

## Evaluasi Efektivitas SMOTE dan Random Under Sampling pada Klasifikasi Emosi Tweet

I Komang Dharmendra<sup>1,\*</sup>, I Made Agus Wirahadi Putra<sup>2</sup>, Yohanes Priyo Atmojo<sup>3</sup>

<sup>1</sup>Fakultas Informatika dan Komputer; Sistem Informasi; Institut Teknologi dan Bisnis STIKOM Bali; Denpasar; Indonesia; (0361) 244445; e-mail: dharmendra@stikom-bali.ac.id<sup>1</sup>,

<sup>2</sup>Fakultas Bisnis dan Vokasi, Manajemen Informatika, Institut Teknologi dan Bisnis STIKOM Bali, Denpasar, Indonesia; e-mail: wirahadi@stikom-bali.ac.id<sup>2</sup>,

<sup>3</sup>Fakultas Informatika dan Komputer; Teknologi Informasi; Institut Teknologi dan Bisnis STIKOM Bali; Denpasar; Indonesia; (0361) 244445; e-mail: yohanes@stikom-bali.ac.id<sup>3</sup>

\* Korespondensi: e-mail: dharmendra@stikom-bali.ac.id

Diterima: 8 Oktober 2024 ; Review: 19 November 2024; Disetujui: 15 Desember 2024

Cara sitasi: Dharmarendra IK, Putra IMAW, Atmojo YP. 2024. Evaluasi Efektivitas SMOTE dan Random Under Sampling pada Klasifikasi Emosi Tweet. Informatics for Educators and Professionals : Journal of Informatics. Vol.9 (2): 182 – 193.

**Abstrak:** Penelitian ini bertujuan untuk mengevaluasi efektivitas dua teknik model sampling, yaitu SMOTE (*Synthetic Minority Over-sampling Technique*) dan *Random Under Sampling* (RUS), dalam meningkatkan performa beberapa model klasifikasi dalam menangani ketidakseimbangan data pada klasifikasi emosi tweet berbahasa Indonesia. Penelitian ini menggunakan pendekatan eksperimen dengan menerapkan teknik oversampling dan undersampling pada dataset yang berisi tweet yang telah dianalisis emosinya. Subjek penelitian terdiri dari lima model klasifikasi, yaitu *Maximum Entropy*, SVM, *Random Forest*, *Neural Network*, dan Naive Bayes. Data diperoleh dari pengumpulan tweet yang diproses menggunakan teknik text mining dan dianalisis dengan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa SMOTE secara konsisten memberikan peningkatan yang lebih signifikan dibandingkan RUS, terutama pada model *Random Forest* dan *Neural Network*. *Maximum Entropy* dan SVM terbukti sebagai model dengan performa terbaik di kedua skenario, sementara Naive Bayes meskipun efisien dalam hal waktu, menunjukkan performa lebih rendah. Penelitian ini memberikan kontribusi bagi pengembangan teknik penanganan ketidakseimbangan data dalam klasifikasi emosi, khususnya dalam konteks bahasa Indonesia, serta memberikan rekomendasi untuk pengembangan model yang lebih akurat dan sensitif terhadap ketidakseimbangan kelas. Dampak penelitian ini diharapkan dapat meningkatkan kualitas analisis emosi pada media sosial dan menyediakan solusi praktis bagi organisasi yang ingin memahami emosi publik secara lebih responsif.

**Kata kunci:** Ketidakseimbangan Data, Klasifikasi Emosi, *Maximum Entropy* Random Under Sampling, SMOTE

**Abstract:** This study aims to evaluate the effectiveness of two sampling techniques, namely SMOTE (*Synthetic Minority Over-sampling Technique*) and *Random Under Sampling* (RUS), in improving the performance of several classification models for addressing class imbalance in emotion classification of Indonesian-language tweets. The study adopts an experimental approach by applying oversampling and undersampling techniques to a dataset containing tweets with annotated emotional categories. The research subjects include five classification models: *Maximum Entropy*, SVM, *Random Forest*, *Neural Network*, and Naive Bayes. Data were collected from tweets processed using text mining techniques and analyzed using evaluation metrics such as accuracy, precision, recall, and F1-score. The results show that SMOTE consistently provides more significant performance improvements than RUS, particularly for the *Random Forest* and

*Neural Network models. Maximum Entropy and SVM demonstrated the best performance across both scenarios, whereas Naive Bayes, while computationally efficient, exhibited lower performance metrics. This study contributes to advancing techniques for handling class imbalance in emotion classification, particularly in the context of the Indonesian language, and offers recommendations for developing more accurate and sensitive models. The findings are expected to enhance the quality of emotion analysis on social media and provide practical solutions for organizations seeking to understand public emotions more responsively.*

**Keywords:** SMOTE; Random Under Sampling; Maximum Entropy; Data Imbalance; Emotion Classification

## 1. Pendahuluan

Dalam era digital yang semakin berkembang, media sosial, khususnya Twitter, telah menjadi salah satu sumber utama komunikasi dan penyebaran informasi, di mana jutaan pengguna mengungkapkan opini, perasaan, dan tanggapan terhadap berbagai isu, produk, atau peristiwa setiap hari[1]. Namun, tantangan utama yang dihadapi di lapangan adalah bagaimana memahami dan mengelola luapan emosi dari data Twitter yang sangat besar, tidak terstruktur, dan heterogen[2]. Perusahaan, organisasi, maupun peneliti sering kesulitan dalam menganalisis emosi ini secara cepat dan akurat, sehingga respons terhadap perubahan sentimen publik kerap terlambat atau tidak tepat sasaran. Idealnya, analisis emosi dari data media sosial dapat dilakukan secara cepat, akurat, dan menyeluruh untuk membantu organisasi dalam mengidentifikasi sentimen masyarakat, memahami dinamika sosial, dan merancang strategi yang lebih responsif. Sayangnya, analisis ini sering terkendala oleh ketidakseimbangan data, di mana emosi tertentu seperti “senang” atau “marah” lebih dominan dibandingkan dengan emosi lain seperti “cemas” atau “sedih,” yang dapat mengurangi akurasi model klasifikasi emosi[3].

Sebagai solusi terhadap permasalahan ini, penelitian ini menawarkan pendekatan inovatif dalam klasifikasi emosi menggunakan *text mining* yang memanfaatkan algoritma pembelajaran mesin dengan teknik penanganan ketidakseimbangan data[4]. Fokus penelitian ini adalah pada analisis emosi dalam tweet berbahasa Indonesia, yang relatif masih jarang dieksplorasi secara mendalam dibandingkan dengan teks dalam bahasa Inggris. Penelitian sebelumnya telah menunjukkan bahwa *text mining* efektif dalam analisis emosi. Misalnya, penelitian yang menemukan bahwa analisis sentimen dapat membantu memahami pandangan masyarakat terhadap isu-isu terkini[5], [6]. Namun, kendala utama yang dihadapi adalah distribusi data yang tidak seimbang, yang membuat model bias terhadap kelas emosi tertentu. Selain itu, studi lain mengungkapkan bahwa klasifikasi emosi pada teks berbahasa Indonesia menghadapi tantangan tambahan berupa nuansa bahasa dan variasi gaya penulisan[7].

