 2023). Fungsi kerugian yang dipilih adalah Categorical Crossentropy, yang sesuai untuk tugas klasifikasi dengan lebih dari dua kelas, dan fungsi aktivasi yang digunakan di lapisan output adalah Softmax, yang mengkonversi skor yang dihasilkan oleh model menjadi probabilitas (Garcin et al., 2023).

Evaluasi

Setelah pelatihan selesai, model akan dievaluasi untuk mengukur kinerjanya. Evaluasi ini akan mencakup metrik seperti akurasi, presisi, *recall*, dan F1-*score* untuk menilai kemampuan model dalam mendeteksi ujaran kebencian (Sarwar & Murdock, 2022).

C. Hasil dan Pembahasan

Penerapan Oversampling Menggunakan SMOTE

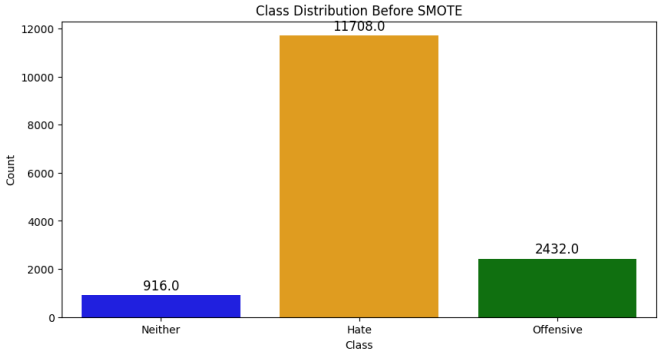
Pada tahap penerapan *oversampling* menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), langkah-langkah yang terdiri dari pemisahan teks dan label kelas dari dataset, tokenisasi teks menggunakan *tokenizer* BERT, aplikasi SMOTE untuk meresampling dataset, tokenisasi ulang pada dataset yang sudah *diresampling*, dan persiapan data untuk pelatihan dan pengujian model. *Tokenizer* BERT digunakan untuk melakukan tokenisasi terhadap teks ujaran. Tokenisasi ini memungkinkan kita untuk mewakili teks sebagai urutan token atau subkata-kata yang sesuai dengan kosakata yang digunakan oleh model BERT. Hasil dari tokenisasi dapat diliat pada Tabel 4.

Tabel 4. Sampel Hasil Tokenisasi

Teks	Hasil Tokenisasi
as a woman you shouldnt complain about cleaning up your house	[101 1037 1037 2450 2017 5807 2102 17612 2055 9344 2039 2115 2160 102]
boy dat coldtyga dwn bad for cuffin dat hoe in the st place	[101 2879 23755 3147 3723 3654 1040 7962 2919 2005 26450 2378 23755 7570 2063 1999 1996 2358 2173 102]
dawg you ever fuck a bitch and she start to cry you be confused a shit	[101 4830 27767 2017 2412 6616 1037 7743 1998 2016 2707 2000 5390 2017 2022 5457 1037 4485 102]

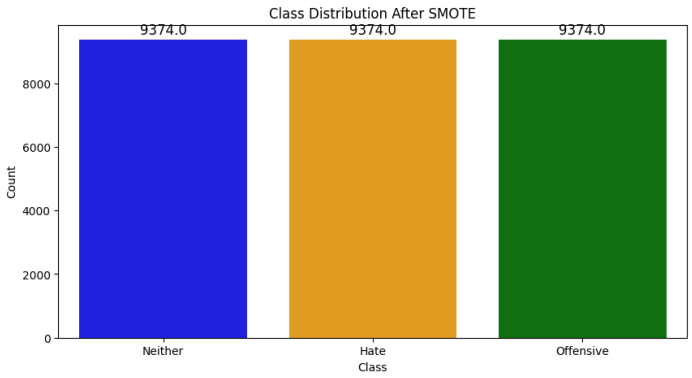
Proses selanjutnya melibatkan aplikasi SMOTE untuk meresampling dataset, di mana sampel sintetis baru dibuat dari kelas minoritas dengan cara menggabungkan

atribut dari sampel minoritas yang sudah ada dalam ruang fitur yang ada. Sebelum menerapkan SMOTE, distribusi jumlah data per kelas dalam dataset klasifikasi ujaran kebencian tidak seimbang. Kelas *Hate* memiliki jumlah sampel terbanyak, yaitu sebanyak 11.708, diikuti oleh kelas *Offensive* dengan 2.432 sampel, dan kelas *Neither* dengan hanya 916 sampel. Seperti pada Gambar 2.



