Jurnal Dinamika Informatika
Volume 14, No 2, Oktober 2025
ISSN 1978-1660
ISSN *online* 2549-8517

SMOTE untuk Meningkatkan Performa Naïve Bayes dan Random Forest dalam Analis Sentimen aplikasi Digitalent

Yusril Muhamad Izha Mahendra*[1], Ahmad Faqih[2], Kaslani[3]
Teknik Informataika STMIK IKMI CIREBON[12], Komputerisasi Akuntansi, STMIK IKMI CIREBON[3]
e-mail: yusrilmizham51@gmail.com[1] ahmadfaqih367@gmail.com[2] kaslani@ikmi.ac.id [3]

Intisari

Analisis sentimen sangat penting untuk memahami bagaimana suatu aplikasi, seperti aplikasi pelatihan digital seperti Digitalent, dilihat oleh pengguna. Ulasan pengguna yang tersedia di platform distribusi aplikasi memberikan data yang cukup untuk analisis ini. Namun, dalam analisis sentimen, ketidakseimbangan data menjadi masalah umum; ulasan positif cenderung lebih banyak dibandingkan dengan ulasan negatif dan netral. Ketidakseimbangan ini dapat berdampak pada model pembelajaran mesin, yang dapat menyebabkan prediksi yang tidak akurat terhadap kelas mayoritas. Tujuan penelitian ini adalah untuk memecahkan masalah ini dengan menggunakan teknik SMOTE (Teknik Pemilihan Minoritas Sintetis) dalam analisis sentimen ulasan aplikasi Digitalent dan membandingkan kinerja dua algoritma pembelajaran mesin, Naive Bayes dan Random Forest. Data penelitian dikumpulkan dari ulasan pengguna berbahasa Indonesia dari platform Digitalent. Sebelum diproses untuk analisis, data melalui proses pra-pemrosesan seperti pembersihan, tokenisasi, dan normalisasi. Teknik SMOTE diterapkan untuk menyeimbangkan jumlah ulasan untuk setiap kelas sentimen. Selanjutnya, algoritma Naive Bayes dan Random Forest digunakan untuk mengkategorikan sentimen. Hasil penelitian penerapan SMOTE berhasil meningkatkan proporsi kelas negatif dan netral, sehingga distribusi dataset menjadi seimbang. Hasil pengujian menunjukkan bahwa akurasi Naïve Bayes meningkat dari 68,25% menjadi 92,16%, sementara Random Forest meningkat dari 80,16% menjadi 92,12%. Meskipun Naïve Bayes mengalami peningkatan yang lebih besar, Random

Forest tetap lebih akurat secara keseluruhan. Penerapan SMOTE terbukti efektif dalam meningkatkan kinerja kedua algoritma dalam klasifikasi data tidak seimbang.

Kata kunci: Analisis Sentimen, Naïve Bayes, Random Forest, Penerapan Smote, Digitalent

Abstract

Sentiment analysis is critical to understanding how an app, such as a digital training app like Digitalent, is viewed by users. User reviews available on app distribution platforms provide ample data for this analysis. However, in sentiment analysis, data imbalance is a common problem; positive reviews tend to outnumber negative and neutral reviews. This imbalance can impact machine learning models, which can lead to inaccurate predictions of the majority class. The purpose of this research is to solve this problem by using SMOTE (Synthetic Minority Selection Technique) technique in sentiment analysis of Digitalent app reviews and comparing the performance of two machine learning algorithms, Naive Bayes and Random Forest. The research data was collected from Indonesian user reviews from the Digitalent platform. Before being processed for analysis, the data went through pre-processing processes such as cleaning, tokenization, and normalization. SMOTE technique was applied to balance the number of reviews for each sentiment class. Furthermore, Naive Bayes and Random Forest algorithms are used to categorize the sentiment. The results of the SMOTE application research successfully increased the proportion of negative and neutral classes, so that the distribution of the dataset became balanced. The test results show that the accuracy of Naïve Bayes increased from 68.25% to 92.16%, while Random Forest increased from 68.25% to 92.16%.

