

# ANALISIS SENTIMEN REVIEW APLIKASI SATU SEHAT MOBILE MENGGUNAKAN MODEL SAMPLING TOMEK LINKS

Ida Ayu Mirah Cahya Dewi<sup>1)</sup>, I Komang Dharmendra<sup>2)</sup>, Ni Wayan Setiasih<sup>3)</sup>

Program Studi Sistem Informasi<sup>1) 2) 3)</sup>

Fakultas Informatika dan Komputer, Institut Teknologi dan Bisnis STIKOM Bali, Bali<sup>1) 2) 3)</sup>  
mirahcahyadewi@stikom-bali.ac.id, dharmendra@stikom-bali.ac.id, setiasih@stikom-bali.ac.id

## ABSTRACT

*This study focuses on sentiment analysis in texts, particularly concerning user opinions on the Satu Sehat Mobile Application in Indonesia. Sentiment analysis is instrumental in understanding the emotions expressed in user opinions about the service. Both quantitative and qualitative methods are employed for text analysis, involving statistical techniques and interpretative approaches. The Satu Sehat Mobile Application enhances healthcare accessibility by providing disease information, nearby clinics, and online reservation services. Imbalanced data can influence sentiment analysis, leading to model bias. The solution involves using sampling methods like Tomek Links to balance the data. Models using the Random Forest, Neural Network, KNN, and SVM algorithms are constructed and evaluated using accuracy, precision, recall, and F1-Score metrics for improved sentiment analysis, particularly on imbalanced data from user opinions on the Satu Sehat Mobile Application. Evaluation of performance is conducted on data without and with sampling. SVM yields the highest accuracy without sampling (80.528%), as it emphasizes margin optimization. However, Random Forest excels in Precision, Recall, and F1-Score for minority classes. In the sampling model scenario, all models experience accuracy improvement, notably SVM with a 0.489% increase. The application of Tomek Links sampling method enhances matrix evaluation and addresses class imbalance. The sampling model also reduces processing time by working with a smaller dataset after instance removal. This method effectively mitigates dataset imbalance, enhances model performance, and reduces processing time.*

**Keywords:** Sentiment Analysis, Satu Sehat Mobile, Review, Imbalanced Dataset, Tomek Links.

## ABSTRAK

Penelitian ini fokus pada analisis sentimen dalam teks, khususnya terkait opini pengguna Aplikasi Satu Sehat *Mobile* di Indonesia. Analisis sentimen berguna untuk memahami emosi yang terungkap dalam opini pengguna terhadap layanan. Metode kuantitatif dan kualitatif digunakan untuk menganalisis teks, dengan teknik statistik dan pendekatan interpretatif. Aplikasi Satu Sehat *Mobile* membantu akses kesehatan dengan informasi penyakit, klinik terdekat, dan layanan reservasi online. Kondisi data tidak seimbang dapat memengaruhi analisis sentimen, menghasilkan bias dalam model. Solusinya adalah menggunakan metode sampling seperti *Tomek Links* untuk seimbangkan data. Model algoritma *Random Forest*, *Neural Network*, *KNN*, dan *SVM* dibangun dan dievaluasi dengan metrik akurasi, presisi, *recall*, dan *F1-Score* untuk analisis sentimen yang lebih tepat, khususnya pada data tidak seimbang dari opini pengguna Aplikasi Satu Sehat *Mobile*. Dimana evaluasi performa dilakukan pada data tanpa dan dengan model sampling. Penggunaan *SVM* menghasilkan akurasi tertinggi tanpa sampling (80,528%), karena fokus pada margin. Namun, *Random Forest* memiliki *Precision*, *Recall*, dan *F1-Score* lebih unggul untuk kelas minoritas. Dalam skenario model sampling, semua model mengalami peningkatan akurasi, terutama *SVM* dengan peningkatan 0,489%. Penerapan metode sampling *Tomek Links* memperbaiki evaluasi matriks dan mengatasi ketidakseimbangan kelas. Model sampling juga mengurangi waktu proses dengan ukuran dataset yang lebih kecil setelah penghapusan instance. Metode ini efektif mengatasi masalah ketidakseimbangan dataset, meningkatkan kinerja model, dan mengurangi waktu proses.

