JURNAL FASILKOM
Volume 16 No. 2 | Agustus 2026: 264-272

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Perbandingan Algoritma *Support Vector Machine* dan *Random Forest* untuk Analisis Sentimen terkait Makan Bergizi Gratis (MBG)

Alfian Saputra[1], Nuralamsah Zulkarnaim[2], Farid Wajidi[3]
[1,2,3]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Sulawesi Barat
[1]alfiansaputra355@gmail.com, [2]Nuralamsah@unsulbar.ac.id*, [3]Faridwajidi@unsulbar.ac.id*

Abstract

Advances in digital technology have made social media the primary platform for the public to voice their opinions on government policies, including the Free Nutritious Meals (MBG) program. This study aims to compare the performance of the Support Vector Machine (SVM) and Random Forest algorithms in conducting sentiment analysis of public opinion on the X platform regarding this policy, in order to provide an objective overview for policymakers. A total of 2,164 tweets were collected via crawling using Tweet Harvest from May to August 2025. The methodological steps included preprocessing, BERT-based automatic labeling, Eliminate the neutral class to focus the analysis on binary classification (positive and negative), TF-IDF feature extraction, and the application of SMOTE to address class imbalance in the dataset. Model optimization was performed using Grid Search with a 5-Fold Cross Validation testing scheme and an 80:20 data split. The results of the study indicate that the majority of public responses to the MBG program were positive. Based on the final evaluation, the SVM with a linear kernel proved superior with an accuracy of 78.24% and a macro average F1-score of 0.76, outperforming the Random Forest (300 estimators), which achieved an accuracy of 75.00% and an F1-score of 0.73. The application of SMOTE has proven to be crucial in improving the negative class recall, enabling the model to identify minority sentiments more accurately. This study concludes that SVMs demonstrate more stable and objective generalization capabilities in mapping public opinion following data balancing.

Keywords: sentiment analysis, free nutritious meal, SVM, random forest, SMOTE

Abstrak

Kemajuan teknologi digital telah menjadikan media sosial sebagai wadah utama bagi masyarakat untuk menyuarakan aspirasi terhadap kebijakan pemerintah, termasuk program Makan Bergizi Gratis (MBG). Penelitian ini bertujuan untuk membandingkan performa algoritma *Support Vector Machine* (SVM) dan *Random Forest* dalam melakukan analisis sentimen terhadap opini publik di *platform* X terkait kebijakan tersebut guna memberikan gambaran objektif bagi pemangku kebijakan. Data sebanyak 2.164 *tweet* dikumpulkan melalui teknik *crawling* menggunakan *Tweet Harvest* pada periode Mei hingga Agustus 2025. Tahapan metodologi mencakup *preprocessing*, pelabelan otomatis berbasis BERT, eliminasi kelas netral untuk memfokuskan analisis pada klasifikasi biner (positif dan negatif), ekstraksi fitur TF-IDF, serta penerapan *SMOTE* untuk mengatasi ketidakseimbangan kelas pada *dataset*. Optimasi model dilakukan menggunakan *Grid Search* dengan skema pengujian 5-*Fold Cross Validation* dan pembagian data 80:20. Hasil penelitian menunjukkan bahwa mayoritas respon publik terhadap program MBG bersifat positif. Berdasarkan evaluasi akhir, SVM dengan kernel linear terbukti lebih unggul dengan akurasi 78,24% dan *macro average F1-score* 0,76, mengungguli *Random Forest* (300 *estimators*) yang mencapai akurasi 75,00% dan *F1-Score* 0,73. Penerapan *SMOTE* terbukti krusial dalam meningkatkan nilai *recall* kelas negatif, sehingga model mampu mengenali sentimen minoritas dengan lebih akurat. Penelitian ini menyimpulkan bahwa SVM memiliki kemampuan generalisasi yang lebih stabil dan objektif dalam memetakan opini publik setelah dilakukan penyeimbangan data.

Kata kunci: analisis sentimen, makan bergizi gratis, SVM, *random forest*, *SMOTE*

1. Pendahuluan

Kemajuan teknologi digital telah mentransformasi cara masyarakat berinteraksi, khususnya dalam merespons berbagai isu *public*, media sosial kini tidak hanya berfungsi sebagai saluran informasi, tetapi juga menjadi wadah bagi warga untuk menyuarakan aspirasi terhadap kebijakan pemerintah, hal ini memungkinkan persepsi kolektif masyarakat terhadap suatu fenomena dapat terpantau secara langsung melalui *platform* digital tersebut [1]. Data dari *We Are Social* pada Januari 2024 menunjukkan bahwa penetrasi media sosial di Indonesia telah mencapai 167 juta pengguna aktif, atau mencakup sekitar 60,4% dari seluruh populasi.Dalam ekosistem digital tersebut, masyarakat Indonesia tercatat paling aktif berinteraksi melalui *platform WhatsApp, Instagram, Facebook, TikTok, Telegram, dan X (Twitter)* [2]. Walaupun sejumlah literatur terdahulu melaporkan angka penetrasi pengguna X yang signifikan, data terkini dari *We Are Social* per Januari 2024 menunjukkan realitas yang berbeda. Tercatat bahwa pengguna X di Indonesia berjumlah 24,69 juta jiwa, yang mana angka tersebut merepresentasikan 8,9% dari keseluruhan populasi nasional[3]. Dengan begitu, Penelitian ini hanya fokus pada media sosial *platform X(Twitter).*

Kebijakan pemberian makanan bergizi gratis telah diterapkan di berbagai negara termasuk di Indonesia, program Makan Bergizi Gratis (MBG) diterapkan pada januari 2025 dibawah pemerintahan presiden Prabowo