Keywords: *Sentiment Analysis, Naïve Bayes, Random Forest, Smote Implementation, Digitalent*

Jurnal Dinamika Informatika
Volume 14, No 2, Oktober 2025
ISSN 1978-1660
ISSN *online* 2549-8517

PENDAHULUAN

Perkembangan pesat di bidang informatika telah membawa dampak signifikan pada berbagai aspek kehidupan, termasuk teknologi, bisnis, dan pendidikan. Di era digital, sekarang kemajuan teknologi menjadi bagian penting dari pendidikan [1]. Aplikasi mobile, platform pendidikan digital, dan media sosial kini menjadi bagian tak terpisahkan dari interaksi pengguna dengan teknologi [2]. Salah satu aplikasi pendidikan yang berkembang pesat di Indonesia adalah Digitalent, yang menarik perhatian banyak pengguna sebagai sarana edukasi dan pelatihan. Selain memberikan manfaat pembelajaran, aplikasi ini juga menyediakan data berharga melalui ulasan pengguna. Ulasan tersebut memberikan wawasan penting mengenai pengalaman dan persepsi pengguna yang dapat dimanfaatkan untuk meningkatkan kualitas produk dan strategi pengembangan aplikasi.

Penelitian ini berfokus pada analisis sentimen ulasan pengguna aplikasi Digitalent di Google Play Store dengan menggunakan pendekatan berbasis SMOTE (Synthetic Minority Oversampling Technique). Metode ini diterapkan untuk mengatasi masalah ketidakseimbangan distribusi data sentimen, di mana ulasan sering kali lebih dominan di salah satu kategori, seperti positif, negatif, atau netral. Ketidakseimbangan ini dapat menyebabkan kecenderungan pada Machine learning, yang lebih condong memprediksi kelas mayoritas dan mengabaikan kelas minoritas. Dengan menerapkan SMOTE, diharapkan performa model analisis sentimen dapat ditingkatkan secara signifikan, sehingga memberikan hasil yang lebih akurat dan representatif. Pada penelitian sebelumnya seperti yang dilakukan oleh [3] menunjukkan bahwa hasil dari penerapan SMOTE, membuktikan bahwa operator SMOTE memang efektif untuk mengatasi kondisi data tidak seimbang, terbukti dengan kenaikan accuracy dari 78% menjadi sebesar 85.87%. Adapun penelitian yang dilakukan oleh [4] Hasil penelitiannya menunjukkan bahwa metode optimasi Information Gain dan SMOTE mampu meningkatkan nilai akurasi, recall, dan F1-score secara rata-rata sebesar 6.25%, 23.9%, dan 25.44%. Adapun [5] menyatakan bahwa teknik SMOTE dapat mengatasi

ketidakseimbangan kelas dan meningkatkan prediksi untuk kelas minoritas, yang menghasilkan hasil yang lebih relevan dan akurat dalam proses klasifikasi.

