

Komparasi Beberapa Algoritma Machine Learning untuk Klasifikasi Sentimen Pengguna X (Twitter) Terkait Isu Kabur Aja Dulu

Rayhan Dwi Ramadhan, Fatchul Arifin
Universitas Negeri Yogyakarta
Email: rayhandwi.2021@student.uny.ac.id

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen pengguna media sosial X (Twitter) terhadap isu "Kabur Aja Dulu" serta membandingkan performa beberapa algoritma machine learning dalam mengklasifikasikan sentimen tersebut. Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka kerja SEMMA (Sample, Explore, Modify, Model, Assess). Data dikumpulkan melalui teknik crawling menggunakan tools Tweet Harvest dan diproses melalui tahapan preprocessing seperti pembersihan data, normalisasi, tokenisasi, dan stemming. Pelabelan sentimen dilakukan dengan pendekatan lexicon-based menggunakan InSet Lexicon, sementara proses ekstraksi fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Untuk mengatasi ketidakseimbangan kelas, diterapkan metode Synthetic Minority Oversampling Technique (SMOTE). Enam algoritma diuji, yaitu Support Vector Machine (SVM), Logistic Regression, Naive Bayes, Random Forest, Decision Tree, dan K-Nearest Neighbors (KNN). Hasil penelitian menunjukkan bahwa mayoritas sentimen bersifat negatif. Algoritma SVM menghasilkan akurasi terbaik sebesar 80%, diikuti oleh Logistic Regression (77%) dan Random Forest (75%). Temuan ini menunjukkan bahwa SVM paling efektif digunakan dalam klasifikasi sentimen terkait isu sosial, dan dapat menjadi acuan dalam pengembangan sistem pemantauan opini publik.

Kata kunci: analisis sentimen, Kabur Aja Dulu, media sosial X, machine learning, SEMMA

ABSTRACT

This research aims to analyze the sentiment of X (Twitter) users regarding the "Kabur Aja Dulu" issue and to compare the performance of several machine learning algorithms in classifying these sentiments. A quantitative approach was employed using the SEMMA (Sample, Explore, Modify, Model, Assess) framework. The data were collected through a crawling technique using the Tweet Harvest tool and processed through several preprocessing stages, including data cleaning, normalization, tokenization, stopword removal, and stemming. Sentiment labelling was conducted using a lexicon-based approach with the InSet Lexicon, and feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF). To address class imbalance in the dataset, the Synthetic Minority Oversampling Technique (SMOTE) was applied. Six machine learning algorithms were evaluated: Support Vector Machine (SVM), Logistic Regression, Naive Bayes, Random Forest, Decision Tree, and K-Nearest Neighbors (KNN). The results showed that the majority of the sentiments were negative. The SVM algorithm achieved the highest accuracy at 80%, followed by Logistic Regression (77%) and Random Forest (75%).

Keywords: sentiment analysis, Kabur Aja Dulu, social media X, machine learning, SEMMA

PENDAHULUAN

Era digital saat ini ditandai dengan perkembangan informasi yang sangat cepat dan mudah diakses melalui

berbagai platform media sosial. Aplikasi X (sebelumnya Twitter) menjadi salah satu media sosial utama tempat masyarakat menyampaikan opini dan pendapat mereka. Opini-opini ini dapat berupa

pujian, kritik, hingga ujaran kebencian atau informasi palsu (hoax) yang berpotensi memicu perdebatan di media sosial (Adinda Salsabila, 2023). Riset yang dilakukan oleh We Are Social pada Februari 2025 menunjukkan bahwa pengguna media sosial di Indonesia mencapai 143 juta jiwa dari total populasi 285 juta jiwa. Secara spesifik, pengguna media sosial X di Indonesia tercatat mencapai 25,2 juta jiwa dan menduduki peringkat ke-7 sebagai platform yang paling sering digunakan. Pengguna X sering berbagi pesan singkat yang dikenal sebagai tweet, yang dapat diakses oleh pengguna lain. Tweet yang merupakan hasil dari penyaluran opini dan komentar, merupakan resource yang dapat digunakan untuk menganalisis sentimen (Fikri & Sabrila, 2020). Media sosial X selalu memperlihatkan trending topik yang dapat memudahkan penggunaanya dalam melihat topik baik di dalam negeri maupun di luar negeri yang sedang ramai diperbincangkan sehingga penggunaanya merasa up to date (Aulia Girmnafa & Susilo, 2022).