Pendekatan seperti *Synthetic Minority Over-sampling Technique (SMOTE)*[8] dan *data augmentation*[9] telah digunakan untuk mengatasi ketidakseimbangan data, tetapi efektivitasnya pada teks pendek seperti tweet masih menjadi tantangan tersendiri. Oleh karena itu, penelitian ini menawarkan kontribusi baru dengan mengevaluasi kombinasi teknik penanganan ketidakseimbangan data pada dataset berbahasa Indonesia, serta membandingkan kinerjanya menggunakan metrik evaluasi yang lebih sensitif terhadap kelas minoritas. Tujuan utama penelitian ini adalah mengembangkan model klasifikasi emosi yang lebih akurat dan sensitif terhadap ketidakseimbangan data pada tweet berbahasa Indonesia, mengevaluasi efektivitas berbagai teknik penanganan ketidakseimbangan data, termasuk oversampling, undersampling, dan algoritma berbobot, serta memberikan rekomendasi praktis untuk organisasi dalam memahami emosi publik melalui analisis *text mining* terhadap data media sosial. Dengan pendekatan ini, penelitian ini diharapkan memberikan kontribusi dalam meningkatkan kualitas analisis emosi di media sosial, khususnya untuk konteks Indonesia, serta menawarkan solusi yang relevan terhadap kendala yang ditemukan dalam penelitian sebelumnya.

## 2. Metode Penelitian

### 2.1 Twitter

Twitter adalah platform media sosial sekaligus mikroblogging yang memungkinkan penggunaannya untuk mengirim pesan singkat berbentuk teks hingga 280 karakter, yang dikenal sebagai “tweet”[10]. Sejak didirikan pada tahun 2006, Twitter telah berkembang menjadi salah satu platform digital paling populer di dunia dengan jutaan pengguna aktif setiap hari. Platform ini menawarkan kemudahan bagi pengguna untuk berbagi informasi secara cepat, menemukan

konten berdasarkan topik atau hashtag tertentu, serta berinteraksi dengan pengguna lain melalui tweet yang tampil di halaman profil mereka[11].

Twitter memainkan peran signifikan di berbagai bidang, termasuk politik, bisnis, dan hiburan. Politisi sering menggunakannya untuk berkomunikasi langsung dengan pemilih, sementara perusahaan memanfaatkannya untuk promosi dan menjaga hubungan dengan pelanggan. Di dunia hiburan, selebriti memanfaatkan Twitter untuk berinteraksi dengan penggemar serta mempromosikan proyek terbaru mereka. Meski Twitter efektif dalam penyebaran informasi secara cepat, keterbatasan karakter sering kali mempermudah penyebaran berita palsu atau informasi yang salah, yang dapat menurunkan tingkat kepercayaan publik.

## 2.2 *Imbalanced Dataset*

Dataset yang tidak seimbang adalah situasi di mana distribusi kelas dalam suatu dataset sangat tidak proporsional, dengan salah satu kelas memiliki jumlah data yang jauh lebih sedikit dibandingkan kelas lainnya. Misalnya, jika dari 1000 data, hanya 10 di antaranya yang termasuk dalam kelas minoritas, maka dataset tersebut disebut sebagai *Imbalanced Dataset* [8]. Ketidakseimbangan ini sering terjadi pada masalah klasifikasi, di mana meskipun kelas minoritas jarang muncul, kelas tersebut sering kali memiliki kepentingan yang lebih besar dibandingkan kelas mayoritas. Contoh yang umum adalah deteksi penipuan kartu kredit, di mana transaksi penipuan menjadi kelas minoritas, sedangkan transaksi sah adalah kelas mayoritas[12].

Ketidakseimbangan dalam dataset menimbulkan tantangan besar dalam pengembangan model prediksi[13]. Model yang dilatih dengan dataset yang tidak seimbang cenderung kurang efektif dalam mengklasifikasikan kelas minoritas dan lebih condong untuk memprediksi kelas mayoritas karena dominasi data dari kelas tersebut. Beberapa pendekatan dapat digunakan untuk mengatasi masalah ini, seperti undersampling[14], yang mengurangi jumlah data dari kelas mayoritas, atau oversampling[15], yang menambah data dari kelas minoritas. Meskipun demikian, kedua metode ini memiliki kekurangan masing-masing dan perlu dipertimbangkan dengan hati-hati sebelum diterapkan.

## 2.3 SMOTE

*Synthetic Minority Oversampling Technique* (SMOTE) adalah metode yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset. Ketidakseimbangan kelas terjadi ketika jumlah sampel dalam satu kelas jauh lebih sedikit dibandingkan dengan kelas lainnya, yang dapat menyebabkan bias dalam model pembelajaran mesin. SMOTE bekerja dengan cara menghasilkan sampel sintesis dari kelas minoritas dengan menginterpolasi antara sampel minoritas yang ada. Teknik ini membantu meningkatkan representasi kelas minoritas dalam dataset, sehingga model pembelajaran mesin dapat belajar lebih baik dan menghasilkan prediksi yang lebih akurat untuk kelas minoritas[8], [16], [17].

Dalam konteks *text mining*, SMOTE memainkan peran penting dalam meningkatkan akurasi klasifikasi teks yang tidak seimbang. Misalnya, dalam klasifikasi teks, sering kali terdapat ketidakseimbangan antara jumlah dokumen yang termasuk dalam kategori mayoritas dan minoritas. Dengan menggunakan SMOTE, sampel teks dari kategori minoritas dapat ditingkatkan secara sintesis, sehingga model pembelajaran mesin dapat lebih efektif dalam mengenali dan mengklasifikasikan teks dari kategori minoritas. Hal ini sangat penting dalam aplikasi seperti analisis sentimen, di mana ulasan positif mungkin jauh lebih banyak daripada ulasan negatif, atau dalam deteksi spam, di mana email spam mungkin lebih sedikit dibandingkan dengan email non-spam[18], [19], [20].

## 2.4 RUS

*Random Under-Sampling* (RUS) adalah salah satu teknik yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset. Ketidakseimbangan kelas terjadi ketika jumlah instance dalam satu kelas jauh lebih banyak dibandingkan dengan kelas lainnya, yang dapat menyebabkan model prediktif cenderung mengabaikan kelas minoritas. RUS bekerja dengan cara mengurangi jumlah instance dari kelas mayoritas sehingga distribusi kelas menjadi lebih seimbang. Meskipun metode ini dapat membantu mengurangi bias terhadap kelas mayoritas, RUS juga memiliki kelemahan, yaitu potensi kehilangan informasi penting dari kelas mayoritas yang dapat mempengaruhi kinerja model secara keseluruhan[21].