P-ISSN : 2089-3353
E-ISSN : 2808-9162

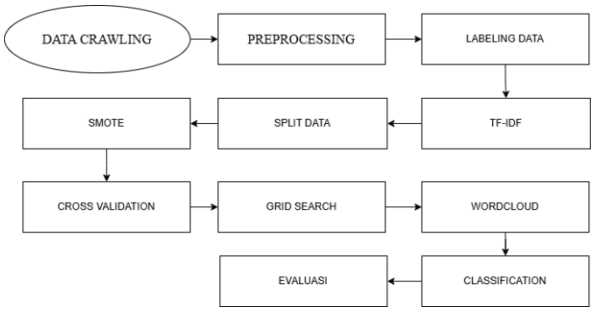
Subianto, program ini berfokus pada pemenuhan gizi pelajar dan ibu hamil dan program ini menargetkan 82,9 juta penerima manfaat dalam kurun waktu hingga tahun 2029 [4]. Pemerintah telah menganggarkan dana sebesar 71 triliun untuk program ini pada tahun 2025 [5]. Dalam implementasinya, kebijakan ini banyak menuai pro dan kontra salah satu isu yang beredar di media berita bahwa kebijakan ini memangkas anggaran Pendidikan hampir sepertiga bagian untuk dialihkan ke program makanan bergizi gratis, Kebijakan ini dinilai melanggar aturan negara karena jaminan 20% APBN untuk sektor pendidikan merupakan kewajiban konstitusional yang tidak boleh dikurangi [6]. Analisis sentimen merupakan metode pengolahan teks yang bertujuan untuk mendeteksi serta mengklasifikasikan polaritas emosi seperti positif, negatif, atau netral yang terdapat dalam suatu data tekstual, melalui pendekatan teknik analisis teks, metode ini secara luas diimplementasikan untuk mengevaluasi persepsi, opini, maupun aspirasi masyarakat terhadap berbagai objek, mulai dari komoditas produk dan layanan hingga kebijakan atau gagasan tertentu[7].

Dari penelitan sebelumnya yang membandingkan metode *Support Vector Machine* (SVM) dan *random forest* terkait sentimen PSBB, hasil yang diperoleh menunjukkan bahwa meskipun algoritma *Random Forest* memiliki tingkat akurasi yang lebih tinggi. *Support Vector Machine* (SVM) dianggap lebih unggul dalam penelitian ini karena mampu mengidentifikasi sentimen berlabel "Positif" yang gagal dideteksi oleh model *Random Forest* [8]. Adapun penelitan lainnya yang membandingan kedua algoritma *Support Vector Machine* (SVM) dan *Random Forest* menunjukkan bahwa implementasi teknik *SMOTE* secara signifikan meningkatkan efektivitas klasifikasi sentimen publik mengenai *metaverse*, dengan perolehan akurasi mencapai 91% pada algoritma *Random Forest* dan 90% pada *Support Vector Machine.* Meskipun *Random Forest* menunjukkan performa akurasi yang lebih unggul, penerapan *SMOTE* terbukti krusial dalam mempertajam identifikasi sentimen positif pada kedua model [9]. Pada penelitian lainnya menyimpulkan bahwa implementasi algoritma *Support Vector Machine* (SVM) dengan ekstraksi fitur TF-IDF efektif dalam memetakan sentimen publik terhadap kebijakan penanganan pandemi, seperti *lockdown* dan PSBB. Berdasarkan distribusi data yang didominasi oleh sentimen positif (68,75%), model ini berhasil mencapai tingkat akurasi sebesar 74%, dengan nilai *precision* 75%, *recall* 92%, dan *F1-Score* sebesar 83% [10]. Kemudian penelitian selanjutnya, kombinasi algoritma naïve bayes dengan pembobotan TF-IDF mendapatkan hasil yang baik dengan akurasi 84% yang menandakan bahwa algoritma *Naïve Bayes* cukup andal dalam klasifikasi teks pendek dan bersifat informal [1]. Adapun penelitan lainnya yang membandingan kedua algoritma *Naïve Bayes* dan KNN hasil akurasi menunjukkan bahwa algoritma *Naïve Bayes* lebih ungggul dengan akurasi 79,61% sedangkan algoritma KNN mendapat akurasi 73,30% ini menunjukkan bahwa algortima *Naïve Bayes* lebih handal dalam analisis sentimen teks berbahasa Indonesia dengan *dataset* yang besar [11].

Meskipun penelitian mengenai analisis sentimen kebijakan pemerintah menggunakan SVM atau *Random Forest* sudah banyak dilakukan, penelitian ini memiliki perbedaan mendasar dalam aspek metodologi. Berbeda dengan penelitian terdahulu yang sering mengabaikan masalah ketidakseimbangan kelas dan menggunakan parameter standar, penelitian ini mengintegrasikan *SMOTE* untuk penanganan *class imbalance* serta menerapkan *Grid Search* untuk optimasi hyperparameter pada kedua algoritma. Pendekatan ini memastikan komparasi performa antara SVM dan *Random Forest* dilakukan pada kondisi model yang paling optimal dan seimbang, sehingga menghasilkan akurasi yang lebih valid.

2. Metode Penelitian

Alur penelitian ini diinisiasi melalui proses ekstraksi data secara langsung dari media sosial X dengan memanfaatkan instrumen *tweet harvest*. Data yang diperoleh kemudian masuk ke dalam fase pra-pemrosesan, dimana pembersihan teks dieksekusi melalui beberapa rangkain prosedur yang meliputi *cleansing*, *case folding*, serta normalisasi. Selanjutnya , dilakukan pula tahapan *tokenization*, eliminasi *stopword*, hingga proses *stemming*. Pada tahap labeling data, setiap tweet di kelompokkan ke dalam kategori sentimen: positif dan negatif. Selanjutnya dilakukan ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk mentransformasi data teks menjadi representasi vektor numerik. Berikutnya melakukan pembagian data, data latih dan data uji. Tahap berikutnya dikalukan klasifikasi menggunakan algortima *Support Vector Machine (SVM) dan Random Forest.* Rangkaian penelitian ini ditutup dengan tahap evaluasi, di mana performa kedua model diukur dan dibandingkan dengan hasil yang terbaik



Gambar 1. Alur Penelitian

2.1. *Crawling* data

Proses pengumpulan data dilakukan melalui teknik *crawling* dengan bantuan Tool *Tweet Harvest*. Tahapan ini menjadi bagian fundamental dalam analisis data, di mana peneliti memanfaatkan *Application Programming Interface* (API) resmi Twitter untuk

P-ISSN : 2089-3353
E-ISSN : 2808-9162

mengakses serta mengekstraksi data teks yang diunggah oleh pengguna secara sistematis [12]. Pada tahap awal, pengumpulan data menggunakan kata kunci "MBG" karena intensitas kemunculannya yang tinggi. Pengambilan data dilakukan selama periode 20 Mei hingga 30 Agustus 2025, dengan total perolehan sebanyak 2.164 *tweet* mentah. *Dataset* tersebut kemudian disimpan dalam format CSV untuk kemudian masuk ke tahap pemrosesan (*processing*).