Dalam penelitian ini, pendekatan yang akan digunakan mengintegrasikan metode penelitian eksperimental dengan teknik analisis data yang relevan dalam bidang Informatika, khususnya dalam konteks analisis sentimen. Penelitian ini akan menerapkan 2 metode analisis sentimen utama, Naive Bayes dan Random Forest untuk mengevaluasi ulasan pengguna aplikasi Digitalent di Google Play Store [6]. Seperti yang dikatakan [7] kenapa menggunakan algoritma Naive bayes karena memiliki performa yang cepat dalam melatih data, mudah dalam implementasinya, serta memiliki efektifitas yang tinggi, terbukti juga dari jurnal [8] bahwa hasil dari penelitiannya menunjukan algoritma Naive Bayes cukup baik dengan nilai akurasi sebesar 81,15% dibandingkan dengan Decision Tree memiliki nilai akurasi 68,19%. Algoritma Random Forest juga mampu mengklasifikasikan sentimen dengan baik [9]. Setiap metode akan diimplementasikan secara sistematis, dimulai dari tahap preprocessing data yang mencakup pembersihan data (Cleaning), normalisasi huruf (case folding), normalisasi kata, penyaringan data (filtering) , dan pemisahan atau penguraian (Tokenizing) sesuai dengan pendekatan yang telah terbukti efektif dalam literatur sebelumnya (Hasugian et al., 2023) Evaluasi kinerja akan menggunakan metrik standar dalam analisis sentimen, termasuk akurasi, presisi, dan recall [10] yang akan dihitung melalui prosedur validasi silang k-fold untuk memastikan kekuatan hasil seperti pada penelitian tersebut. Analisis komparatif akan dilengkapi dengan uji signifikansi statistik untuk menentukan apakah perbedaan kinerja antar metode signifikan secara statistik.

METODE PENELITIAN

2.1. Metode Pengumpulan Data

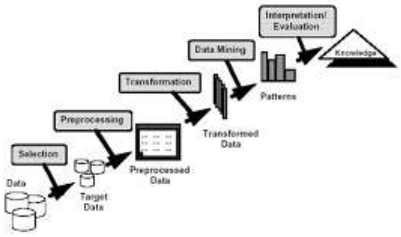
Metode pengumpulan data dalam penelitian ini dilakukan melalui teknik web scraping pada ulasan aplikasi Digitalent yang tersedia di Google Play Store. Web scraping adalah proses otomatisasi pengambilan data dari situs web menggunakan skrip atau perangkat lunak khusus. Dalam konteks ini, Google Play Scraper digunakan untuk mengakses dan mengekstrak data ulasan aplikasi

Jurnal Dinamika Informatika
Volume 14, No 2, Oktober 2025
ISSN 1978-1660
ISSN *online* 2549-8517

Digitalent, termasuk teks ulasan, rating, serta informasi relevan lainnya yang dapat digunakan untuk analisis sentimen. Teknik ini memungkinkan pengumpulan data dalam jumlah besar dan terstruktur, sehingga memudahkan proses analisis lebih lanjut. Data yang diperoleh melalui scraping kemudian diolah dan dianalisis untuk mengetahui sentimen pengguna terhadap aplikasi tersebut [11].

2.2. Metode Pengolahan Data

Dalam penelitian analisis sentimen ulasan aplikasi *Digitalent*, pendekatan yang digunakan adalah *Knowledge Discovery in Databases* (*KDD*), sebuah proses sistematis yang terdiri dari beberapa tahapan yang bertujuan untuk mengekstrak informasi berharga dari data yang besar dan kompleks, seperti ulasan aplikasi Digitalent. [6] penelitian tersebut juga menggunakan metode KDD untuk metode analisis



Gambar 1. Metode Analisis

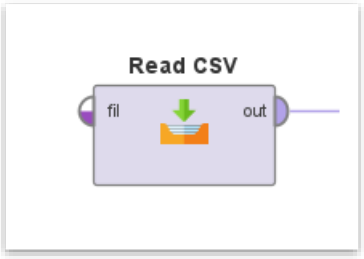
HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Data pada penelitian ini diambil melalui scraping data ulasan aplikasi *Digitalent* dari *Google Play Store* menggunakan *Google Colaboratory*. Proses ini dilakukan untuk memperoleh ulasan pengguna yang relevan dengan data yang diperoleh hanya mencakup 630 ulasan terbaru dalam bahasa Indonesia, yang dikumpulkan untuk analisis lebih lanjut. Berikut adalah tampilan proses *Scrapping*

3.1.1. Data *Selection*

Dalam memasukan dataset yang telah kita ambil melalui *Scrapping* menggunakan *Google Colaboratory*, kita memerlukan operator *Read CSV* dalam Tools *Altair AI Studio* untuk mengimpor data.