RUS memastikan bahwa model pembelajaran mesin tidak hanya fokus pada kelas mayoritas. Dengan mengurangi jumlah instance dari kelas mayoritas, RUS membantu model untuk lebih memperhatikan dan mempelajari pola dari kelas minoritas. Hal ini sangat penting dalam aplikasi *text mining* seperti deteksi spam, analisis sentimen, dan klasifikasi dokumen di mana ketidakseimbangan kelas sering terjadi. Namun, penting untuk dicatat bahwa meskipun RUS dapat meningkatkan kinerja model pada kelas minoritas, metode ini harus digunakan dengan hati-hati karena dapat menyebabkan hilangnya informasi yang berharga dari kelas mayoritas, yang pada akhirnya dapat mempengaruhi akurasi keseluruhan model[22].

### 2.5 Random Forest

*Random Forest* adalah salah satu algoritma machine learning yang digunakan untuk prediksi dan klasifikasi. Algoritma ini memanfaatkan metode ensemble learning, di mana banyak model pohon keputusan (*decision trees*) digabungkan menjadi satu "hutan" (*forest*) dengan tujuan meningkatkan akurasi dan performa prediksi. Kelebihan utama dari *Random Forest* adalah kemampuannya mengatasi overfitting dan meningkatkan kemampuan generalisasi model[5].

Pohon keputusan adalah struktur hierarkis yang membantu pengambilan keputusan berdasarkan aturan tertentu. Dalam *Random Forest*, beberapa pohon keputusan dibangun secara acak menggunakan sampel data dan subset fitur yang berbeda. Prediksi akhir dihasilkan dari kombinasi prediksi setiap pohon, melalui metode voting mayoritas atau rata-rata. Pendekatan ini mengurangi variabilitas dan risiko overfitting, sehingga membuat *Random Forest* menjadi model yang kuat dan andal.

### 2.6 Maximum Entropy

*Maximum Entropy* adalah pendekatan dalam statistika dan teori informasi yang digunakan untuk menentukan distribusi probabilitas yang paling sesuai dengan sejumlah batasan yang ada. "*Entropi*" dalam konteks ini mengacu pada tingkat ketidakpastian atau keacakan dalam suatu sistem. Prinsip utama *Maximum Entropy* adalah memilih distribusi probabilitas dengan entropi tertinggi, yang berarti menghindari asumsi atau pengetahuan awal yang tidak dapat dibuktikan. Pendekatan ini dapat diterapkan dalam berbagai bidang, seperti pemodelan bahasa alami, pengenalan pola, atau pemrosesan sinyal. Persamaan *Maximum Entropy* tercantum dalam Persamaan 1.

$$H(X) = - \sum_x \frac{P(x)}{\log P(x)} P(x) = \frac{e^{(wx)}}{Z(w)} \dots\dots\dots (1)$$

### 2.7 Neural Network

*Neural Network* adalah model komputasi yang terinspirasi oleh struktur dan fungsi otak manusia. Model ini dirancang untuk mengenali pola, mengklasifikasikan data, dan membuat prediksi dengan belajar dari contoh. Jaringan saraf telah diterapkan di berbagai bidang, termasuk statistika, geofisika, radiologi, dan pembelajaran mesin.

Klasifikasi emosi menggunakan *Neural Network* pada data Twitter telah menjadi area penelitian yang penting karena kaya akan konten emosional yang diekspresikan dalam tweet. Berbagai arsitektur pembelajaran mendalam telah dieksplorasi untuk meningkatkan akurasi dan efisiensi dalam klasifikasi emosi.

*Neural Network* telah terbukti sangat efektif untuk klasifikasi emosi dalam tweet. Model-model ini melampaui pendekatan pembelajaran mesin tradisional dan mendapatkan manfaat signifikan dari teknik prapemrosesan dan embeddin[23]. Kerangka kerja pembelajaran multimodal dan multitugas lebih meningkatkan akurasi dengan memanfaatkan keterkaitan antara emosi dan sentimen. Secara keseluruhan, integrasi metode pembelajaran mendalam dalam analisis emosi tweet menawarkan pendekatan yang kuat untuk memahami dan mengklasifikasikan konten emosional dalam data media sosial[24].

### 2.8 Support Vector Machine

*Support Vector Machine* (SVM) adalah algoritma pembelajaran mesin yang sangat efektif dalam memisahkan kelas-kelas dalam tugas klasifikasi. Algoritma ini bekerja dengan mencari hiperbidang optimal yang memisahkan dua kelas data yang berbeda dalam ruang berdimensi tinggi. Hiperbidang ini disebut "margin maksimal" dan dipilih sedemikian rupa agar jarak antara hiperbidang dan titik data terdekat dari kedua kelas, yang dikenal sebagai vektor pendukung

(*support vector*), adalah sebesar mungkin. Keunggulan ini membuat SVM ideal untuk menangani masalah klasifikasi pada data yang tidak linier dan berdimensi tinggi[25].

Pendekatan SVM didasarkan pada optimasi matematis dengan dua tujuan utama: memaksimalkan margin antar kelas dan meminimalkan kesalahan klasifikasi. SVM juga menggunakan fungsi kernel untuk mentransformasi data ke dalam ruang fitur yang lebih tinggi, sehingga memungkinkan data yang awalnya tidak dapat dipisahkan secara linier menjadi dapat dipisahkan. Beberapa jenis fungsi kernel yang digunakan mencakup kernel *linier*, kernel *polinomial*, dan kernel *radial basic function* (RBF), yang dipilih berdasarkan karakteristik data. Persamaan SVM dapat dilihat pada Persamaan 2.

$$(f(x) = wx + b) = \frac{wx+b}{\|w\|} \text{Margin} = \frac{2}{\|w\|} \text{minimize} \frac{1/2w^2}{\|w\|} \text{subject to } (wx + b) > 1 \dots\dots\dots(2)$$

## 2.9 Naïve Bayes Classification

*Naive Bayes Classification* adalah metode klasifikasi dalam machine learning yang didasarkan pada Teorema Bayes[26]. Metode ini sangat bermanfaat dalam berbagai tugas klasifikasi, seperti pengenalan teks, pengkategorian email sebagai spam atau tidak, dan banyak aplikasi lainnya. Naive Bayes didasarkan pada asumsi "naif" yang menyatakan bahwa setiap fitur dalam data saling independen, meskipun dalam kenyataannya hal ini sering kali tidak sepenuhnya benar. Meskipun asumsi independensi ini jarang sepenuhnya terpenuhi, metode ini tetap memberikan hasil yang sangat baik dalam praktik[27]. Persamaan *Naive Bayes Classification* dapat dilihat pada Persamaan 3.

$$\left( \frac{P(C|X)}{P(X|C)} \right) = \frac{P(X|C)P(C)}{P(X)} \dots\dots\dots(3)$$

## 2.10 Dataset

Penelitian ini akan menggunakan dataset Twitter dari penelitian yang telah dilakukan sebelumnya[28], yang dapat diakses melalui tautan berikut: [https://s.id/29duL](https://s.id/29duL). Dataset tersebut terdiri dari lima kelas emosi, yaitu: anger dengan 1101 data tweet, happy dengan 1017 data tweet, sadness dengan 997 data tweet, fear dengan 649 data tweet, dan love dengan 637 data tweet.