2.2. *Preprocessing*

Preprocessing merupakan tahapan krusial untuk menormalisasi istilah dalam teks guna menghasilkan data latih yang berkualitas dan memastikan fitur yang diekstraksi relevan dengan tujuan penelitian. Proses ini bertujuan menyederhanakan kompleksitas data, mengingat opini di media sosial Twitter sering kali mengandung kata tidak baku, istilah di luar kamus, maupun penggunaan bahasa daerah. Melalui *preprocessing* atau normalisasi, ekspresi atipikal dieliminasi untuk mengembalikan teks ke bentuk naturalnya, sehingga mampu meminimalkan *noise* pada tahap analisis selanjutnya [13]. Sebelum masuk ke tahap *processing* data *crawling* yang telah di ambil dilakukan pembersihan data duplikat terlebih dahulu agar memudahkan dalam proses klasifikasi. Berikut beberapa proses pada tahap *processing*; Pertama *Cleansing,* Tahap ini bertujuan untuk mengeliminasi seluruh karakter non-tekstual yang tidak relevan, meliputi simbol, tautan URL, emotikon, serta karakter numerik. Proses pembersihan ini dilakukan guna memastikan data yang akan diolah hanya terdiri dari karakter alfabet yang diperlukan untuk analisis [14]. Kedua *Case Folding*, yang bertujuan untuk menyeragamkan seluruh karakter teks dalam *dataset* menjadi huruf kecil (*lowercase*) [14]. Ketiga Normalisasi, Tahapan ini bertujuan menyelaraksan kosakata nonformal maupun Bahasa slang agar merujuk pada ketentuan ejaan resmi EYD edisi V serta Kamus Besar Bahasa Indonesia (KBBI). Hal ini krusial untuk menjaga konsistensi semantik dalam data [15]. Keempat tokenisasi, adalah prosedur segmentasi teks yang memecah rangkaian kalimat atau dokumen menjadi unit-unit atomik yang lebih kecil, yang dikenal sebagai token, Tahapan ini berfungsi untuk mempermudah analisis struktural pada setiap kata atau elemen bahasa secara individu [9]. Kelima *Stopword Removal*, yaitu menyeleksi dan mengeliminasi kosakata yang dianggap minim bobot makna dalam pemrosesan teks. Kata-kata fungsional seperti konjungsi, preposisi, dan pronomina dihapus guna memfokuskan pemrosesan pada kata-kata kunci yang mengandung substansi makna [15]. Terakhir tahap *Stemming*, yaitu mentransformasi kata-kata berimbuhan menjadi bentuk dasarnya guna meningkatkan konsistensi dalam proses analisis. Prosedur ini dilakukan dengan mengeliminasi prefiks, infiks, maupun sufiks, sehingga kata seperti 'berlari' akan dikonversi menjadi kata dasar 'lari' [11].

2.3. *Labeling* data

Proses pelabelan bertujuan untuk memilah data tekstual ke dalam kelas sentimen tertentu. Melalui analisis kontekstual pada setiap baris *dataset*, data diberikan label positif, netral, atau negatif guna menentukan polaritas opini yang direpresentasikan oleh pengguna [16]. Penelitian ini menggunakan metode pelabelan *Bidirectional Encoder Representations from Transformers* (BERT) dengan nama taufiqdp/indonesian-*sentiment*. Model ini disempurnakan untuk melakukan analisis sentimen pada komentar dan ulasan berbahasa Indonesia.

2.4. *Split* data

Split data merupakan prosedur pembagian *dataset* ke dalam beberapa sub-himpunan (*subset*) yang terpisah. Tahapan ini dilakukan untuk memisahkan data yang akan digunakan dalam proses pembangunan model dengan data yang akan digunakan untuk menguji performa model tersebut [17]. Dalam prosedur pembagian data, satu subset dialokasikan sebagai data latih (*training set*) untuk membangun model, sementara subset lainnya difungsikan sebagai data uji (*testing set*) untuk mengevaluasi performa model yang telah dihasilkan [17]. *Split* data terbagi beberapa 90:10, 80:20, 70:30, dan 60:40. Pemilihan split data ditetapkan sesuai keinginan pengguna. Pada penelitian ini menggunakan pembagian 80:20.

2.5. TF-IDF

Untuk mengubah data teks ke dalam bentuk numerik yang dapat diproses oleh algoritma, penelitian ini menerapkan pembobotan TF-IDF dalam tahap vektorisasi. Langkah ini berfungsi untuk menetapkan nilai bobot pada setiap fitur berdasarkan frekuensi kemunculannya, sehingga signifikansi tiap kata dalam *dataset* dapat terukur secara objektif [18]. Hasil akhir berupa perkalian antara TF dan IDF.

2.6. *SMOTE*

Metode *SMOTE* (*Synthetic Minority Over-Sampling Technique*) diintegrasikan sebagai teknik resampling guna menyelaraskan proporsi data antar-kelas yang tidak seimbang. Metode ini bekerja dengan cara mengidentifikasi sampel pada kelas minoritas dan membangkitkan data sintetis baru melalui teknik interpolasi di antara sampel-sampel tersebut, sehingga jumlah data pada kelas minoritas meningkat secara proporsional[19].

2.7. *Grid Search*

Grid search merupakan salah satu teknik optimasi yang bertujuan untuk menentukan kombinasi parameter paling ideal bagi sebuah model pembelajaran mesin. Algoritma ini bekerja secara sistematis dengan menguji seluruh kemungkinan kombinasi dari nilai parameter yang telah ditentukan sebelumnya oleh peneliti. Setiap kombinasi tersebut disimpan dalam sebuah matriks (*grid*) untuk dievaluasi, di mana parameter terbaik

P-ISSN : 2089-3353
E-ISSN : 2808-9162

dipilih berdasarkan nilai tingkat kesalahan (*error rate*) terkecil atau skor performa tertinggi [20].