Gambar 2. *Operator CSV*

Gambar 1 menampilkan operator *read CSV* untuk menampilkan data, hal ini bertujuan untuk menampilkan hasil atau proses dari pemilihan data. Setelah impor data berhasil, langkah berikutnya adalah memilih atribut, menghapus atribut, dan memilih label.

3.1.2. *Preprocessing Data*

Setelah data diimpor menggunakan *Read CSV*, Langkah selanjutnya adalah melakukan proses analisis, berikut adalah lagkah-langkah yang dapat dilakukan sebagai berikut:

Gambar 3. Pembersihan data

Tahapan preproses data bertujuan untuk mempersiapkan data sebelum dianalisis. Pertama, data dibersihkan dengan menangani nilai yang hilang. Kemudian, pengubahan data nominal menjadi teks. Dengan demikian, data nominal yang berbentuk kategori atau label akan dikonversi ke dalam format teks untuk memudahkan analisis lebih lanjut, terutama pada langkah-langkah seperti eksplorasi data, pelabelan, dan pemodelan. Langkah berikutnya adalah

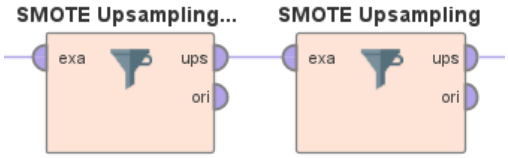
Jurnal Dinamika Informatika
Volume 14, No 2, Oktober 2025
ISSN 1978-1660
ISSN *online* 2549-8517

melakukan proses dokumen yang mencakup beberapa rangkaian operasi seperti *Tokenize, Transform Case, Filter Stopwords, Filter Tokens,* dan *Stem*.

3.1.3. *Transformation*

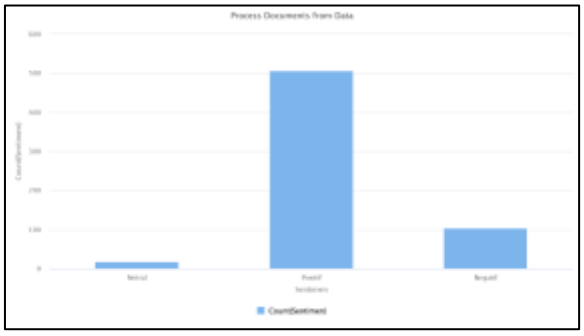
transformasi data ini adalah untuk membuat data lebih siap untuk pelatihan model dengan algoritma *Naive Bayes* dan *Random Forest*, yang digunakan dalam analisis sentimen ulasan aplikasi *Digitalent*. Proses transformasi yang dilakukan termasuk:

a. SMOTE Upsampling



Gambar 4. Operator *SMOTE*

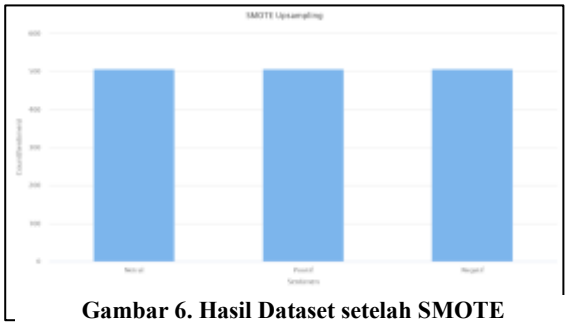
Tujuan dari tahapan ini untuk mengatasi ketidakseimbangan kelas dalam *dataset*. Metode *SMOTE (Synthetic Minority Over-sampling Technique)*, digunakan untuk meningkatkan jumlah sampel pada kelas minoritas. dua operator *SMOTE* digunakan secara berurutan, setiap operator diatur secara manual pada kelas negatif dan netral, yang memungkinkan fleksibilitas dalam mengatur jumlah sampel yang dibutuhkan. Berikut adalah data yang tidak seimbang