## 2.11 Praproses

Proses prapemrosesan data teks adalah langkah-langkah untuk membersihkan dan mempersiapkan data teks agar siap digunakan dalam analisis data, yang bertujuan meningkatkan akurasi dan kinerja model pembelajaran mesin[29]. Tahapan dalam proses prapemrosesan meliputi:

1. Mengubah semua teks menjadi huruf kecil.
2. Menghapus karakter khusus, tanda baca, angka, dan simbol.
3. Menghilangkan stopwords (kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti "dan", "atau").
4. Melakukan *stemming*, yaitu menghapus imbuhan pada kata dan mengembalikannya ke bentuk dasar.

## 2.12 Vektorisasi

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode vektorisasi yang populer digunakan dalam klasifikasi emosi[30]. TF-IDF mengukur seberapa penting suatu kata dalam sebuah dokumen dengan mempertimbangkan frekuensi kemunculan kata tersebut dalam dokumen serta jumlah dokumen yang mengandung kata tersebut[31].

Dalam klasifikasi emosi, vektorisasi menggunakan TF-IDF dilakukan untuk menghitung bobot setiap kata dalam dokumen. Kata dengan bobot yang lebih tinggi dianggap lebih signifikan dalam menentukan emosi dari dokumen tersebut, sehingga membantu meningkatkan akurasi dalam analisis emosi.

$$\text{TF-IDF} = f \times N \times \log(Nn) \dots\dots\dots(4)$$

### 2.13 Penggunaan Model Sampling

Model sampling adalah pendekatan yang digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam pembelajaran mesin. Ketidakseimbangan kelas terjadi ketika satu kelas memiliki jumlah sampel yang jauh lebih sedikit dibandingkan kelas lain, yang dapat menyebabkan model pembelajaran mesin cenderung bias terhadap kelas mayoritas dan gagal menghasilkan prediksi yang akurat untuk kelas minoritas.

Model sampling melibatkan teknik oversampling dan undersampling untuk menyeimbangkan distribusi kelas dalam dataset. Oversampling menambahkan sampel ke kelas minoritas dengan menghasilkan sampel sintesis baru yang mirip dengan data yang sudah ada dalam kelas tersebut. Sementara itu, undersampling mengurangi jumlah sampel dari kelas mayoritas dengan menghapus sebagian sampel, sehingga jumlahnya seimbang dengan kelas minoritas.

Dalam penelitian ini, dua skenario dilakukan:

1. Pembangunan model menggunakan SMOTE (Synthetic Minority Over-sampling Technique) untuk oversampling kelas minoritas.
2. Pembangunan model menggunakan RUS (Random Under Sampler) untuk undersampling kelas mayoritas.

Tahap pembangunan model dengan empat algoritma yang berbeda, yaitu *Random Forest*, *Maximum Entropy*, *Neural Network*, *SVM*, dan *Naive Bayes Classification* (NBC), dalam penelitian ini bertujuan untuk menangani masalah ketidakseimbangan data pada klasifikasi emosi tweet berbahasa Indonesia. Setiap algoritma ini dipilih untuk mengeksplorasi pendekatan yang berbeda dalam menangani ketidakseimbangan kelas dan memberikan gambaran menyeluruh mengenai dampak metode ini terhadap performa klasifikasi.

Langkah-langkah ini melibatkan pengujian setiap algoritma dalam berbagai skenario untuk melihat bagaimana ketidakseimbangan data memengaruhi akurasi, sensitivitas, dan kemampuan model dalam mengklasifikasikan emosi dari tweet. Dengan menerapkan beberapa algoritma tersebut, penelitian dapat memberikan wawasan yang mendalam tentang bagaimana penanganan ketidakseimbangan data, seperti melalui model sampling, dapat meningkatkan kinerja klasifikasi emosi tweet berbahasa Indonesia.

Penelitian ini tidak hanya membandingkan kinerja antar algoritma, tetapi juga menunjukkan metode terbaik dalam menangani ketidakseimbangan data di tiap model, sehingga hasilnya dapat menjadi referensi yang komprehensif bagi masalah serupa.

## 3. Hasil dan Pembahasan

Berdasarkan hasil pengujian, performa dari keempat model klasifikasi (*Random Forest*, *Maximum Entropy*, *SVM*, *Naive Bayes Classification*, dan *Neural Network*) menunjukkan variasi pada setiap skenario penanganan ketidakseimbangan kelas (imbalanced class). Hal ini tercermin dalam tabel matriks evaluasi yang disajikan pada Tabel 1. Variasi performa ini dapat terlihat dari metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang menggambarkan seberapa baik setiap model dalam menangani ketidakseimbangan kelas dalam klasifikasi emosi tweet berbahasa Indonesia.

Dengan membandingkan performa tiap model di bawah kondisi yang berbeda, dapat diidentifikasi metode mana yang lebih efektif dalam menangani ketidakseimbangan data dan memberikan prediksi yang lebih akurat serta seimbang antara kelas mayoritas dan minoritas. Hasil ini membantu memberikan pemahaman yang lebih baik tentang kekuatan dan kelemahan setiap model dalam konteks data yang tidak seimbang.

### 4.1 Akurasi

Dalam evaluasi akurasi dari kedua skenario model sampling, *Maximum Entropy* menunjukkan performa yang konsisten sebagai model dengan akurasi tertinggi. Pada skenario Random Under Sampling (RUS), model ini mencapai akurasi sebesar 0,631, diikuti oleh *SVM* dengan akurasi 0,622. Kedua model ini menunjukkan kinerja yang lebih baik dibandingkan model lainnya dalam skenario RUS. *Neural Network* dan *Random Forest* menunjukkan akurasi yang lebih rendah, masing-masing dengan nilai 0,574 dan 0,568, sementara *Naive Bayes Classification* memiliki akurasi terendah dengan 0,553.

Dalam skenario SMOTE, akurasi *Maximum Entropy* sedikit meningkat menjadi 0,638, tertinggi di antara semua model. *SVM* juga mengalami peningkatan akurasi menjadi 0,625, menunjukkan kinerja yang stabil. *Random Forest* mengalami peningkatan akurasi yang cukup

signifikan menjadi 0,607 dari sebelumnya 0,568 dalam skenario RUS. *Neural Network* dan *Naive Bayes Classification* juga menunjukkan peningkatan akurasi, masing-masing mencapai 0,595 dan 0,580.

*Maximum Entropy* tetap menjadi model dengan akurasi tertinggi dalam kedua skenario. Penerapan SMOTE meningkatkan akurasi pada semua model, terutama pada *Random Forest* dan *Neural Network*, dibandingkan dengan skenario RUS, yang cenderung menghasilkan performa akurasi yang lebih rendah.