Tabel 1. Konfigurasi Parameter *Grid Search* Untuk SVM Dan *Random Forest*

Algoritma	Parameter	Nilai Yang Diuji
SVM	Kernel	Linear (terpilih), rbf, poly, sigmoid
	Nilai C	0.1, 1 (terpilih), 10
	Gamma	*scale* (terpilih), *auto*
Random Forest	n_*estimators*	100, 200, 300 (terpilih)
	max_depth	*none* (terpilih), 10, 20

Tabel 1 menunjukkan bahwa SVM mencapai performa paling optimal menggunakan *kernel linear*, yang artinya fitur data sentimen dapat dipisahkan secara linear pada dimensi tinggi. Sedangkan pada *Random Forest*, performa terbaik diperoleh pada nilai n_*estimators* = 300, memberikan stabilitas prediksi tanpa memicu *overfitting*.

2.8. *Wordcloud*

Wordcloud adalah teknik visualisasi dalam *text mining* yang digunakan untuk merepresentasikan intensitas kemunculan kata dalam suatu kumpulan data teks. Representasi ini mengandalkan tipografi untuk menunjukkan bobot data; kata dengan ukuran huruf yang lebih besar menandakan frekuensi kemunculan yang tinggi, sementara ukuran huruf yang lebih kecil mencerminkan frekuensi penggunaan kata yang lebih rendah dalam dokumen tersebut [21].

2.9. *Classification*

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang efektif untuk tugas klasifikasi maupun regresi. Dalam implementasinya, SVM memanfaatkan fungsi *kernel* yang terbagi menjadi kategori *linear* dan non-*linear*. Pemisahan dua kategori data pada ruang fitur berdimensi tinggi dilakukan oleh kernel *linear* SVM dengan cara mengonstruksi garis pembatas (*hyperplane*) terbaik, khususnya bagi data yang bersifat *linearly separable*. Sebaliknya, SVM non-*linear* diterapkan pada distribusi data yang lebih kompleks dengan mengintegrasikan pendekatan fungsi *kernel* guna memetakan fitur awal ke ruang dimensi yang lebih tinggi agar dapat dipisahkan secara akurat [22]. Pada penelitian ini diuji menggunakan beberapa jenis kernel dalam pengujian, yaitu linear, *Radial Basis Function* (RBF), *Polynomial* dan *Sigmoid*. Pemilihan kernel terbaik dilakukan secara otomatis melalui proses *Grid Search* untuk mendapatkan model yang paling adaptif terhadap karakteristik data sentimen Makan Bergizi Gratis

Random Forest dikategorikan sebagai algoritma *machine learning* berbsasis *ensemble* mengintegrasikan beberapa teknik pembelajaran mesin menjadi satu model prediktif yang kuat. Mekanisme kerja algoritma ini melibatkan pembangunan sejumlah besar pohon keputusan (*decision trees*) secara simultan. Hasil prediksi akhir ditentukan melalui proses pengambilan suara terbanyak (*majority voting*) dari seluruh prediksi yang dihasilkan oleh pohon-pohon tersebut, sehingga mampu meningkatkan akurasi dan stabilitas model [23]. Dalam penelitian ini dioptimasi menggunakan teknik *Grid Search* dengan menguji variasi banyaknya pohon keputusan (n_*estimators*) dan Batasan kedalaman pohon secara maksimal *(max_depth)*. Hal ini bertujuan untuk menemukan keseimbangan optimal antara akurasi model dan pencegahan *overfitting* pada data sentimen program Makan Bergizi Gratis.

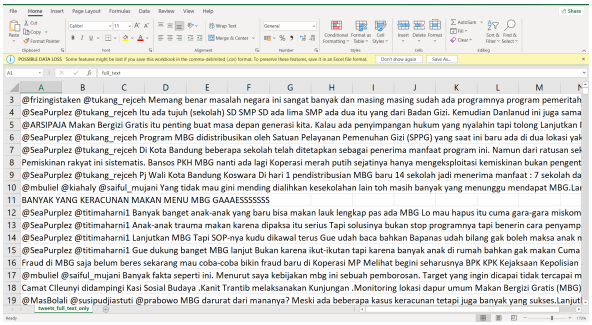
2.10. Evaluasi

Metode *Confusion Matrix* diterapkan untuk mengukur kinerja dan reliabilitas proses klasifikasi. Pengujian dilakukan dengan menghitung nilai *accuracy* untuk melihat tingkat kebenaran umum, serta nilai *precision* dan *recall* untuk mengevaluasi kualitas prediksi pada setiap kelas sentimen secara lebih spesifik [24].

3. Hasil dan Pembahasan

3.1. *Crawling* data

Proses penarikan data dalam penelitian ini dilakukan melalui *Google Colab* menggunakan tool *tweet harvest* berbasis Bahasa pemrograman *python*. Dengan menggunakan kata kunci "MBG" Selama periode 20 Mei hingga 30 Agustus 2025, didapatkan data *tweet* mentah sebanyak 2.164 tweet. Kemudian, hasil *crawling* tersebut disimpan ke dalam file format CSV. *Dataset* komprehensif dapat ditinjau melalui tautan https://bit.ly/4sLmyoX serta disajikan secara visual pada gambar 2 dibawah.



Gambar 2. *Dataset*

3.2. *Preprocessing*

Tahap pertama melakukan proses penghapusan data ganda untuk menjamin setiap *tweet* bersifat unik. Langkah ini bertujuan agar data yang diolah merupakan data yang unik dan tidak memengaruhi hasil perhitungan sentimen. Sebelum proses penghapusan duplikasi, jumlah data awal yang tersedia adalah sebanyak 2164 *tweet,* dan sesudah dilakukan proses penghapusan tersebut, jumlah data akhir menjadi 1972 *tweet.*

Tahap kedua *cleansing* atau pembersihan teks bertujuan untuk mengeliminasi seluruh karakter non-

P-ISSN : 2089-3353
E-ISSN : 2808-9162

tekstual yang tidak relevan, meliputi simbol, tautan URL, emotikon, serta karakter numerik

Tabel 2. Contoh Hasil *Cleaning*

Sebelum	Sesudah
@mbuliel @saiful_mujani Banyak fakta seperti ini. Menurut saya kebijakan mbg ini sebuah pemborosan. Target yang ingin dicapai tidak tercapai malahan yang terjadi pemborosan karena makanan banyak tidak habis dan akhirnya dibuang. Coba kebijakan mbg ditujukan kpd murid yg benar² membutuhkan lbh efektif	Banyak fakta seperti ini Menurut saya kebijakan mbg ini sebuah pemborosan Target yang ingin dicapai tidak tercapai malahan yang terjadi pemborosan karena makanan banyak tidak habis dan akhirnya dibuang Coba kebijakan mbg ditujukan kpd murid yg benar² membutuhkan lbh efektif

Tahap ketiga *case folding* yaitu menyeragamkan semua teks yang awalnya huruf besar menjadi huruf kecil (*Lowering Case*).