Salah satu topik yang belakangan ini ramai diperbincangkan di platform X adalah isu “Kabur Aja Dulu”. Tren ini mencerminkan kekecewaan masyarakat, khususnya generasi muda Indonesia, terhadap kondisi sosial, ekonomi, dan keadilan di dalam negeri. Faktor-faktor seperti ketidakstabilan ekonomi, tingginya angka pengangguran dibandingkan peluang kerja, serta gaji yang tidak sepadan dengan jam kerja menjadi pemicu viralnya isu ini. Fenomena “Kabur Aja Dulu” ini juga merupakan ekspresi

keputusasaan, kemarahan, dan bentuk protes terhadap kebijakan pemerintah yang dinilai tidak berpihak kepada rakyat, terutama terkait alokasi anggaran yang memprioritaskan program tertentu sementara pendidikan dan kesehatan menjadi prioritas pendukung, dapat terlihat pada Gambar 1 arah kebijakan BPP Tahun Anggaran 2026.



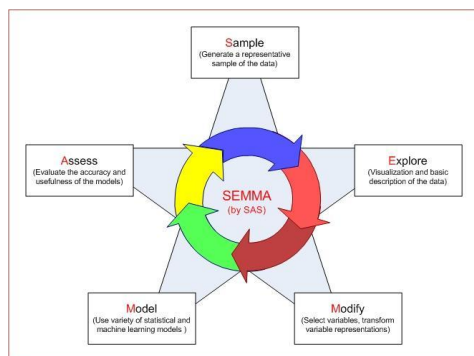
Gambar 1. Arah Kebijakan BPP Tahun Anggaran 2026

Analisis sentimen, sebagai cabang dari *Natural Language Processing* (NLP), adalah metode yang efektif untuk mengekstraksi dan mengidentifikasi emosi, opini, dan sikap dari teks (Fathur Rahman et al., 2024). Analisis sentimen juga bisa digunakan sebagai mengekspresikan keadaan emosional seperti kegembiraan, kesedihan, dan kemarahan (Asri et al., 2022). Tujuan yang paling utama yaitu dengan fokus pada pengelompokan sentimen sebagai negatif, positif, atau netral (Herjanto & Carudin, 2024). Penelitian ini menerapkan kerangka kerja SEMMA (*Sample, Explore, Modify, Model, Assess*) untuk menganalisis sentimen publik terhadap tren Kabur Aja Dulu, serta membandingkan kinerja enam algoritma *machine learning*, yaitu SVM, *Logistic*

Regression, Naive Bayes, Random Forest, Decision Tree, dan K-Nearest Neighbors (KNN).

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksploratif berbasis kerangka kerja SEMMA (*Sample, Explore, Modify, Model, Assess*) seperti yang ditunjukkan pada Gambar 2. Data dikumpulkan melalui teknik *crawling* menggunakan *tools Tweet Harvest* berbasis *Node.js* dan *auth token* yang terhubung ke media sosial X. Metode ini memfasilitasi pengumpulan data dari berbagai sumber secara sistematis dan efisien, sehingga data tersebut bisa digunakan untuk keperluan seperti penelitian, analisis, maupun pengembangan aplikasi (Algifari Rismawan & Syahidin, 2023). Data yang diambil berupa tweet berbahasa Indonesia yang mengandung kata kunci “Kabur Aja Dulu” dan variasinya, dalam rentang waktu 14–19 Februari 2025.



Gambar 2. Tahapan Metode SEMMA

HASIL DAN PEMBAHASAN

Pada bagian ini menyajikan hasil-hasil penelitian secara objektif dan

Proses preprocessing data meliputi beberapa tahapan, yaitu data *cleaning*, *case folding*, normalisasi, *tokenization*, *stopword removal*, dan *stemming*. Preprocessing data dilakukan untuk membersihkan data agar data tersebut menjadi lebih bersih dan terstruktur sehingga lebih efisien dan efektif untuk dilakukan analisis lebih lanjut (Khairani et al., 2024). Untuk pelabelan sentimen, digunakan pendekatan *lexicon-based* dengan menggunakan kamus *InSet Lexicon*. Pendekatan Lexicon Based digunakan untuk pengklasifikasi sentimen dari setiap opini, sehingga suatu kalimat sentimen dapat di kelompokkan sesuai kategori sentimen negatif dan positif (Mahendrajaya et al., 2019).