Tabel 1. Matrik Evaluasi

Model	Accuracy	Precision	Recall	F1-Score	Elapsed Time (s)
Model Sampling Random Under Sampling					
<i>Random Forest</i>	0,568	0,567	0,568	0,561	1,514
<i>Maximum Entropy</i>	0,631	0,630	0,631	0,630	0,344
SVM	0,622	0,625	0,622	0,623	1,470
<i>Naive Bayes Classification</i>	0,553	0,599	0,553	0,548	0,005
<i>Neural Network</i>	0,574	0,580	0,574	0,574	5,413
Model Sampling SMOTE					
<i>Random Forest</i>	0,607	0,614	0,607	0,605	2,021
<i>Maximum Entropy</i>	0,638	0,637	0,638	0,637	0,434
SVM	0,625	0,633	0,625	0,627	3,497
<i>Naive Bayes Classification</i>	0,580	0,611	0,580	0,576	0,011
<i>Neural Network</i>	0,595	0,601	0,595	0,597	8,655

Sumber: Hasil Penelitian (2024)

#### 4.2 Presisi

Dalam evaluasi presisi dari kedua skenario model sampling, *Maximum Entropy* konsisten mencapai presisi tertinggi. Pada skenario Random Under Sampling (RUS), model ini memperoleh presisi sebesar 0,630, sedikit lebih tinggi dibandingkan SVM dengan presisi 0,625. Kedua model ini menunjukkan kemampuan yang baik dalam mengidentifikasi kelas positif. Di sisi lain, model seperti *Neural Network* dan *Naive Bayes Classification* memiliki presisi yang lebih rendah dalam skenario RUS, masing-masing sebesar 0,580 dan 0,599, sementara *Random Forest* mencatatkan nilai presisi terendah dengan 0,567.

Pada skenario SMOTE, *Maximum Entropy* tetap unggul dengan presisi 0,637, dan SVM juga tetap kompetitif dengan nilai presisi 0,633. *Random Forest* dan *Neural Network* mengalami peningkatan presisi yang signifikan dibandingkan dengan skenario RUS, dengan masing-masing nilai 0,614 dan 0,601. *Naive Bayes Classification* juga mengalami peningkatan presisi menjadi 0,611.

Secara keseluruhan, penerapan SMOTE memberikan dampak positif pada presisi semua model, dengan peningkatan yang lebih jelas dibandingkan dengan RUS. *Maximum Entropy* dan SVM secara konsisten menjadi model dengan presisi tertinggi dalam kedua skenario.

#### 4.3 Recall

Dalam evaluasi recall dari kedua skenario model sampling, *Maximum Entropy* menunjukkan performa yang konsisten sebagai model dengan nilai recall tertinggi. Pada skenario Random Under Sampling (RUS), *Maximum Entropy* mencapai recall sebesar 0,631, diikuti oleh SVM dengan nilai recall 0,622. Kedua model ini menunjukkan kemampuan yang baik dalam mendeteksi kelas positif. *Neural Network* dan *Random Forest* memiliki nilai recall yang lebih rendah dalam skenario ini, masing-masing sebesar 0,574 dan 0,568, sementara *Naive Bayes Classification* menunjukkan performa terendah dengan nilai recall 0,553.

Pada skenario SMOTE, *Maximum Entropy* tetap memimpin dengan nilai recall sebesar 0,638, sedikit lebih tinggi dibandingkan skenario RUS. SVM juga mengalami peningkatan menjadi 0,625, sementara *Random Forest* menunjukkan perbaikan signifikan dengan nilai recall 0,607, naik dari 0,568 pada skenario RUS. *Neural Network* dan *Naive Bayes Classification* juga mengalami peningkatan recall masing-masing menjadi 0,595 dan 0,580.

Penerapan SMOTE cenderung meningkatkan recall pada semua model, terutama pada *Random Forest* dan *Neural Network*. *Maximum Entropy* dan SVM terus menunjukkan kinerja

terbaik dalam kedua skenario, dengan recall yang tinggi dan stabil, menandakan kemampuan yang baik dalam mengidentifikasi kelas positif secara menyeluruh.

#### 4.4 F1-Score

F1-score pada kedua skenario model sampling, *Maximum Entropy* menunjukkan performa yang paling konsisten dengan nilai tertinggi. Pada skenario Random Under Sampling (RUS), *Maximum Entropy* mencapai F1-score sebesar 0,630, diikuti oleh SVM dengan nilai 0,623, menandakan keseimbangan yang baik antara precision dan recall pada kedua model ini. *Neural Network* dan *Random Forest* menunjukkan F1-score yang lebih rendah, masing-masing sebesar 0,574 dan 0,561, sementara *Naive Bayes Classification* memiliki F1-score terendah pada skenario ini dengan nilai 0,548.

Pada skenario SMOTE, *Maximum Entropy* tetap unggul dengan F1-score 0,637, sedikit lebih tinggi dibandingkan skenario RUS. SVM juga mengalami peningkatan, mencapai F1-score 0,627. *Random Forest* mencatat peningkatan signifikan dalam F1-score menjadi 0,605, naik dari 0,561 pada skenario RUS. *Neural Network* dan *Naive Bayes Classification* juga mengalami peningkatan dengan F1-score masing-masing sebesar 0,597 dan 0,576.

Penerapan SMOTE cenderung meningkatkan F1-score untuk semua model, dengan *Maximum Entropy* dan SVM mempertahankan performa terbaik dalam kedua skenario. *Random Forest* dan *Neural Network* juga mengalami peningkatan signifikan dalam F1-score setelah penerapan SMOTE, sementara *Naive Bayes Classification* tetap berada di posisi terendah namun menunjukkan sedikit peningkatan.

#### 4.4 Waktu Eksekusi (*Elapsed Time*)

Dalam evaluasi waktu eksekusi (*elapsed time*) pada kedua skenario model sampling, terdapat variasi yang signifikan di antara model-model yang diuji. Pada skenario Random Under Sampling (RUS), *Naive Bayes Classification* menunjukkan waktu eksekusi tercepat dengan hanya 0,005 detik, menjadikannya model yang paling efisien dari segi waktu. *Maximum Entropy* juga memiliki waktu eksekusi yang cepat, yaitu 0,344 detik, diikuti oleh SVM dan *Random Forest* dengan waktu masing-masing 1,470 detik dan 1,514 detik. *Neural Network*, sebagai model yang lebih kompleks, menunjukkan waktu eksekusi paling lama, yaitu 5,413 detik, yang mencerminkan biaya komputasi yang lebih tinggi.

Dalam skenario SMOTE, waktu eksekusi semua model meningkat, terutama karena proses oversampling yang lebih intensif. *Naive Bayes Classification* masih menjadi model tercepat dengan 0,011 detik, sementara *Maximum Entropy* tetap cukup cepat dengan 0,434 detik. *Random Forest* mengalami sedikit peningkatan waktu eksekusi menjadi 2,021 detik, dan SVM memiliki waktu yang cukup tinggi, yaitu 3,497 detik. *Neural Network* kembali menjadi model dengan waktu eksekusi terlama, yaitu 8,655 detik, hampir dua kali lebih lama dibandingkan skenario RUS.

Secara keseluruhan, *Naive Bayes Classification* dan *Maximum Entropy* adalah model yang paling efisien dari segi waktu di kedua skenario. *Neural Network*, meskipun kuat dari segi kinerja, membutuhkan waktu eksekusi yang jauh lebih lama, terutama ketika menggunakan SMOTE. Model seperti SVM dan *Random Forest* menunjukkan waktu eksekusi yang moderat, tetapi masih membutuhkan lebih banyak waktu dibandingkan dengan model linear yang lebih sederhana.