Tabel 3. Contoh Hasil *Lowering Case*

Sebelum	Sesudah
Banyak fakta seperti ini Menurut saya kebijakan mbg ini sebuah pemborosan Target yang ingin dicapai tidak tercapai malahan yang terjadi pemborosan karena makanan banyak tidak habis dan akhirnya dibuang Coba kebijakan mbg ditujukan kpd murid yg benar² membutuhkan lbh efektif	banyak fakta seperti ini menurut saya kebijakan mbg ini sebuah pemborosan target yang ingin dicapai tidak tercapai malahan yang terjadi pemborosan karena makanan banyak tidak habis dan akhirnya dibuang coba kebijakan mbg ditujukan kpd murid yg benar² membutuhkan lbh efektif

Tahap keempat normasisasi yang bertujuan mengubah kata yang tidak baku menjadi ke kata baku sesuai Pedoman Umum Ejaan Bahasa Indonesia (PUEBI) atau Kamus Besar Bahasa Indonesia (KBBI). Pada penelitian ini menggunakan kamus slang dari https://github.com/krisnawans/Kode-Analisis-Sentimen-Program-MBG-Bi-GRU. Jumlah kata sebanyak 6286.

Tabel 4. Contoh Hasil Normalisasi

Sebelum	Sesudah
banyak fakta seperti ini menurut saya kebijakan mbg ini sebuah pemborosan target yang ingin dicapai tidak tercapai malahan yang terjadi pemborosan karena makanan banyak tidak habis dan akhirnya dibuang coba kebijakan mbg ditujukan kpd murid yg benar² membutuhkan lbh efektif	banyak fakta seperti ini menurut saya kebijakan makanan bergizi gratis ini sebuah pemborosan target yang ingin dicapai tidak tercapai malahan yang terjadi pemborosan karena makanan banyak tidak habis dan akhirnya dibuang coba kebijakan makanan bergizi gratis ditujukan kepada murid yang benar-benar membutuhkan lebih efektif

Tahap kelima tokenisasi, Setiap kalimat diurai menjadi bagian-bagian lebih kecil yang dinamakan token. Penelitian ini menggunakan *library Natural language tool kit* (NLTK).

Tabel 5. Contoh Hasil Tokenisasi

Sebelum	Sesudah
banyak fakta seperti ini menurut saya kebijakan makanan bergizi gratis ini sebuah pemborosan target yang ingin dicapai tidak tercapai malahan yang terjadi pemborosan karena makanan banyak tidak habis dan akhirnya dibuang coba kebijakan makanan bergizi gratis ditujukan kepada murid yang benar-benar membutuhkan lebih efektif	['banyak', 'fakta', 'seperti', 'ini', 'menurut', 'saya', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ini', 'sebuah', 'pemborosan', 'target', 'yang', 'ingin', 'dicapai', 'tidak', 'tercapai', 'malahan', 'yang', 'terjadi', 'pemborosan', 'karena', 'makanan', 'banyak', 'tidak', 'habis', 'dan', 'akhirnya', 'dibuang', 'coba', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ditujukan', 'kepada', 'murid', 'yang', 'benar-benar', 'membutuhkan', 'lebih', 'efektif']

Tahap keenam *Stopword Removal*, menyeleksi dan mengeliminasi kata-kata yang dianggap tidak memiliki bobot. Menggunakan *library python* Sastrawi

Tabel 6. Contoh Hasil *Stopword*

Sebelum	Sesudah
['banyak', 'fakta', 'seperti', 'ini', 'menurut', 'saya', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ini', 'sebuah', 'pemborosan', 'target', 'yang', 'ingin', 'dicapai', 'tidak', 'tercapai', 'malahan', 'yang', 'terjadi', 'pemborosan', 'karena', 'makanan', 'banyak', 'tidak', 'habis', 'dan', 'akhirnya', 'dibuang', 'coba', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ditujukan', 'kepada', 'murid', 'yang', 'benar-benar', 'membutuhkan', 'lebih', 'efektif']	['banyak', 'fakta', 'saya', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ini', 'sebuah', 'pemborosan', 'target', 'ingin', 'dicapai', 'tercapai', 'malahan', 'terjadi', 'pemborosan', 'makanan', 'banyak', 'habis', 'akhirnya', 'dibuang', 'coba', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ditujukan', 'murid', 'benar-benar', 'membutuhkan', 'lebih', 'efektif']

Tahap terahkir *stemming* mengubah kata imbuhan ke bentuk dasarnya

Tabel 7. Contoh Hasil *Stemming*

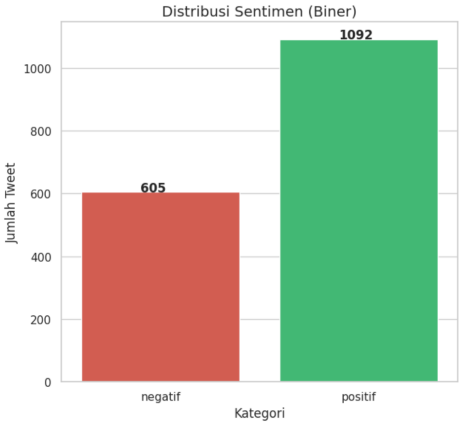
Sebelum	Sesudah
['banyak', 'fakta', 'saya', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ini', 'sebuah', 'pemborosan', 'target', 'ingin', 'dicapai', 'tercapai', 'malahan', 'terjadi', 'pemborosan', 'makanan', 'banyak', 'habis', 'akhirnya', 'dibuang', 'coba', 'kebijakan', 'makanan', 'bergizi', 'gratis', 'ditujukan', 'murid', 'benar-benar', 'membutuhkan', 'lebih', 'efektif']	['banyak', 'fakta', 'saya', 'bijak', 'makan', 'gizi', 'gratis', 'ini', 'buah', 'boros', 'target', 'ingin', 'capai', 'capai', 'malah', 'jadi', 'boros', 'makan', 'banyak', 'habis', 'akhir', 'buang', 'coba', 'bijak', 'makan', 'gizi', 'gratis', 'tuju', 'murid', 'benar', 'butuh', 'lebih', 'efektif']