Gambar 2. Distribusi Kelas Sebelum SMOTE

Proses SMOTE dimulai dengan mengidentifikasi instansi kelas minoritas. Kemudian, untuk setiap instansi kelas minoritas, SMOTE menemukan k tetangga terdekat (umumnya k=5). Setelah itu, sampel sintetis dibuat dengan menginterpolasi antara instansi minoritas dan salah satu tetangga terdekatnya. Interpolasi dilakukan dengan memilih titik acak sepanjang segmen garis yang menghubungkan instansi minoritas dengan tetangganya. Hasilnya adalah instansi baru yang berada di antara dua titik tersebut. Sebelum menerapkan SMOTE, dataset memiliki 916 sampel untuk kelas *'Neither'*, 11708 sampel untuk kelas *'Hate'*, dan 2432 sampel untuk kelas *'Offensive'*. Setelah menerapkan SMOTE, jumlah sampel di setiap kelas menjadi seimbang, yaitu masing-masing 9374 sampel untuk kelas 'Neither', 'Hate', dan 'Offensive'. Ini berarti SMOTE telah berhasil menambahkan sampel sintetis untuk kelas 'Neither' dan 'Offensive' sehingga jumlah sampel di setiap kelas menjadi sama seperti yang dilihat pada Gambar 3.



Gambar 3. Distribusi Kelas Sesudah SMOTE

Hasil Pelatihan Model

Dua eksperimen dilakukan untuk membandingkan kinerja model klasifikasi ujaran kebencian, satu tanpa menggunakan *oversampling* dan satu lagi dengan

menggunakan teknik SMOTE untuk *oversampling*. Berikut adalah hasil dari kedua eksperimen yang bisa dilihat pada Tabel 5 dan Tabel 6.

Tabel 5. Hasil Pelatihan Tanpa Penerapan SMOTE

Class	*Precision*	*Recall*	*F1-Score*
Offensive	0.33	0.45	0.38
Hate	0.94	0.89	0.91
Neither	0.77	0.86	0.81
Accuracy			0.85
Macro Avg	0.68	0.73	0.70
Weighted Avg	0.87	0.85	0.86

Tabel 6. Hasil Pelatihan Dengan Penerapan SMOTE

Class	*Precision*	*Recall*	*F1-Score*
Offensive	0.45	0.28	0.35
Hate	0.92	0.94	0.93
Neither	0.80	0.83	0.82
Accuracy			0.88
Macro Avg	0.72	0.68	0.70
Weighted Avg	0.87	0.88	0.87

Berdasarkan Tabel 5 dapat dilihat bahwa model secara konsisten memperbaiki performanya pada data pelatihan, dengan penurunan yang stabil dalam tingkat *loss* dan peningkatan yang sejalan dalam akurasi. Namun, pada tahap-tahap akhir pelatihan, terjadi perbedaan yang signifikan antara kinerja pada data pelatihan dan data validasi, menunjukkan kemungkinan terjadinya *overfitting*. Dalam eksperimen ini, *early stopping* diimplementasikan dengan patiensi sebesar 5. Pada *epoch* 7, terjadi peningkatan *loss* pada data validasi yang menandakan adanya *overfitting*. Early *stopping* kemudian menghentikan pelatihan pada epoch 8, sehingga model yang memiliki akurasi validasi tertinggi adalah pada *epoch* 3 dengan akurasi *sebesar* 89.11%.