#### 4.5 Analisa

Dalam penelitian ini, evaluasi dilakukan terhadap akurasi dari beberapa model klasifikasi dalam dua skenario model sampling, yaitu Random Under Sampling (RUS) dan SMOTE. Hasilnya menunjukkan bahwa *Maximum Entropy* adalah model dengan akurasi tertinggi pada kedua skenario. Pada skenario RUS, model ini mencapai akurasi sebesar 0,631, sedikit lebih unggul dari SVM yang memiliki akurasi 0,622. Model lain seperti *Neural Network* dan *Random Forest* mencatat akurasi yang lebih rendah, masing-masing sebesar 0,574 dan 0,568, sementara *Naive Bayes Classification* memiliki akurasi terendah dengan 0,553. Dalam skenario SMOTE, semua model mengalami peningkatan akurasi, dengan *Maximum Entropy* naik menjadi 0,638 dan SVM menjadi 0,625. *Random Forest* dan *Neural Network* juga menunjukkan peningkatan yang signifikan.

Dari segi presisi, *Maximum Entropy* kembali menjadi model dengan presisi tertinggi pada kedua skenario. Pada skenario RUS, presisi model ini mencapai 0,630, diikuti oleh SVM dengan



0,625. *Neural Network* dan *Naive Bayes Classification* memiliki presisi yang lebih rendah, masing-masing 0,580 dan 0,599, sementara *Random Forest* mencatatkan nilai presisi terendah sebesar 0,567. Dengan penerapan SMOTE, presisi untuk semua model mengalami peningkatan, di mana *Maximum Entropy* tetap memimpin dengan 0,637, disusul oleh SVM dengan 0,633. *Random Forest* dan *Neural Network* mencatat peningkatan presisi yang signifikan dibandingkan skenario RUS.

Pada metrik recall, *Maximum Entropy* juga menunjukkan performa yang konsisten dengan nilai tertinggi. Pada skenario RUS, recall model ini adalah 0,631, diikuti oleh SVM dengan 0,622. Model lainnya, seperti *Neural Network* dan *Random Forest*, mencatat nilai recall yang lebih rendah, yaitu 0,574 dan 0,568. *Naive Bayes Classification* menunjukkan performa terendah dengan nilai recall 0,553. Pada skenario SMOTE, nilai recall *Maximum Entropy* meningkat menjadi 0,638, dan SVM juga menunjukkan peningkatan menjadi 0,625. Semua model mengalami peningkatan recall, terutama pada *Random Forest* dan *Neural Network*, yang menunjukkan perbaikan signifikan dalam mendeteksi kelas positif.

Terakhir, untuk F1-score, *Maximum Entropy* sekali lagi tampil sebagai model dengan nilai tertinggi, baik dalam skenario RUS maupun SMOTE. Pada skenario RUS, model ini mencapai F1-score sebesar 0,630, dengan SVM sedikit di bawahnya pada 0,623. Model lainnya, seperti *Random Forest* dan *Neural Network*, memiliki F1-score yang lebih rendah, masing-masing 0,561 dan 0,574. *Naive Bayes Classification* mencatat F1-score terendah dengan 0,548. Dalam skenario SMOTE, semua model mengalami peningkatan F1-score, dengan *Maximum Entropy* meningkat menjadi 0,637, diikuti oleh SVM yang juga naik menjadi 0,627. *Random Forest* dan *Neural Network* juga mencatat peningkatan signifikan dalam F1-score setelah penerapan SMOTE.

Hasil evaluasi menunjukkan bahwa *Maximum Entropy* consistently memberikan performa terbaik dalam hampir semua metrik, termasuk akurasi, presisi, recall, dan F1-score, baik pada skenario Random Under Sampling (RUS) maupun SMOTE. Model ini unggul dalam keseimbangan antara kemampuan mendeteksi kelas positif dan mempertahankan akurasi secara keseluruhan. SVM juga menunjukkan performa yang kompetitif di seluruh metrik, meskipun sedikit di bawah *Maximum Entropy*. Sementara itu, *Random Forest* dan *Neural Network* mengalami peningkatan yang signifikan setelah penerapan SMOTE, terutama dalam recall dan F1-score, meskipun tetap tertinggal dibandingkan *Maximum Entropy* dan SVM. Di sisi lain, *Naive Bayes Classification* meskipun efisien dalam waktu eksekusi, cenderung memiliki performa terendah dalam akurasi dan F1-score, menunjukkan bahwa model ini kurang optimal dalam menangani ketidakseimbangan data meskipun penerapan SMOTE memberikan sedikit perbaikan. Penerapan SMOTE secara umum berhasil meningkatkan performa model, terutama pada *Random Forest* dan *Neural Network*, sementara RUS cenderung menghasilkan performa yang lebih rendah.

Temuan dalam penelitian ini menunjukkan bahwa *Maximum Entropy* secara konsisten memberikan performa terbaik di hampir semua metrik evaluasi, baik pada skenario *Random Under Sampling (RUS)* maupun *SMOTE*. Hal ini menandakan bahwa model ini memiliki keunggulan dalam menyeimbangkan kemampuan mendeteksi kelas positif sekaligus mempertahankan akurasi keseluruhan pada dataset yang tidak seimbang. Peningkatan performa yang dihasilkan oleh *Maximum Entropy* dalam skenario *SMOTE*, dengan akurasi sebesar 0,638 dan F1-score 0,637, memberikan bukti bahwa teknik penanganan ketidakseimbangan data dapat memperbaiki hasil klasifikasi, khususnya untuk teks pendek seperti tweet berbahasa Indonesia. Bila dibandingkan dengan penelitian-penelitian sebelumnya, hasil ini lebih unggul karena kebanyakan studi terdahulu hanya mengeksplorasi teknik dasar seperti oversampling atau undersampling tanpa mengevaluasi kombinasi teknik tersebut dengan berbagai algoritma klasifikasi pada data berbahasa Indonesia. Penelitian ini juga menyoroti bahwa model seperti SVM tetap kompetitif dalam sebagian besar metrik, namun *Maximum Entropy* mampu memberikan hasil yang lebih stabil di semua skenario, menjadikannya pilihan yang lebih andal untuk tugas klasifikasi emosi pada dataset yang tidak seimbang.

Penelitian ini menawarkan evaluasi komprehensif terhadap dua teknik penanganan ketidakseimbangan data, yaitu RUS dan *SMOTE*, yang diterapkan pada tweet berbahasa Indonesia, sebuah domain yang masih jarang diteliti secara mendalam. Tidak hanya mengevaluasi kinerja berbagai algoritma klasifikasi seperti *Maximum Entropy*, SVM, *Random Forest*, *Neural Network*, dan *Naive Bayes*, penelitian ini juga memberikan wawasan mengenai respons masing-masing model terhadap perubahan distribusi data setelah penerapan teknik

penyeimbangan. Selain itu, penelitian ini memanfaatkan dataset berbahasa Indonesia yang menghadirkan tantangan unik, seperti variasi gaya bahasa, penggunaan slang, dan struktur sintaks, yang membedakannya secara signifikan dari dataset berbahasa Inggris yang lebih sering menjadi fokus dalam penelitian sebelumnya.