3.3. *Labeling* data

Setelah tahap *preprocessing* kemudian lanjut ke tahap pelabelan, pada tahap ini proses pelabelan secara otomatis menggunakan *Bidirectional Encoder Representations from Transformers* (BERT). Dapat di

P-ISSN : 2089-3353
E-ISSN : 2808-9162

lihat pada laman berikut; https://huggingface.co/taufiqdp/indonesian-sentiment. Berdasarkan hasil pelabelan otomatis menggunakan BERT didapatkan hasil, positif sebanyak 1092 data, negatif 605 data dan netral sebanyak 275. Pada penelitian ini sentimen netral di hapus.



Gambar 3. Hasil Pelabelan *Bert*

3.4. TF-IDF

TF-IDF atau *Term Frequency-Inverse Document Frequency* digunakan untuk me-ngektraksi fitur. Sebelum dilakukan ektraksi fitur kita lakukan pembagian data atau split data dengan perbandingan 80:20. Diperoleh data uji 340 dan data latih sebanyak 1357.

```
========================================
HASIL PEMBAGIAN DATA (80:20) & TF-IDF
========================================
Total Data Keseluruhan     : 1697
Jumlah Data Latih (80%)    : 1357
Jumlah Data Uji (20%)      : 340
Jumlah Fitur (Kata Unik)   : 3157
----------------------------------------
```

Gambar 4. Hasil Total Data Uji Dan Data Latih

3.5. *Balancing* data

Proses penyeimbangan data menggunakan metode *SMOTE* agar mendapatkan hasil yg lebih baik. Karena terjadi tidak keseimbangan data pada hasil pelabelan *BERT*, Hasil pelabelan *BERT* menghasilkan nilai positif lebih tinggi daripada nilai negatif. *SMOTE* hanya diterapkan pada data latih untuk mencegah data leakage. Data uji tetap menggunakan distribusi asli sebanyak 219 sampel positif dan 121 sampel negatif. Maka dari itu diperlukan peningkatan pada nilai negatif agar memperoleh hasil model yang lebih optimal. Hasil *oversampling* dengan menerapkan *SMOTE* dapat dilihat pada tabel berikut

Tabel 8. Hasil *Oversampling Smote*

Label	Sebelum *SMOTE*	Setelah *SMOTE*
Positif	873	873
Negatif	484	873

3.6. *Cross Validation*

Cross validation adalah metode evaluasi statistik yang digunakan untuk menilai generalisasi model dengan cara membagi *dataset* menjadi lima subset atau *fold* yang berukuran sama. Dalam skema 5-*fold cross validation* ini, pengujian dijalankan dalam lima tahap, di mana setiap tahapnya satu *fold* berperan sebagai data uji, sementara empat *fold* tersisa digunakan sebagai data latih. Prosedur ini bertujuan untuk memastikan bahwa seluruh bagian data digunakan dalam proses pelatihan maupun pengujian, sehingga mampu menghasilkan estimasi performa model yang lebih stabil dan bebas dari bias [25].

Tabel 9. Hasil *Validation* Model Menggunakan *Cross Validation* Dengsn Optimasi *Smote*

Iterasi	SVM + *SMOTE* + *Grid Search*	*Random Forest* + *SMOTE* + *Grid Search*
1	79.78%	79.41%
2	79.78%	77.57%
3	78.23%	79.34%
4	80.81%	78.23%
5	78.23%	78.60%
Rata - Rata	79.37%	78.63%

Berdasarkan tabel 8, hasil pengujian menggunakan metode 5-*Fold Cross Validation* dengan bantuan optimasi *Grid Search* algoritma SVM memperoleh rata-rata akurasi sebesar 79,37%, di mana penggunaan *Kernel linear* ditemukan sebagai parameter yang paling optimal dalam memisahkan kelas sentimen pada *dataset* ini. Sementara itu, algoritma *Random Forest* menghasilkan rata-rata akurasi sebesar 78,63% dengan menggunakan konfigurasi n_*estimators*= 300. Secara deskriptif, SVM mencatatkan nilai rata-rata akurasi yang sedikit lebih tinggi dengan selisih 0,74%. Penggunaan *kernel linear* pada SVM terbukti optimal dalam memisahkan batas keputusan pada dimensi tinggi, sedangkan *Random Forest* dengan 300 *trees* memberikan performa yang stabil setelah *dataset* diseimbangkan menggunakan *SMOTE*. Namun, untuk memastikan apakah selisih performa yang tipis ini berdampak signifikan atau tidak, pengujian signifikansi statistik dilakukan lebih lanjut pada bagian berikutnya.

Tabel 10. Hasil Uji Statistic (*Paired t-Test*)

Parameter	Nilai
T-Statistic	1.0237
P-Value	0.3638

Tabel 9 menyajikan hasil uji signifikansi statistik menggunakan *Paired t-Test* untuk membandingkan performa SVM dan *Random Forest* dari skor 5-*Fold Cross Validation*. Hasil pengujian menghasilkan nilai *t-statistic* = 1.0237 dan *p-value* = 0.3638. Karena batas ambang nilai *p-value* > 0.05, dapat disimpulkan bahwa secara statistik tidak terdapat perbedaan performa yang signifikan antara algoritma SVM dan *Random Forest*. Selisih akurasi yang terjadi tergolong tipis dan dipengaruhi oleh variasi acak pembagian data, sehingga kedua algoritma terbukti memiliki tingkat

P-ISSN : 2089-3353
E-ISSN : 2808-9162

keandalan yang seimbang dalam memprediksi sentimen publik.

3.7. *Wordcloud*

WordCloud merupakan teknik visualisasi data teks yang menampilkan kata-kata dari sebuah dokumen atau kumpulan data, di mana ukuran font kata mencerminkan frekuensi kemunculannya. Berikut hasil output visualisasi teks sentimen positif maupun negatif.