Selanjutnya, dilakukan proses *feature extraction* menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengubah data teks menjadi representasi numerik. Untuk mengatasi ketidakseimbangan kelas dalam data, diterapkan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Tahap pemodelan dilakukan dengan menerapkan enam algoritma *machine learning*, yaitu *Support Vector Machine* (SVM), *Logistic Regression*, *Naive Bayes*, *Random Forest*, *Decision Tree*, dan *K-Nearest Neighbors* (KNN). Evaluasi performa model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*, berdasarkan *confusion matrix*.

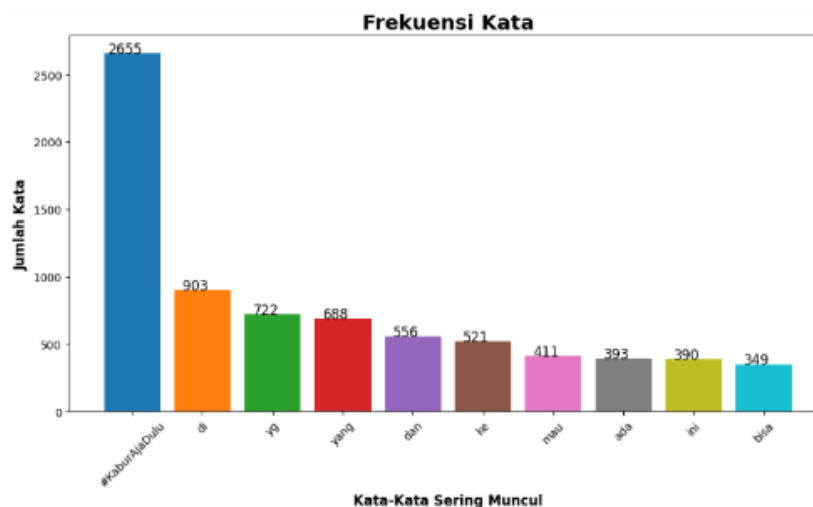
pembahasan ilmiah terkait klasifikasi sentimen pengguna X (Twitter) terhadap isu "Kabur Aja Dulu".

Sample

Pada Tahap *sample* berhasil mengumpulkan data tweet dari media sosial X. Pengambilan data dilakukan menggunakan *tweet harvest*, sebuah *tools* yang dibangun dengan *Node.js* dan memerlukan *twitter auth token* untuk mengakses *API Twitter*. Langkah-langkah *crawling* data dimulai dengan mendapatkan *twitter auth token* melalui inspeksi halaman X, kemudian menghubungkan Google Colab dengan *auth token* tersebut, dan menginstal *Pandas* serta *Node.js*. Proses *crawling* dilakukan dengan kata kunci "#KaburAjaDulu" berbahasa Indonesia, menghasilkan 3.081 data yang disimpan dalam file CSV.

Explore

Setelah proses *crawling*, data mentah yang terkumpul memiliki 15 kolom, Untuk mempermudah analisis, penelitian ini memfokuskan pada atribut *full_text* saja. Data kemudian dibersihkan dari duplikasi pada kolom *full_text*, mengurangi jumlah data dari 3.081 menjadi 3.075. Sebelum preprocessing, visualisasi wordcloud dan frekuensi kata dilakukan untuk melihat kata-kata yang paling sering muncul. Pada Gambar 3 dapat dilihat hasil visualisasi menunjukkan kata "#KaburAjaDulu" menjadi yang paling sering muncul dengan 2.655 kemunculan.



Gambar 3. Visualisasi Frekuensi Kata

Modify

Setelah proses *crawling*, data mentah yang terkumpul memiliki 15 kolom, Untuk mempermudah analisis, penelitian ini memfokuskan pada atribut *full_text* saja. Data kemudian dibersihkan dari duplikasi pada kolom *full_text*,

mengurangi jumlah data dari 3.081 menjadi 3.075. Sebelum preprocessing, visualisasi wordcloud dan frekuensi kata dilakukan untuk melihat kata-kata yang paling sering muncul. Pada Gambar 3 dapat dilihat hasil visualisasi menunjukkan kata "#KaburAjaDulu"

menjadi yang paling sering muncul dengan 2.655 kemunculan.