Namun, penelitian ini juga menghadapi beberapa kendala yang perlu diperhatikan. Salah satu kendala utama adalah ketidakseimbangan data yang ekstrem dalam beberapa kelas emosi, yang bahkan setelah penerapan SMOTE atau RUS, tetap memengaruhi performa model, khususnya pada kelas minoritas dengan jumlah sampel yang sangat kecil. Selain itu, meskipun SMOTE mampu meningkatkan kinerja model secara keseluruhan, teknik ini dapat menyebabkan overfitting pada beberapa algoritma, terutama Random Forest, karena data sintesis yang dihasilkan tidak selalu merepresentasikan pola alami dalam dataset asli. Kendala lainnya adalah waktu komputasi yang cukup tinggi untuk model kompleks seperti Neural Network, terutama pada skenario SMOTE yang membutuhkan pengolahan data tambahan. Terakhir, pengaruh nuansa bahasa Indonesia, termasuk penggunaan kata-kata slang dan akronim, menjadi tantangan tersendiri yang dapat memengaruhi akurasi model, mengingat sebagian besar algoritma tidak dirancang secara khusus untuk menangani variasi bahasa tersebut.

Meskipun terdapat kendala-kendala tersebut, penelitian ini memberikan kontribusi penting dalam memahami pengaruh teknik penanganan ketidakseimbangan data pada tugas klasifikasi emosi dalam tweet berbahasa Indonesia dan menawarkan arah baru bagi penelitian lebih lanjut dalam domain ini.

#### 4. Kesimpulan

Kesimpulan dari perbandingan penggunaan model sampling menunjukkan bahwa penerapan SMOTE secara keseluruhan lebih efektif dalam meningkatkan performa model dibandingkan dengan Random Under Sampling (RUS). SMOTE berhasil meningkatkan akurasi, presisi, recall, dan F1-score pada hampir semua model, terutama pada *Random Forest* dan *Neural Network* yang menunjukkan peningkatan signifikan. Sebaliknya, RUS cenderung menghasilkan performa yang lebih rendah karena pengurangan sampel dari kelas mayoritas yang dapat menghilangkan informasi penting. *Maximum Entropy* dan SVM tetap menjadi model yang paling unggul dalam kedua skenario, namun peningkatan performa lebih terlihat ketika menggunakan SMOTE. Hal ini menunjukkan bahwa SMOTE adalah teknik yang lebih efektif dalam menangani ketidakseimbangan data dibandingkan RUS.

Penelitian selanjutnya dapat difokuskan pada pengembangan teknik penanganan ketidakseimbangan data yang lebih canggih, seperti ADASYN atau metode berbasis *Generative Adversarial Networks (GANs)*, untuk menghasilkan data sintesis yang lebih representatif terhadap kelas minoritas. Selain itu, pengembangan model yang dirancang khusus untuk bahasa Indonesia, seperti penyesuaian model berbasis transformer misalnya *IndoBERT*, dapat menjadi langkah penting dalam mengatasi tantangan linguistik unik. Peningkatan kualitas dataset juga menjadi prioritas, dengan mengumpulkan data dari berbagai sumber dan memastikan anotasi yang lebih akurat oleh para ahli. Pendekatan baru seperti integrasi data multi-modal yang mencakup teks, gambar, atau metadata, juga dapat memberikan pemahaman yang lebih kaya terhadap emosi dalam konteks sosial tertentu. Penelitian longitudinal untuk mempelajari perubahan pola emosi dari waktu ke waktu juga dapat dilakukan, khususnya dalam menanggapi isu-isu sosial atau tren tertentu. Selain itu, pengembangan sistem real-time untuk memprediksi dan memantau emosi publik dapat membantu dalam pengambilan keputusan cepat di berbagai bidang. Untuk memastikan hasil yang lebih mendalam, evaluasi dengan metrik tambahan seperti AUC atau *cost-sensitive metrics* dapat digunakan untuk mengukur keseimbangan kinerja model secara lebih menyeluruh. Dengan langkah-langkah ini, penelitian di masa depan dapat menghasilkan kontribusi yang lebih komprehensif dan aplikatif di berbagai domain.

#### Ucapan Terima Kasih (Opsional)

Peneliti mengucapkan banyak terimakasih ITB STIKOM Bali yang telah memberikan pembiayaan penelitian ini dan LPPM ITB STIKOM Bali telah mendukung selama berjalannya kegiatan penelitian.