**Kata Kunci :** Analisa Sentimen, Satu Sehat *Mobile*, Review, Dataset Tidak Seimbang, *Tomek Links*.

## PENDAHULUAN

Analisa sentimen adalah proses identifikasi, ekstraksi, dan klasifikasi emosi yang diungkapkan dalam teks[1]. Analisa sentimen dapat digunakan untuk berbagai tujuan, seperti untuk memahami opini pengguna tentang produk atau layanan, untuk mendeteksi *cyberbullying* atau *trolling*, dan untuk menganalisis sentimen publik tentang suatu topik[2].

Opini pengguna aplikasi dapat menunjukkan bagaimana pengalaman pengguna dalam menggunakan aplikasi[3]. Analisa sentimen dapat digunakan untuk memahami emosi yang diungkapkan dalam opini pengguna aplikasi, yang dapat membantu pengembang aplikasi untuk meningkatkan kualitas aplikasi yang dikembangkan seperti yang dilakukan pada SATU SEHAT *Mobile*.

Aplikasi Satu Sehat *Mobile* adalah sebuah aplikasi *Mobile* yang dirancang untuk membantu masyarakat Indonesia dalam mengakses informasi kesehatan secara mudah dan cepat[4]. Aplikasi ini merupakan pengembangan dari aplikasi Peduli Lindungi yang dibuat oleh Kementerian Kesehatan Indonesia dengan tujuan untuk meningkatkan kesadaran masyarakat akan kesehatan serta memberikan akses mudah ke layanan kesehatan. Aplikasi Satu Sehat *Mobile* menyediakan berbagai fitur, antara lain informasi mengenai penyakit dan gejala, informasi klinik dan rumah sakit terdekat, serta tips kesehatan. Aplikasi ini juga dilengkapi dengan fitur pencarian dokter dan fasilitas kesehatan, sehingga memudahkan pengguna dalam mencari layanan kesehatan terdekat.

Selain itu, aplikasi ini juga memungkinkan pengguna untuk melakukan reservasi layanan kesehatan seperti konsultasi dokter, pemeriksaan laboratorium, dan pemeriksaan radiologi secara online. Aplikasi Satu Sehat *Mobile* tersedia pada platform Android dan IOS, dimana pada platform Android pengguna dapat memberikan penilaian berupa rating dan opini mengenai penggunaan Aplikasi Satu Sehat *Mobile* pada setiap versinya.

Dalam pengolahan data untuk analisa sentimen terkadang ditemukan kondisi data yang tidak

seimbang, dimana jumlah data cenderung lebih banyak pada kelas tertentu. Data yang tidak seimbang dapat menjadi masalah untuk analisa sentimen, karena dapat menyebabkan model analisa sentimen menjadi bias[5][6]. Model analisa sentimen yang bias dapat membuat kesalahan dalam memprediksi sentimen dari opini pengguna aplikasi[7], [8].

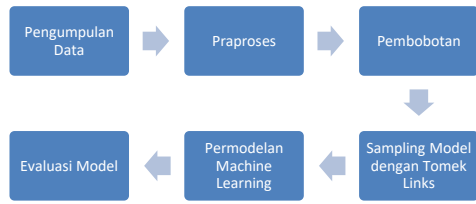
Salah satu solusi untuk menangani kondisi tersebut adalah melakukan metode sampling seperti *Tomek Links*. *Tomek Links* adalah teknik sampling *undersampling* yang dapat digunakan untuk mengatasi masalah data aplikasi yang tidak seimbang[9]. *Tomek Links* bekerja dengan menghilangkan pasangan data yang paling mirip antara dua kelas yang berbeda. *Tomek Links* dapat membantu untuk menyeimbangkan data aplikasi dan mengurangi bias pada model analisa sentimen.