Pada tahap *modify* melibatkan serangkaian proses *preprocessing* untuk mempersiapkan data agar mudah dipahami dan dikenali oleh komputer. Tahapan ini meliputi:

1. Data *Cleaning*, hasilnya terlihat pada Tabel 1, proses untuk menghapus data yang tidak bermakna dan tidak relevan, seperti hashtag, retweet, mention, link, emoji, dan karakter khusus.

Tabel 1. Contoh Hasil Data *Cleaning*

Sebelum	Sesudah
Ideologi Pancasila para menteri saya ragukan nih @bahlillahadalia dan menteri lainnya... #KaburAjaDulu #KaburDuluAja #KaburSajaDulu #Indonesia	Ideologi Pancasila para menteri saya ragukan nih dan menteri lainnya KaburAjaDulu KaburDuluAja KaburSajaDulu Indonesia

2. *Case Folding*, semua huruf dalam teks diubah menjadi huruf kecil yang disebut *lower case* untuk menyamakan kata dan mencegah perbedaan makna, prosesnya terlihat pada Tabel 2.

Tabel 2. Contoh Hasil Dari *Case Folding*

Sebelum	Sesudah
Ideologi Pancasila para menteri saya ragukan nih dan menteri lainnya KaburAjaDulu KaburDuluAja KaburSajaDulu Indonesia	ideologi pancasila para menteri saya ragukan nih dan menteri lainnya kaburajadulu kaburduluaja kabursajadulu indonesia

3. Normalisasi, pada Gambar 3 ditampilkan kata-kata tidak baku, singkatan, atau tidak standar diubah

menjadi kata baku dan standar. Kamus kata baku diambil dari Kaggle.

Tabel 3. Contoh Hasil Dari Normalisasi

Sebelum	Sesudah
oya ga smua yg mau ke ln itu kaburajadulu ig kesana dalam posisi gada kerjaan dah dapet dr sini baru berangkat	oya tidak semua yang mau ke ln itu kaburajadulu juga kesana dalam posisi tidak ada pekerjaan sudah dapat dari sini baru berangkat

4. *Tokenization*, teks dibagi menjadi bagian-bagian yang lebih kecil atau kata individual, yang disebut token, seperti terlihat pada Tabel 4.

Tabel 4. Contoh Hasil Dari *Tokenization*

Sebelum	Sesudah
ideologi pancasila para menteri saya ragukan nih dan menteri lainnya kaburajadulu kaburduluaja kabursajadulu indonesia	[ideologi, pancasila, para, menteri, saya, ragukan, nih, dan, menteri, lainnya, kaburajadulu, kaburduluaja, kabursajadulu, indonesia]

5. *Stopword Removal*, pada Tabel 5 terlihat kata-kata umum yang tidak signifikan dalam analisis (misalnya "dan", "di", "ke", "yang") dihapus dari teks.

Tabel 5. Contoh Hasil Dari Stopword Removal

Sebelum	Sesudah
[ideologi, pancasila, para , menteri, saya , ragukan, nih, dan , menteri, lainnya , kaburajadulu, kaburajadulu, kaburajadulu, Indonesia]	[ideologi, pancasila, menteri, ragukan, nih, menteri, kaburajadulu, kaburajadulu, kaburajadulu, Indonesia]

6. *Stemming*, terakhir pada Tabel 6 ditampilkan kata-kata yang diubah menjadi bentuk dasarnya dengan menghilangkan imbuhan (awalan, sisipan, atau akhiran). Setelah