Gambar 5. *Wordcloud* Sentimen Positif Dan Negatif

Gambar 5 memperlihatkan perbedaan pandangan masyarakat yang sangat jelas terhadap program Makan Bergizi Gratis. Pada bagian sentimen positif, kosakata yang paling menonjol ialah "lanjut", "sehat", "baik", dan "manfaat", yang menandakan bahwa banyak orang mendukung program ini agar terus dijalankan karena dianggap sangat berguna bagi kesehatan anak-anak. Sebaliknya, pada bagian sentimen negatif, muncul kata "racun" sebagai kekhawatiran utama masyarakat terkait keamanan makanan yang diberikan, serta kata "korupsi", "salah", dan "beban" yang menunjukkan adanya rasa tidak percaya terhadap pengelolaan anggaran dan cara pelaksanaan program di lapangan. Secara keseluruhan, gambar ini menggambarkan bahwa dukungan masyarakat muncul karena harapan akan perbaikan gizi, sementara penolakan muncul karena rasa takut akan risiko keracunan dan penyalahgunaan dana.

3.8. Evaluasi

Langkah terakhir yang dilakukan yaitu tahap evaluasi membandingkan hasil akurasi kedua algoritma *Support Vector Machine* (SVM) dan *Random Forest* dengan menggunakan *confusion matrix.*

Hasil pengujian menggunakan algoritma *Support Vector Machine* (SVM) dengan Kernel Linear yang dikombinasikan dengan teknik *SMOTE* menunjukkan performa yang solid dengan nilai akurasi sebesar 78,24%. Model ini terbukti sangat optimal dalam mengklasifikasi sentimen positif dengan nilai *precision* sebesar 0,83, *recall* 0,83, dan *f1-score* 0,83, yang menandakan kemampuan tinggi dalam mengenali opini mendukung terhadap program Makan Bergizi Gratis (MBG). Sementara itu, pada klasifikasi sentimen negatif, model menunjukkan hasil yang konsisten dengan nilai *precision*, *recall*, dan *f1-score* masing-masing sebesar 0,69, di mana penggunaan *SMOTE* terbukti efektif meminimalisir bias meskipun terdapat perbedaan jumlah sampel antara kelas positif (219 data) dan negatif (121 data). Secara keseluruhan, nilai *macro average* sebesar 0,76 dan *weighted average* sebesar 0,78 menegaskan bahwa model SVM memiliki kinerja yang *robust* dan seimbang dalam menangani ketidakseimbangan data pada penelitian ini.

```
============================
HASIL EVALUASI: SVM + SMOTE (KERNEL LINEAR)
============================
=== Confusion Matrix ===
[[ 84  37]
 [ 37 182]]

=== Classification Report ===
              precision    recall  f1-score   support

     negatif       0.69      0.69      0.69       121
     positif       0.83      0.83      0.83       219

    accuracy                           0.78       340
   macro avg       0.76      0.76      0.76       340
weighted avg       0.78      0.78      0.78       340

Akurasi SVM: 78.24%
```

Gambar 6. Hasil *Support Vector Machine*

Hasil pengujian menggunakan algoritma *Random Forest* dengan konfigurasi 300 *estimators* dan teknik *SMOTE* menunjukkan performa yang cukup baik dengan nilai akurasi sebesar 75,00%. Model ini memiliki kemampuan klasifikasi yang lebih kuat pada sentimen positif, dengan nilai *precision* sebesar 0,80, *recall* 0,81, dan *f1-score* 0,81, yang mencerminkan efektivitas model dalam mengenali mayoritas opini mendukung program Makan Bergizi Gratis (MBG). Sementara itu, pada klasifikasi sentimen negatif, model menunjukkan performa yang moderat dengan nilai *precision* sebesar 0,65, serta *recall* dan *f1-score* sebesar 0,64. Meskipun hasil ini berada di bawah performa SVM, penggunaan 300 pohon keputusan tetap memberikan stabilitas yang cukup baik dalam menangani data uji asli yang tidak seimbang (219 positif dan 121 negatif). Secara keseluruhan, nilai *macro average* sebesar 0,73 dan *weighted average* 0,75 mengindikasikan bahwa *Random Forest* mampu memberikan prediksi yang cukup konsisten, meskipun masih terdapat ruang untuk optimasi dalam membedakan fitur-fitur sentimen negatif pada *dataset* ini.

```
============================
HASIL EVALUASI: RANDOM FOREST + SMOTE (300 ESTIMATORS)
============================
=== Confusion Matrix ===
[[ 77  44]
 [ 41 178]]

=== Classification Report ===
              precision    recall  f1-score   support

     negatif       0.65      0.64      0.64       121
     positif       0.80      0.81      0.81       219

    accuracy                           0.75       340
   macro avg       0.73      0.72      0.73       340
weighted avg       0.75      0.75      0.75       340

Akurasi Random Forest: 75.00%
```

Gambar 7. Hasil *Random Forest*

4. Kesimpulan

Melalui penelitian yang telah dilakukan, temuan bahwa *Support Vector Machine* (SVM) dengan Kernel Linear

P-ISSN : 2089-3353
E-ISSN : 2808-9162

mengungguli *Random Forest* dalam hal akurasi klasifikasi sentimen mengenai program Makan Bergizi Gratis. Hal ini dibuktikan dengan capaian akurasi SVM sebesar 78% dengan nilai *f1-score* yang seimbang sebesar 0,83 pada kelas positif dan 0,69 pada kelas negatif. Implementasi teknik *SMOTE* dalam penelitian ini terbukti sangat krusial dalam mengatasi ketidakseimbangan data, di mana SVM mampu mendeteksi sentimen minoritas (negatif) dengan nilai *recall* mencapai 69%, mengungguli *Random Forest* yang hanya mencapai akurasi 75% dengan *recall* negatif sebesar 0,64. Dengan demikian, kombinasi SVM dan *SMOTE* memberikan keseimbangan performa yang paling optimal dalam mengenali pola opini publik di *platform* X dibandingkan penggunaan algoritma *ensemble* seperti *Random Forest*.