Setelah data diseimbangkan dengan *Tomek Links*, dilanjutkan dengan melakukan membangun model dengan menggunakan algoritma *Random Forest*[10], *Neural Network*[11], *KNN*[12], dan *SVM*[13]. Kemudian untuk mengukur performa dari setiap algoritma dengan menggunakan *matrix evaluasi* yang meliputi akurasi, presisi, recall, dan *F1-Score*. Evaluasi ini membantu dalam memahami sejauh mana model-model tersebut dapat melakukan analisa sentimen opini[14], terutama opini pengguna Aplikasi Satu Sehat *Mobile* pada data yang tidak seimbang.

Melalui penelitian ini diharapkan mampu memberikan pemahaman dalam melakukan analisa sentimen pada penggunaan model sampling *Tomek Links* dalam menangani data yang tidak seimbang, dan bagaimana perbandingan kinerja algoritma *Random Forest*, *Neural Network*, *KNN*, dan *SVM* dalam melakukan analisa sentimen, terutama analisa sentimen opini pengguna Aplikasi Satu Sehat *Mobile*.

## METODOLOGI PENELITIAN

Penelitian ini bertujuan menerapkan model sampling *Tomek Links* pada analisa sentimen masyarakat pada aplikasi SATU SEHAT *Mobile* dan membandingkan kinerja model *Random Forest*, *Neural Network*, *KNN*, dan *SVM* seperti yang terlihat pada gambar 1.



Gambar 1. Metode Penelitian

**Data Penelitian**

Data yang digunakan sebagai penelitian adalah review pengguna aplikasi SATU SEHAT *Mobile* pada perangkat Android yang discrease melalui halaman *Play Store*. Pada tabel 1 ditunjukkan 15 versi aplikasi dengan jumlah review terbanyak.

Tabel 1. Versi dengan Jumlah Data Terbanyak

No	Versi	Jumlah
1	3.4.6	10614
2	4.0.2	5171
3	3.4.5	4557
4	5.2.0	3951
5	3.4.3	3722
6	5.0.0	3527
7	3.4.1	3478
8	3.2.6	2761
9	5.3.2	2016
10	3.4.0	1819
11	3.4.4	1692
12	5.2.1	1608
13	4.0.5	1606
14	5.3.1	1347
15	4.0.7	1326

Untuk memfokuskan penelitian, analisa sentimen akan menggunakan data dengan review terbanyak yang ada pada versi 3.4.6 dengan jumlah 10.614 review. Pada pemberian review, pengguna aplikasi memberikan review berupa kalimat tentang pengalaman penggunaan aplikasi dan nilai bintang dengan skala 1 sampai 5. Untuk memudahkan analisa sentimen skala bintang dikonversi menjadi 2 jenis opini dengan kondisi bintang 1-2 menjadi opini negatif, dan bintang

3-5 menjadi opini positif seperti terlihat pada tabel 2

Tabel 2. Perubahan Skala Bintang Menjadi Sentimen

Bintang	Jumlah	Opini	Jumlah Opini
1	6340	Negatif	8440
2	1160		
3	940		
4	566	Positif	2174
5	1608		

**Praproses Data**

Dilakukan praproses dataset dengan mengubah semua karakter menjadi huruf kecil, menghilangkan karakter non-alphabet (angka, karakter khusus), menghapus emoticon, melakukan penghapusan stopword (kata yang sering muncul dalam sebuah teks namun tidak memiliki makna khusus atau relevan dalam konteks analisis teks) dan melakukan stemming menggunakan model Sastrawi yang bertujuan mengubah kata dalam setiap kalimat menjadi kata dasar. Tabel 3 menunjukkan contoh kalimat review sebelum praproses dan setelah praproses.