## Referensi

- [1] A. K. Fauziyyah, 'ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON', *J. Ilm. SINUS*, vol. 18, no. 2, Art. no. 2, Jul. 2020, doi: 10.30646/sinus.v18i2.491.
- [2] A. C. Sanders *et al.*, 'Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of COVID-19 Twitter discourse', *AMIA Summits Transl. Sci. Proc.*, vol. 2021, pp. 555–564, May 2021.
- [3] E. M. O. N. Haryanto, A. K. A. Estetikha, and R. A. Setiawan, 'IMPLEMENTASI SMOTE UNTUK MENGATASI IMBALANCED DATA PADA SENTIMEN ANALISIS SENTIMEN HOTEL DI NUSA TENGGARA BARAT DENGAN MENGGUNAKAN ALGORITMA SVM', *Inf. Interaktif*, vol. 7, no. 1, Art. no. 1, Jan. 2022.
- [4] A. Nurhopipah and C. Magnolia, 'PERBANDINGAN METODE RESAMPLING PADA IMBALANCED DATASET UNTUK KLASIFIKASI KOMENTAR PROGRAM MBKM', *J. Publ. Ilmu Komput. Dan Multimed.*, vol. 2, no. 1, Art. no. 1, Jan. 2023, doi: 10.55606/jupikom.v2i1.862.
- [5] B. B. Baskoro, I. Susanto, and S. Khomsah, 'Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)', *INISTA J. Inform. Inf. Syst. Softw. Eng. Appl.*, vol. 3, no. 2, Art. no. 2, Jun. 2021, doi: 10.20895/inista.v3i2.218.
- [6] R. Bayari and A. Bensefia, 'Text Mining Techniques for Cyberbullying Detection: State of the Art', *Adv. Sci. Technol. Eng. Syst. J.*, vol. 6, no. 1, pp. 783–790, Feb. 2021, doi: 10.25046/aj060187.
- [7] I. K. Dharmendra, R. A. N. Diaz, M. Samsudin, I. G. A. N. R. Semadi, and I. M. A. W. Putra, 'TEXT MINING UNTUK MENDETEKSI EMOSI PENGGUNA TERHADAP "NUSANTARA" SEBAGAI NAMA IKN', *J. Teknol. Inf. Dan Komput.*, vol. 9, no. 5, Oct. 2023, doi: 10.36002/jutik.v9i5.2639.
- [8] N. A. Azhar, M. S. Mohd Pozi, A. Mohamed Din, and A. Jatowt, 'An Investigation of SMOTE based Methods for Imbalanced Datasets with Data Complexity Analysis', *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2022, doi: 10.1109/TKDE.2022.3179381.
- [9] C. E. Lopez and C. Gallemore, 'An augmented multilingual Twitter dataset for studying the COVID-19 infodemic', *Soc. Netw. Anal. Min.*, vol. 11, no. 1, p. 102, Oct. 2021, doi: 10.1007/s13278-021-00825-0.
- [10] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, 'ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI', *J. Teknoinfo*, vol. 14, no. 2, Art. no. 2, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [11] E. Utami, I. Oyong, S. Raharjo, A. Dwi Hartanto, and S. Adi, 'Supervised learning and resampling techniques on DISC personality classification using Twitter information in Bahasa Indonesia', *Appl. Comput. Inform.*, vol. ahead-of-print, no. ahead-of-print, Jan. 2021, doi: 10.1108/ACI-03-2021-0054.
- [12] C. V. Priscilla and D. P. Prabha, 'A two-phase feature selection technique using mutual information and XGB-RFE for credit card fraud detection', *Int. J. Adv. Technol. Eng. Explor.*, vol. 8, no. 85, Dec. 2021, doi: 10.19101/IJATEE.2021.874615.
- [13] V. H. Barella, L. P. F. Garcia, M. C. P. De Souto, A. C. Lorena, and A. C. P. L. F. De Carvalho, 'Assessing the data complexity of imbalanced datasets', *Inf. Sci.*, vol. 553, pp. 83–109, Apr. 2021, doi: 10.1016/j.ins.2020.12.006.
- [14] M. Koziarski, 'CSMOUTE: Combined Synthetic Oversampling and Undersampling Technique for Imbalanced Data Classification', Apr. 01, 2020. doi: 10.48550/arXiv.2004.03409.
- [15] F. Last, G. Douzas, and F. Bacao, 'Oversampling for Imbalanced Learning Based on K-Means and SMOTE', *Inf. Sci.*, vol. 465, pp. 1–20, Oct. 2018, doi: 10.1016/j.ins.2018.06.056.
- [16] B. S. Raghuvanshi and S. Shukla, 'SMOTE based class-specific extreme learning machine for imbalanced learning', *Knowl.-Based Syst.*, vol. 187, p. 104814, Jan. 2020, doi: 10.1016/j.knosys.2019.06.022.
- [17] T. Pan, J. Zhao, W. Wu, and J. Yang, 'Learning imbalanced datasets based on SMOTE and Gaussian distribution', *Inf. Sci.*, vol. 512, pp. 1214–1233, Feb. 2020, doi: 10.1016/j.ins.2019.10.048.

- [18] S. Shaikh, S. M. Daudpota, A. S. Imran, and Z. Kastrati, 'Towards Improved Classification Accuracy on Highly Imbalanced Text Dataset Using Deep Neural Language Models', *Appl. Sci.*, vol. 11, no. 2, p. 869, Jan. 2021, doi: 10.3390/app11020869.
- [19] A. Khurana and O. P. Verma, 'Optimal Feature Selection for Imbalanced Text Classification', *IEEE Trans. Artif. Intell.*, vol. 4, no. 1, pp. 135–147, Feb. 2023, doi: 10.1109/TAI.2022.3144651.
- [20] Asniar, N. U. Maulidevi, and K. Surendro, 'SMOTE-LOF for noise identification in imbalanced data classification', *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3413–3423, Jun. 2022, doi: 10.1016/j.jksuci.2021.01.014.
- [21] J. M. Johnson and T. M. Khoshgoftaar, 'The Effects of Data Sampling with Deep Learning and Highly Imbalanced Big Data', *Inf. Syst. Front.*, vol. 22, no. 5, pp. 1113–1131, Oct. 2020, doi: 10.1007/s10796-020-10022-7.
- [22] J. Zhao, J. Jin, S. Chen, R. Zhang, B. Yu, and Q. Liu, 'A weighted hybrid ensemble method for classifying imbalanced data', *Knowl.-Based Syst.*, vol. 203, p. 106087, Sep. 2020, doi: 10.1016/j.knosys.2020.106087.
- [23] M. A. Tocoglu, O. Ozturkmenoglu, and A. Alpkocak, 'Emotion Analysis From Turkish Tweets Using Deep Neural Networks', *IEEE Access*, vol. 7, pp. 183061–183069, 2019, doi: 10.1109/ACCESS.2019.2960113.
- [24] Naufal Hilmiaji, Kemas Muslim Lhaksmana, and Mahendra Dwifabri Purbolaksono, 'Identifying Emotion on Indonesian Tweets using Convolutional Neural Networks', *J. REST/Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 3, pp. 584–593, Jun. 2021, doi: 10.29207/resti.v5i3.3137.
- [25] S. Styawati, A. Nurkholis, A. A. Aldino, S. Samsugi, E. Suryati, and R. P. Cahyono, 'Sentiment Analysis on Online Transportation Reviews Using Word2Vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm', in *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, Jan. 2022, pp. 163–167. doi: 10.1109/ISMODE53584.2022.9742906.
- [26] V. Ariandi, F. F. H. Hadi, and Shary Armonitha Lusinia, 'Sistem Pakar Untuk Mendiagnosa Penyakit Dada Dengan Menggunakan Metode Naives Bayes Dan Bahasa Pemrograman Php Dan Mysql', *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 1, Art. no. 1, Jan. 2022, doi: 10.47233/jteksis.v4i1.345.
- [27] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, 'SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE', *J. Ilmu Komput. Dan Inf.*, vol. 13, no. 2, pp. 109–122, Jul. 2020, doi: 10.21609/jiki.v13i2.885.
- [28] M. S. Saputri, R. Mahendra, and M. Adriani, 'Emotion Classification on Indonesian Twitter Dataset', in *2018 International Conference on Asian Language Processing (IALP)*, Nov. 2018, pp. 90–95. doi: 10.1109/IALP.2018.8629262.
- [29] I. K. Dharmendra, N. N. U. Januhari, I. P. Ramayasa, and I. M. A. W. Putra, 'Uji Komparasi Sentiment Analysis Pada Opini Alumni Terhadap Perguruan Tinggi', *J. Tek. Inform. UNIKA St. Thomas*, pp. 1–6, May 2022, doi: 10.54367/jtiust.v7i1.1748.
- [30] Riza Adrianti Supono and Muhammad Azis Suprayogi, 'Perbandingan Metode TF-ABS dan TF-IDF Pada Klasifikasi Teks Helpdesk Menggunakan K-Nearest Neighbor', *J. REST/Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 5, pp. 911–918, Oct. 2021, doi: 10.29207/resti.v5i5.3403.
- [31] M. Das, S. K., and P. J. A. Alphonse, 'A Comparative Study on TF-IDF feature Weighting Method and its Analysis using Unstructured Dataset', 2023, *arXiv*. doi: 10.48550/ARXIV.2308.04037.