Penelitian ini memiliki keterbatasan. Pertama, sebanyak 275 data yang berlabel netral hasil seleksi BERT langsung dihapus dan tidak dianalisis lebih lanjut. Kedua, pengambilan data hanya menggunakan satu kata kunci saja, yaitu 'MBG', sehingga data yang didapat mungkin belum menggambarkan seluruh obrolan masyarakat di media sosial.

Untuk penelitian selanjutnya, disarankan bagi peneliti berikutnya untuk mengeksplorasi teknik ekstraksi fitur berbasis *word embedding* seperti *Word2Vec* atau *FastText* guna menangkap konteks semantik kata yang lebih mendalam pada data media sosial. Selain itu, pengembangan model dapat dilakukan dengan menguji algoritma *Deep Learning* seperti *Long Short-Term Memory* (LSTM) atau melakukan *fine-tuning* lebih lanjut pada model BERT untuk meningkatkan sensitivitas klasifikasi pada sentimen negatif yang memiliki tingkat ambiguitas tinggi.

Daftar Rujukan

[1] K. H. Prastiawan and D. Yuniarto, "Analisis Sentimen Publik terhadap Program Makan Bergizi Gratis dengan Algoritma Naive Bayes," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 5412–5419, Dec. 2025, doi: 10.31004/riggs.v4i4.3652.

[2] V. Rizqika, P. Handoko, and A. Rochmania, "Self disclosure generasi z melalui media x (twitter)," *Jurnal Ilmu Komunikasi UHO : Jurnal Penelitian Kajian Ilmu Komunikasi dan Informasi*, vol. 10, no. 2, pp. 397–411, 2025, doi: 10.52423/jikuho.v10i2.348.

[3] simon Kemp, "Digital 2024: Indonesia," 2024. [Online]. Available: https://datareportal.com/reports/digital-2024-indonesia

[4] S. D. Waluyo, "Kebijakan makanan bergizi gratis: tinjauan ekonomi politik dalam kesejahteraan dan ketahanan pangan," 2025, doi: http://dx.doi.org/10.25157/dak.v12i1.18223.

[5] Associated press, "Indonesia launches free meals program to feed children and pregnant women to fight malnutrition." [Online]. Available: https://apnews.com/article/indonesia-prabowo-subianto-free-meals-children-mothers-213a04587203434f3f85950725e84a8b

[6] BBC, "Anggaran MBG sebesar Rp335 triliun digugat ke MK karena 'memakan' sepertiga dana pendidikan-Seberapa mungkin dikabulkan hakim?," 2026. [Online]. Available: https://www.bbc.com/indonesia/articles/cn42dl7wykpo

[7] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis sentimen ulasan pengguna aplikasi samsat digital nasional (signal) dengan menggunakan metofe naive bayes classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4073.

[8] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. H. Rakhmawati, "Perbandingan Metode Klasifikasi *Random Forest* dan Svm pada analisis sentimen PSBB," 2021, doi: 10.26877/jiu.v7i1.7099.

[9] P. Kumala Sari and R. Randy Suryono, "Komparasi algoritma *support vector machine* dan *random forest* untuk analisis sentimen metaverse," 2024. doi: https://doi.org/10.36040/mnemonic.v7i1.8977.

[10] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "Sentimen analisis publik terhadap kebijakan lockdown pemerintah jakarta menggunakan algoritma svm," *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, doi: https://doi.org/10.33365/jdmsi.v2i1.1021.

[11] N. Sakina, F. Wajidi, and Muh. R. Rasyid, "Evaluasi Algoritma KNN dan Naive Bayes untuk Analisis Sentimen Kebijakan Program Makan Bergizi Gratis," *Jambura Journal of Informatics*, vol. 1, no. 2, pp. 83–97, Oct. 2025, doi: 10.37905/jji.v1i2.34418.

[12] A. Ilham and W. Pramusinto, "Analisis sentimen masyarakat terhadap kesehatan mental pada twitter menggunakan algoritma k-nearest neighbor," 2023, Accessed: Mar. 30, 2026. [Online]. Available: https://senafti.budiluhur.ac.id/senafti/article/view/792

[13] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan *Support Vector Machine*," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY*, vol. 2, no. 1, 2022, doi: 10.46229/jifotech.v2i1.330.

[14] Y. Ansori and K. F. H. Holle, "Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 10, no. 4, p. 429, Dec. 2022, doi: 10.26418/justin.v10i4.51784.

[15] H. Harnelia and A. Saputra, "Analisis sentimen review skincare skintific dengan algoritma *support vector machine* (svm)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4095.

[16] H. Al Rasyid Harpizon *et al.*, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022, doi: 10.32672/jnkti.v5i1.4008.

[17] R. Merdiansah and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI*, vol. 7, no. 1, pp. 221–228, 2024, doi: https://doi.org/10.55338/jikomsi.v7i1.2895.

[18] R. Rahmadani, A. Rahim, and R. Rudiman, "Analisis sentimen ulasan 'ojol the game'di google play store menggunakan algoritma naive bayes dan model ektraksi fitur tf-idf untuk meningkatkan kualitas game," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4988.

[19] C. Candra, K. W. Chandra, and H. Irsyad, "Efektifitas *SMOTE* dalam Mengatasi Imbalanced Class Algoritma K-Nearest Neighbors pada Analisis Sentimen terhadap Starlink," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 31–42, Jul. 2024, doi: 10.54082/jiki.132.

[20] M. Fajri and A. Primajaya, "Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan *Grid Search* dan Random Search," 2023. doi: https://doi.org/10.30871/jaic.v7i1.5004.

[21] P. D. Santoso and W. Wibowo, "Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store," 2022. doi: 10.12962/j23373520.v11i2.72534.

[22] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma *Support Vector Machine* (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bulletin of Information Technology (BIT)*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

[23] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma *Random Forest*," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, Jan. 2023, doi: 10.35970/infotekmesin.v14i1.1751.

[24] G. Rininda, I. H. Santi, and S. Kirom, "Penerapan svm dalam analisis sentimen pada edlink menggunakan pengujian confusion matrix," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 5, 2023, doi: 10.36040/jati.v7i5.7420.

[25] D. A. Sani and M. Z. Sarwani, "A Random *Oversampling* and BERT-based Model Approach for Handling Imbalanced Data in Essay Answer Correction," vol. 16, no. 4, pp. 729–739, 2024, doi: 10.20895/INFOTEL.V16I3.1224.