**TF-IDF**

Selanjutnya dilakukan pembobotan yang bertujuan untuk mengukur pentingnya sebuah kata dalam teks yang dianalisis berdasarkan frekuensinya dan kemunculannya di seluruh dokumen, dengan menggunakan metode TF-IDF dapat membantu dalam mengekstraksi fitur penting yang dapat mempengaruhi sentimen keseluruhan dari teks tersebut[10].

TF dihitung dengan membagi frekuensi kemunculan sebuah kata dalam teks dengan jumlah kata total dalam teks yang ditampilkan dalam persamaan 1.

$$TF = \frac{\text{Jumlah kemunculan kata dalam teks}}{\text{Jumlah kata total dalam teks}} \dots\dots\dots(1)$$

Tabel 3. Contoh Praproses Teks

	Kalimat
Sebelum Praproses	secara keseluruhan aplikasi ini cukup mempermudah dalam penanganan covid hanya ada beberapa kekurangan yg pertama penggunaan data sedikit boros kedua izin layanan lokasi harus terus dinyalakan yg mana membuat

	baterai hp sedikit boros dan terakhir terkadang ketika login aplikasi keluar sendiri mohon kedepannya jika ada perbaikan masalah diatas dapat diperbaiki terimakasih
Setelah Praproses	aplikasi mudah tangan covid kurang yg guna data boros izin layan lokasi nyala yg baterai hp boros terkadang login aplikasi mohon depan baik atas baik terimakasih
Opini	Positif

IDF dihitung dengan menggunakan rumus logaritma dari jumlah total review dalam kumpulan data dibagi dengan jumlah review yang mengandung kata tersebut yang dapat dilihat pada persamaan 2.

$$IDF = \log\left(\frac{\text{Jumlah dokumen}}{\text{Jumlah dokumen yang berisi kata terkait}}\right) \quad (2)$$

Persamaan 3 menampilkan persamaan TF-IDF dimana nilai TF-IDF diperoleh dengan mengalikan nilai TF dengan IDF.

$$TF-IDF = TF * IDF \dots\dots\dots(3)$$

Tabel 4. Kata dengan TF-IDF tertinggi

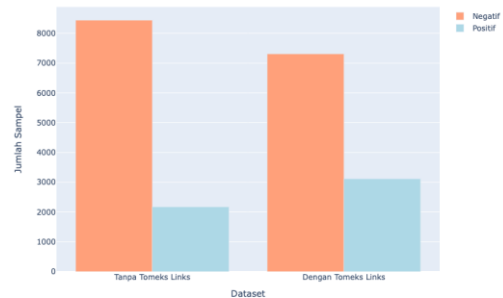
No	Kata	Score TF-IDF
1	boros	0.444978
2	tangan	0.253258
3	terkadang	0.250663
4	atas	0.243834
5	nyala	0.243834
6	kurang	0.235524
7	baik	0.233225
8	depan	0.228766
9	baterai	0.222489
10	izin	0.216063

Dimana setelah pembobotan menggunakan TF-IDF dilakukan, setiap kalimat direpresentasikan sebagai vektor dan tabel 4 menunjukkan 10 kata dengan score TF-IDF tertinggi.

**Model Sampling Tomek Links**

Sampling model adalah proses pengambilan sampel yang representatif dari populasi yang lebih besar. Sampel ini digunakan untuk menganalisis dan membuat prediksi terkait dengan populasi tersebut[15]. Dalam konteks machine learning, model sampling sering digunakan untuk menangani ketidakseimbangan kelas, di mana jumlah sampel dalam kelas yang berbeda tidak seimbang. Model sampling dapat digunakan untuk menghasilkan sampel yang seimbang atau mengurangi jumlah sampel dalam kelas mayoritas agar sebanding dengan jumlah sampel dalam kelas minoritas. Dimana pada penelitian ini terjadi ketidakseimbangan data pada kelas negatif dan positif yang memiliki persentase ratio data 79,52 : 20,48.