Gambar 5. Hasil *Dataset* sebelum *SMOTE*

Dataset yang digunakan terdiri dari 630 ulasan, 507 ulasan positif, 104 ulasan negatif dan 19 ulasan netral. Berikut adalah hasil setelah penerapan smote:

Berdasarkan hasil penelitian pada kedua gambar diatas didapatkan bahwa

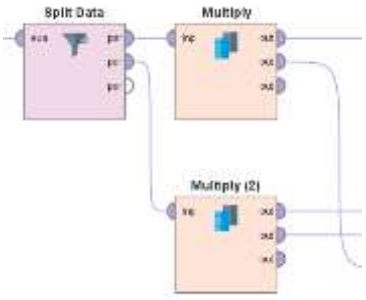


Gambar 6. Hasil Dataset setelah SMOTE

pengaruh *SMOTE* dapat menyeimbangkan data, menjadi 507 pada setiap kelas. seperti [12] menyatakan setelah menerapan teknik *SMOTE,* meningkatkan data dari jumlah berita objektif sebanyak 176 dan berita subjektif sebanyak 24 menjadi 352 datadengan jumlah berita objektif sebanyak 176 data dan jumlah berita subjektif sebanyak 176 data.

Jurnal Dinamika Informatika
Volume 14, No 2, Oktober 2025
ISSN 1978-1660
ISSN *online* 2549-8517

b. Split Data dan *Multiply*

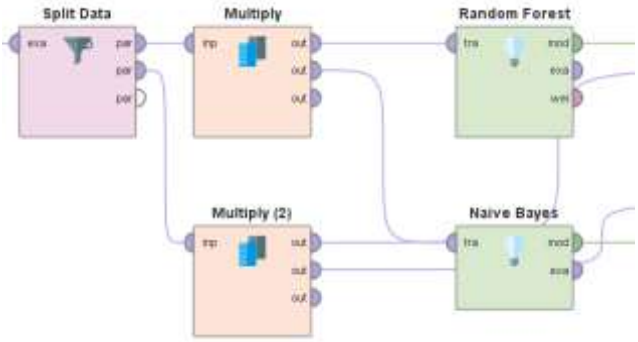


Gambar 7. Pembagian data latih

Operator ini membagi *dataset* menjadi data pelatihan dan pengujian, yang memungkinkan evaluasi model secara lebih akurat. langkah selanjutnya adalah melatih *Algoritma Naive Bayes* dan *Random Forest* menggunakan data data latih dan data uji. Kemudian kita gunakan operator *Multiply* untuk menggabungkan fitur yang ada pada *dataset*.

3.1.4. *Data Mining*

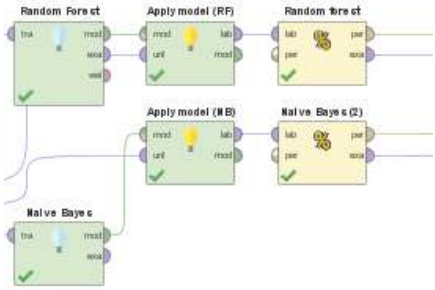
Pada tahap Data Mining, Untuk menganalisis sentimen pada data ulasan, algoritma *Naive Bayes* dan *Random Forest* digunakan. Pada tahap ini, model dibangun dengan tujuan untuk menemukan pola atau hubungan tersembunyi dalam data yang dapat mengungkap sentimen pengguna terhadap aplikasi Digitalent.