Berdasarkan Tabel 5, penerapan *oversampling* menggunakan metode SMOTE dalam pelatihan model menunjukkan perubahan yang signifikan dalam kinerja. Dapat diamati bahwa selama proses pelatihan, terjadi penurunan yang signifikan dalam nilai *loss* dan peningkatan yang konsisten dalam akurasi pada kedua data pelatihan dan validasi. Hal ini menunjukkan bahwa model dapat menggeneralisasi pola-pola yang ditemui dalam data dengan lebih baik setelah penerapan *oversampling*. Salah satu indikator keberhasilan model adalah akurasi pada data validasi. Terlihat bahwa akurasi pada data validasi mencapai puncaknya pada *epoch* ke-8 dengan nilai 96,31%. Setelah itu, akurasi sedikit menurun pada *epoch* berikutnya, tetapi tetap berada pada tingkat yang tinggi. Hal ini menunjukkan bahwa model mampu menghasilkan prediksi yang konsisten pada data validasi.

D. Simpulan

Penerapan teknik oversampling SMOTE pada model klasifikasi ujaran kebencian memberikan hasil yang signifikan dalam memperbaiki kinerja model. Hasil pelatihan menunjukkan bahwa SMOTE meningkatkan akurasi keseluruhan model dari 85% menjadi 88%, menunjukkan bahwa teknik ini membantu model dalam mengklasifikasikan data secara lebih efektif. Pada kelas minoritas *Offensive*,

meskipun precision meningkat dari 0.33 menjadi 0.45, recall mengalami penurunan dari 0.45 menjadi 0.28, yang berarti model lebih baik dalam mengurangi false positives tetapi masih kesulitan mengidentifikasi semua instance dari kelas ini. Namun, F1-score pada kelas *Neither* meningkat, menunjukkan perbaikan dalam keseimbangan antara precision dan recall untuk kelas tersebut. Kinerja pada kelas mayoritas *Hate* tetap stabil dengan penerapan SMOTE, dengan sedikit perubahan dalam metrik *precision*, *recall*, dan F1-*score*, menunjukkan bahwa teknik ini tidak mengganggu kinerja model pada kelas mayoritas yang sudah baik. Rata-rata makro precision juga meningkat dari 0.68 menjadi 0.72, yang berarti ada perbaikan dalam kemampuan model untuk mengklasifikasikan setiap kelas secara lebih seimbang, meskipun rata-rata makro recall sedikit menurun.

E. Referensi

[1] Adityani, S. P., Richasdy, D., & Astuti, W. (2022). Sentiment and Discussion Topic Analysis on Social Media Group using Support Vector Machine. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *6*(3), 1378. https://doi.org/10.30865/mib.v6i3.4233

[2] Ali, A., Azman Mohd Noah, S., Kebangsaan Malaysia, U., & Bangi Lailatul Qadri Zakaria, U. (2023). *A BERT-Based model: Improving Crime News Documents Classiication through Adopting Pre-trained Language Models*. https://doi.org/10.21203/rs.3.rs-2582775/v1

[3] Arief, R., Mutiara, A. B., Kusuma, T. M., & Hustinawaty. (2022). Automated hierarchical classification of scanned documents using convolutional neural network and regular expression. *International Journal of Electrical and Computer Engineering*, *12*(1), 1018–1029. https://doi.org/10.11591/ijece.v12i1.pp1018-1029

[4] Bainbridge, W. A., & Baker, C. I. (2020). Boundaries Extend and Contract in Scene Memory Depending on Image Properties. *Current Biology*, *30*(3), 537-543.e3. https://doi.org/10.1016/j.cub.2019.12.004

[5] Garcin, C., Servajean, M., Joly, A., & Salmon, J. (2023). *A two-head loss function for deep Average-K classification*. http://arxiv.org/abs/2303.18118

[6] Haque, S., Eberhart, Z., Bansal, A., & McMillan, C. (2022). Semantic Similarity Metrics for Evaluating Source Code Summarization. *IEEE International Conference on Program Comprehension*, *2022-March*, 36–47. https://doi.org/10.1145/nnnnnnn.nnnnnnn

[7] Hou, R., Lo, J. Y., Marks, J. R., Hwang, S., & Grimm, L. J. (2023). *Classification performance bias between training and test sets in a limited mammography dataset*. https://doi.org/10.1101/2023.02.15.23285985