Perbedaan Data Sebelum dan Setelah Tomek Links

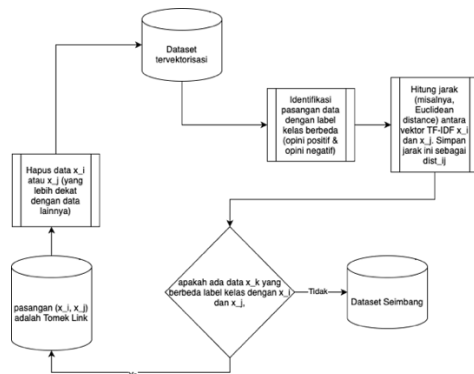


Gambar 2. Perbedaan Jumlah Data Sebelum dan Setelah Penggunaan Tomek Links

Model sampling yang diterapkan pada penelitian ini adalah *Tomek Links*, Metode Sampling *Tomek Links* adalah salah satu teknik undersampling yang digunakan dalam klasifikasi data yang tidak seimbang[9]. Tujuan utama metode ini adalah untuk mengurangi jumlah sampel mayoritas yang berdekatan dengan sampel-sampel minoritas di sekitarnya. Alur dari *Tomek Links* bisa dilihat pada gambar 3.

Dampak penerapan metode Sampling *Tomek Links* pada data review. Data awal terdiri dari 8440 sampel label negatif dan 2174 sampel dengan label positif. Setelah menerapkan metode sampling *Tomek Links*, terjadi perubahan distribusi data menjadi 7310 sampel dengan label negatif dan 3114 sampel dengan label positif yang terlihat pada gambar 2.

Penerapan metode *Sampling Tomek Links* menunjukkan adanya perubahan yang signifikan dalam distribusi kelas setelah penerapan metode *Sampling Tomek Links*. Jumlah sampel dengan label negatif mengalami pengurangan sebesar 1130, sementara jumlah sampel dengan label positif mengalami peningkatan sebesar 940.



Gambar 3. Alur Tomek Links

Perubahan ini mencerminkan proses undersampling yang terjadi, di mana sampel mayoritas yang berdekatan dengan sampel-sampel minoritas yang terisolasi dihapus untuk mencapai keseimbangan kelas yang lebih baik. Dengan mengurangi jumlah sampel mayoritas, diharapkan dapat meningkatkan kemampuan model klasifikasi untuk mengenali dan memprediksi dengan lebih baik sampel-sampel minoritas yang penting. Dalam penelitian ini, perubahan distribusi data setelah menggunakan metode *Sampling Tomek Links* dapat dianggap sebagai strategi untuk mengatasi masalah ketidakseimbangan kelas.

### Pembangunan Model

Pada tahap awal pembangunan model dilakukan pembagian data menjadi data latih dan data uji dengan ratio data 80:20 pada data review asli dan data review yang telah menerapkan model *Sampling Tomek Links* yang dapat dilihat pada tabel 5.

Tabel 5. Pembagian Data

	Data latih (80%)	Data uji (20%)
Tanpa model sampling	8491	2123
Menerapkan model sampling	8339	2085

Pada tahap pembangunan model menggunakan algoritma klasifikasi untuk menguji penggunaan model *Sampling Tomek Links* terhadap analisa sentimen. Penelitian ini menggunakan 4 algoritma *Random Forest*, *Neural Network*, *KNN*, dan *SVM*. Dimana setiap algoritma akan dibangun dan dibandingkan menggunakan matrix evaluasi dan waktu proses, sehingga menunjukkan bagaimana perbedaan model yang menggunakan model *Sampling* dan tidak menggunakan model *Sampling*.