Gambar 8. Menerapkan algoritma

3.1.5. *Evaluasi*

Evaluasi adalah tahapan akhir dari *Knowledge Discovery in Database* (*KDD*). Dari proses ini Dimana kitab isa melihat dan menilai hasil yang telah diperoleh selama pe



Gambar 9. Pengujian hasil

KESIMPULAN

Penelitian ini bertujuan untuk meningkatkan kinerja algoritma *Naive Bayes* dan *Random Forest* dalam analisis sentimen ulasan aplikasi *Digitalent* dengan menggunakan metode *SMOTE*. Hasilnya menunjukkan bahwa:

1. Penerapan *SMOTE* pada algoritma *Naïve Bayes* bertujuan untuk menyeimbangkan distribusi kelas dalam dataset. Sebelum *SMOTE* diterapkan, algoritma ini menunjukkan performa yang baik dalam mengenali kelas mayoritas (positif), namun kesulitan dalam mengenali kelas minoritas (negatif dan netral). *SMOTE* menghasilkan dataset yang lebih seimbang dengan distribusi yang merata pada setiap kelas sentimen (positif, negatif, dan netral), yaitu masing-masing 33,33%. Setelah *SMOTE* diterapkan, algoritma *Naïve Bayes* mampu memberikan prediksi yang lebih akurat terhadap kelas minoritas, dengan recall dan precision pada kelas netral mencapai 100%. Hal ini membuktikan bahwa *SMOTE* secara signifikan membantu *Naïve Bayes* dalam mengidentifikasi pola pada kelas minoritas.
2. *SMOTE* juga diterapkan pada algoritma *Random Forest* untuk mengatasi ketidakseimbangan kelas dalam dataset ulasan aplikasi Digitalent. Tanpa *SMOTE*, Random *Forest* sudah menunjukkan performa yang cukup baik

dibandingkan *Naïve Bayes*, tetapi masih mengalami penurunan akurasi pada kelas minoritas. Dengan *SMOTE*, distribusi data yang lebih merata memungkinkan algoritma untuk belajar secara lebih optimal dari pola pada kelas negatif dan netral. Hasil evaluasi menunjukkan peningkatan signifikan pada *precision*, dan *recall* di seluruh kelas, termasuk pada kelas netral yang sulit diprediksi sebelumnya. Akurasi keseluruhan model setelah penerapan *SMOTE* mencapai 92,11%, menjadikan *Random Forest* algoritma yang lebih stabil dalam menganalisis sentimen dari data yang tidak seimbang.

3. Setelah penerapan *SMOTE*, baik *Naïve Bayes* maupun *Random Forest* menunjukkan peningkatan performa yang signifikan. *Naïve Bayes*, yang sebelumnya menghadapi kendala dalam memprediksi kelas minoritas, mengalami peningkatan akurasi dari sekitar 68-73% menjadi 89-92%. *Recall* dan precision pada kelas netral juga meningkat drastis hingga mencapai nilai maksimal pada berbagai skenario pembagian data. Di sisi lain, *Random Forest*, yang memiliki performa lebih baik sejak awal, menunjukkan peningkatan yang lebih konsisten. Akurasi tertinggi *Random Forest* setelah penerapan *SMOTE* adalah 92,11%, dengan metrik evaluasi lainnya (*precision*, *recall*) menunjukkan hasil yang unggul dibandingkan *Naïve Bayes*. Penerapan *SMOTE* terbukti menjadi langkah penting dalam meningkatkan kemampuan kedua algoritma dalam menangani dataset dengan ketidakseimbangan kelas.

SARAN

Berdasarkan analisis penelitian ini, terdapat beberapa saran yang dapat digunakan untuk pengembangan dan pelaksanaan penelitian di masa depan:

1. Eksplorasi Teknik *Oversampling* Lainnya: Selain *SMOTE*, teknik seperti *ADASYN* (*Adaptive Synthetic Sampling*) dapat digunakan untuk mengatasi ketidakseimbangan kelas selain *SMOTE. ADASYN* memberikan bobot yang lebih besar pada sampel minoritas kelas yang sulit diprediksi, memungkinkan model untuk belajar dari data yang lebih kompleks dengan lebih fokus.