[8] Mifsud, R., Deka, L., & Lahiri, I. (2023). An Optimised BERT Pretraining Approach for Identification of Targeted Offensive Language: Data Imbalance and Potential Solutions. *2023 4th International Conference on Computing and Communication Systems, I3CS 2023*. https://doi.org/10.1109/I3CS58314.2023.10127515

[9] Mondal, M., Silva, L. A., & Benevenuto, F. (2017). A measurement study of hate speech in social media. *HT 2017 - Proceedings of the 28th ACM*

Conference on Hypertext and Social Media, 85–94. https://doi.org/10.1145/3078714.3078723

[10] Nguyen, T., Mengersen, K., Sous, D., & Liquet, B. (2023). SMOTE-CD: SMOTE for compositional data. *PLoS ONE*, *18*(6 June). https://doi.org/10.1371/journal.pone.0287705

[11] Reyad, M., Sarhan, A. M., & Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. *Neural Computing and Applications*, *35*(23), 17095–17112. https://doi.org/10.1007/s00521-023-08568-z

[12] Sarwar, M., & Murdock, V. (2022). *Unsupervised Domain Adaptation for Hate Speech Detection Using a Data Augmentation Approach*. https://www.adl.org

[13] Shah, R., Ali, F. M., Finlay, A. Y., & Salek, M. S. (2021). Correction to: Family reported outcomes, an unmet need in the management of a patient's disease: appraisal of the literature (Health and Quality of Life Outcomes, (2021), 19, 1, (194), 10.1186/s12955-021-01819-4). In *Health and Quality of Life Outcomes* (Vol. 19, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12955-021-01839-0

[14] Shaukat, S., Asad, M., & Akram, A. (2023). Developing an Urdu Lemmatizer Using a Dictionary-Based Lookup Approach. *Applied Sciences (Switzerland)*, *13*(8). https://doi.org/10.3390/app13085103

[15] Su, Y., Xiang, H., Xie, H., Yu, Y., Dong, S., Yang, Z., & Zhao, N. (2020). Application of BERT to Enable Gene Classification Based on Clinical Evidence. *BioMed Research International*, *2020*. https://doi.org/10.1155/2020/5491963

[16] Suparno, B. A., Muktiyo, W., & Susilastuti DN, (Raden Roro) (Dwi). (2023). *Media komunikasi : representasi budaya dan kekuasaan*.

[17] Themeli, C., Giannakopoulos, G., & Pittaras, N. (2021). *A study of text representations in Hate Speech Detection*. http://arxiv.org/abs/2102.04521

[18] Vitianingsih, A. V., Othman, Z., Baharin, S. S. K., Suraji, A., & Maukar, A. L. (2022). Application of the Synthetic Over-Sampling Method to Increase the Sensitivity of Algorithm Classification for Class Imbalance in Small Spatial Datasets. *International Journal of Intelligent Engineering and Systems*, *15*(5), 676–690. https://doi.org/10.22266/ijies2022.1031.58

[19] Wang, Z., & Liu, Q. (2023). Imbalanced Data Classification Method Based on LSSASMOTE. *IEEE Access*, *11*, 32252–32260. https://doi.org/10.1109/ACCESS.2023.3262460

[20] Wei, G., Mu, W., Song, Y., & Dou, J. (2022). An improved and random synthetic minority oversampling technique for imbalanced data. *Knowledge-Based Systems*, *248*. https://doi.org/10.1016/j.knosys.2022.108839

[21] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. M. (2023). *Transformers: State-of-the-Art Natural Language Processing*. https://github.com/huggingface/

[22] Zhou, L., Liu, S., Li, C., Sun, Y., Zhang, Y., Li, Y., Yuan, H., Sun, Y., Xu, F., & Li, Y. (2021). Natural Language Processing Algorithms for Normalizing Expressions of Synonymous Symptoms in Traditional Chinese Medicine. *Evidence-Based Complementary and Alternative Medicine*, *2021*. https://doi.org/10.1155/2021/6676607

[23] Zhu, T., Liu, X., & Zhu, E. (2023). Oversampling With Reliably Expanding Minority Class Regions for Imbalanced Data Learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(6), 6167–6181. https://doi.org/10.1109/TKDE.2022.3171706