### HASIL DAN PEMBAHASAN

Untuk mengetahui kinerja dari model yang dibangun, digunakan metrik Akurasi, *Precision*, *Recall*, dan *F1-Score* untuk mengetahui model mana yang paling baik, dan dihitung juga waktu proses untuk mengetahui durasi waktu yang dibutuhkan setiap model dalam melakukan pengujian. Hasil dari analisa model dapat dilihat dalam matrik evaluasi pada tabel 6.

Pada skenario data tanpa menggunakan model *Sampling*, penggunaan algoritma *SVM* menunjukkan performa yang baik, ditunjukkan dengan nilai akurasi sebesar 80,528% yang merupakan nilai akurasi tertinggi pada skenario tanpa model *Sampling*. Akurasi sendiri adalah rasio prediksi yang benar terhadap total jumlah data. Ini dapat terjadi karena *SVM* berfokus pada mengoptimalkan margin dan tidak secara langsung mengakibatkan penanganan yang lebih baik terhadap ketidakseimbangan kelas. Namun *Random Forest* memiliki nilai tertinggi untuk *Precision*, *Recall*, dan *F1-Score* karena cenderung lebih baik dalam mengidentifikasi dan mengklasifikasikan instance dari kelas minoritas secara akurat, bahkan jika akurasinya sedikit lebih rendah.

Skenario data yang menerapkan model *Sampling* terjadi peningkatan nilai akurasi pada semua model yang dibangun, pada model

Tabel 6. Matrik Evaluasi

Algoritma	Akurasi	Precision	Recall	F1-Score	Waktu Proses
Menggunakan Model Sampling					
<i>Random Forest</i>	80,405%	81,004%	80,405%	78,243%	6,7536
<i>Neural Network</i>	74,706%	74,162%	74,706%	74,384%	0,0018
K-NN	76,401%	75,256%	76,401%	75,321%	0,4515
<i>SVM</i>	81,017%	81,951%	81,017%	78,872%	0,9022
Tidak Menggunakan Model Sampling					
<i>Random Forest</i>	79,760%	79,610%	79,760%	77,991%	9,727
<i>Neural Network</i>	74,484%	74,027%	74,484%	74,221%	0,002
K-NN	74,916%	74,385%	74,916%	74,599%	0,466
<i>SVM</i>	80,528%	81,306%	80,528%	78,394%	0,871

menggunakan algoritma *SVM* terjadi peningkatan akurasi sebesar 0,489%, dan menjadi nilai akurasi tertinggi dibandingkan model lain menggunakan model sampling. Peningkatan pada semua matriks evaluasi setelah menerapkan metode sampling *Tomek Links* bisa terjadi karena efek positif dari metode tersebut dalam mengatasi masalah ketidakseimbangan kelas pada dataset. Sampling *Tomek Links* adalah salah satu teknik yang digunakan untuk memperbaiki ketidakseimbangan kelas dengan cara mengurangi overlap antara kelas mayoritas dan kelas minoritas. Ini dapat berdampak positif pada kinerja model klasifikasi, terutama pada dataset yang memiliki masalah ketidakseimbangan yang signifikan.

Dari sisi waktu proses terjadi pengurangan waktu proses antara model yang menggunakan model sampling dibandingkan dengan model yang tidak menggunakan model sampling. Pengurangan waktu proses terjadi karena dalam penggunaan metode sampling seperti *Tomek Links* dapat mengurangi ukuran keseluruhan dataset dengan menghapus beberapa instance. Dengan ukuran dataset yang lebih kecil, algoritma yang digunakan dapat bekerja lebih cepat karena harus memproses lebih sedikit data.

## SIMPULAN

Hasil penelitian menunjukkan bahwa pada skenario data tanpa menggunakan model sampling, algoritma *SVM* memiliki kinerja yang baik dengan akurasi tertinggi sebesar 80,528%. Meskipun akurasi yang tinggi, *SVM* tidak secara khusus menangani masalah ketidakseimbangan kelas, yang mengakibatkan kelas minoritas tidak diidentifikasi dengan baik. Di sisi lain, model *Random Forest* memiliki nilai tertinggi untuk Precision, Recall, dan F1-Score, menunjukkan kemampuannya dalam mengklasifikasikan instance kelas minoritas secara akurat, bahkan dengan akurasi sedikit lebih rendah.