Gambar 10. Hasil perbandingan Algoritma *Naïve Bayes* dan *Random Forest*

2. Evaluasi Temporal Sentimen: Analisis sentimen berbasis waktu, misalnya, dapat melihat bagaimana perasaan pengguna berubah setelah pembaruan aplikasi atau perubahan kebijakan. Penelitian lanjutan dapat melihat hasilnya. Hal ini akan membantu pengembang memahami bagaimana memenuhi kebutuhan pengguna.
3. Penerapan di Domain Lain: Metode penelitian ini dapat diterapkan pada bidang lain, seperti ulasan produk e-*commerce* atau respons masyarakat terhadap layanan pemerintah. Ini penting untuk mengevaluasi generalisasi strategi yang digunakan dalam berbagai situasi.
4. Algoritma Lain: Disarankan untuk mengeksplorasi algoritma lain seperti *Support Vector Machine (SVM*) atau model *deep learning* seperti *Recurrent Neural Network* (*RNN*)

DAFTAR PUSTAKA

[1] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine Evita," vol. 18, no. 1, pp. 71–80, 2020.

[2] H. Wahyono, "Pemanfaatan Teknologi Informasi dalam Penilaian Hasil Belajar pada Generasi Milenial di Era Revolusi Industri 4 . 0," vol. 3, pp. 192–201, 2019.

[3] K. Kurnia, I. Purnamasari, and D. D. Saputra, "Analisis Sentimen Dengan Metode Naïve Bayes, SMOTE Dan Adaboost Pada Twitter Bank BTN," J. JTIK (Jurnal Teknol. Inf. dan Komunikasi), vol. 7, no. 2, pp. 235–242, 2023, doi: 10.35870/jtik.v7i3.707.

[4] H. Hidayatullah, P. Purwantoro, and Y. Umaidah, "Penerapan Naïve Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt," JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 3, pp. 1546–1553, 2023, doi: 10.36040/jati.v7i3.6887.

[5] M. P. Pulungan, A. Purnomo, and A. Kurniasih, "PENERAPAN SMOTE UNTUK MENGATASI IMBALANCE CLASS DALAM KLASIFIKASI KEPRIBADIAN MBTI MENGGUNAKAN NAIVE BAYES CLASSIFIER," vol. 11, no. 5, pp. 1033–1042, 2024, doi: 10.25126/jtiik.2024117989.

[6] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "ANALISIS SENTIMEN ULASAN APLIKASI MOLA PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," J. Inf. Technol. Comput. Sci., vol. 5, no. 1, pp. 1–7, 2022.

[7] S. K. Wardani, Y. A. Sari, and Indriati, "Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan," vol. 5, no. 12, pp. 5582–5590, 2021.

[8] N. Amalia, T. Suprapti, and G. Dwilestari, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAKSANAAN KURIKULUM MBKM," vol. 18, pp. 57–64, 2023.

[9] S. N. Adhan, G. A. W. Ngurah, D. C. Arisona, I. Yahya, Agusrawati, and Ruslan, "ANALISIS SENTIMEN ULASAN APLIKASI WATTPAD DI GOOGLE PLAY STORE DENGAN METODE RANDOM FOREST," vol. 2, no. 1, pp. 6–15, 2024.

[10] R. Mursyid and A. D. Indriyanti, "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)," J. Informatics Comput. Sci., vol. 06, pp. 371–383, 2024, [Online]. Available: https://ejournal.unesa.ac.id/index.php/jinacs/article/view/61322%0Ahttps://ejournal.unesa.ac.id

[11] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," vol. 6, no. 9, pp. 4305–4313, 2022.

[12] Ridwan, E. H. Hermaliani, and M. Ernawati, "Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian," vol. 4, no. 1, 2024.

[13] C. Agustina and E. Rahmawati, "Optimalisasi Algoritma Random Forest Menggunakan SMOTE untuk Prediksi Pembatalan Tamu Hotel," vol. 12, no. 2, 2024.