Namun, dengan menerapkan metode sampling, terjadi peningkatan signifikan dalam kinerja semua model yang dibangun. Penggunaan metode sampling *Tomek Links* secara khusus meningkatkan akurasi pada model *SVM* sebesar 0,489%, menjadikannya sebagai model dengan akurasi tertinggi dalam skenario data dengan model sampling. Hal ini mengindikasikan bahwa penggunaan metode sampling dapat efektif mengatasi masalah ketidakseimbangan kelas pada dataset review, menghasilkan peningkatan dalam semua matriks evaluasi.

**DAFTAR PUSTAKA**

- [1] K. Dharmendra, K. O. Saputra, and I. N. Pramaita, 'Analisa Sentiment Untuk Opini Alumni Perguruan Tinggi', *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 2, pp. xxxx-xxxx, Jul. 2019, doi: 10.24843/mite.2019.v18i02.p11.
- [2] 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.
- [3] 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.
- [4] 'SATUSEHAT Mobile - Apps on Google Play'. Accessed: Apr. 29, 2023. [Online]. Available: [https://play.google.com/store/apps/details?id=com.telkom.tracencare&hl=en\\_US](https://play.google.com/store/apps/details?id=com.telkom.tracencare&hl=en_US)
- [5] H. Utami, 'Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network', *Indones. J. Appl. Stat.*, vol. 5, no. 1, Art. no. 1, May 2022, doi: 10.13057/ijas.v5i1.56825.
- [6] S. M. Chamzah, M. Lestandy, N. Kasan, and A. Nugraha, 'Penerapan Synthetic Minority Oversampling Technique (SMOTE) untuk Imbalance Class pada Data Text Menggunakan kNN', *Syntax J. Inform.*, vol. 11, no. 02, Art. no. 02, Nov. 2022, doi: 10.35706/syji.v11i02.6940.
- [7] C. Padurariu and M. E. Breaban, 'Dealing with Data Imbalance in Text Classification', *Procedia Comput. Sci.*, vol. 159, pp. 736-745, Jan. 2019, doi: 10.1016/j.procs.2019.09.229.
- [8] F. Abdulloh, A. Aminuddin, M. Rahardi, and S. Anggita, 'Observation of Imbalance Tracer Study Data for Graduates Employability Prediction in Indonesia', *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, pp. 169-174, Sep. 2022, doi: 10.14569/IJACSA.2022.0130820.
- [9] E. F. Swana, W. Doorsamy, and P. Bokoro, 'Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset', *Sensors*, vol. 22, no. 9, Art. no. 9, Jan. 2022, doi: 10.3390/s22093246.
- [10] 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.
- [11] A. Onan, 'Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks', *Concurr. Comput. Pract. Exp.*, vol. 33, no. 23, p. e5909, 2021, doi: 10.1002/cpe.5909.
- [12] F. M. J. Mehedi Shamrat *et al.*, 'Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm', *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 1, p. 463, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp463-470.
- [13] 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.
- [14] I. K. Dharmendra, I. G. N. A. Kusuma, I. A. M. C. Dewi, and Edwar, 'IMPLEMENTASI TEXT MINING UNTUK KLASIFIKASI OPINI ALUMNI PADA PERGURUAN TINGGI', *J. Teknol. Inf. Dan Komput.*, vol. 9, no. 3, Art. no. 3, Jul. 2023, Accessed: Aug. 18, 2023. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jutik/article/view/2504>

- [15] E. Rendón, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutiérrez, 'Data Sampling Methods to Deal With the Big Data Multi-Class Imbalance Problem', *Appl. Sci.*, vol. 10, no. 4, Art. no. 4, Jan. 2020, doi: 10.3390/app10041276.