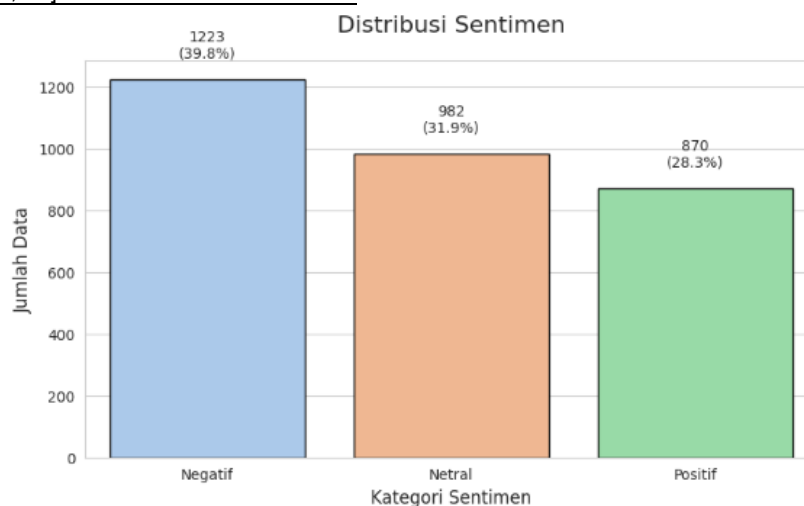
melakukan tahap *preprocessing*, visualisasi *wordcloud* dan frekuensi kata kembali dilakukan. Kata "kaburajadulu" masih menjadi yang paling sering muncul, dengan 3.122 kemunculan. Data yang bernilai kosong (NaN) juga dihapus.

Tabel 6. Contoh Hasil Dari *Stemming*

Sebelum	Sesudah
[gue, voters, bilang, voters, sadarkan , bantu, mengkritik , pemerintahan , sadar, bantu, kritik, join, kaburajadulu, kaitain, cuk, wtf]	gue voters bilang voters sadar bantu kritik perintah sadar bantu kritik join kaburajadulu kaitain cuk wtf

Model

Tahap pemodelan dimulai dengan pelabelan data menggunakan pendekatan *lexicon based*. Kamus *lexicon* positif (3.610 kata) dan negatif (6.610 kata) diperoleh dari *GitHub* akun Fajri91, dengan nilai bobot -5 hingga +5 (Koto & Rahmanningtyas, 2017) Setiap kata dalam tweet dicocokkan dengan kamus ini, dan total skor sentimen menentukan label positif, negatif, atau netral.



Gambar 4. Hasil Sentimen Kabur Aja Dulu

Dari Gambar 4 diatas dapat terlihat hasil pelabelan sentimen menunjukkan distribusi sebagai berikut: 870 kata positif (28.3%), 1.223 kata negatif (39.8%), dan 982 kata netral (31.9%). Dari hasil ini, dapat disimpulkan bahwa sentimen pengguna media sosial X terkait isu "#KaburAjaDulu" antara tanggal 14-19 Februari 2025 cenderung negatif. Hal ini mencerminkan kekecewaan masyarakat, khususnya generasi muda, terhadap

kondisi ekonomi, sosial, dan keadilan di Indonesia, serta tingginya tingkat pengangguran yang tidak diimbangi dengan solusi pemerintah. Visualisasi *wordcloud* dan frekuensi kata untuk masing-masing sentimen (positif, negatif, netral) juga ditampilkan, menunjukkan kata "kaburajadulu" mendominasi di ketiga kategori.

Data kemudian diubah ke format numerik menggunakan TF-IDF (*Term Frequency-Inverse Document*

Frequency). Setelah itu, enam algoritma *machine learning* diuji untuk klasifikasi: *Logistic Regression*, *Decision Tree*, *Random Forest*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Naive Bayes*.

1. Klasifikasi Sebelum SMOTE

Pada pengujian awal menunjukkan akurasi model-model sebelum penerapan SMOTE:

- *Logistic Regression*: 0.69%
- *Decision Tree*: 0.56%
- *Random Forest*: 0.66%
- *Support Vector Machine*: 0.67%
- *K-Nearest Neighbors*: 0.49%
- *Naïve Bayes*: 0.58%

2. SMOTE (*Synthetic Minority Oversampling Technique*)

Berfungsi untuk mengatasi ketidakseimbangan kelas pada dataset, SMOTE menciptakan data sintetis dari kelas minoritas untuk meningkatkan performa klasifikasi. Setelah SMOTE, distribusi kelas menjadi seimbang, dengan masing-masing kelas (Negatif, Positif, Netral) memiliki 1223 data, seperti ditunjukkan pada Gambar 5.

```
Distribusi kelas sebelum SMOTE:
label
Negatif    1223
Netral      982
Positif      870
Name: count, dtype: int64

SMOTE berhasil diterapkan.
Dimensi X setelah SMOTE: (3669, 7325)
Dimensi y setelah SMOTE: (3669,)

Distribusi kelas setelah SMOTE:
label
Negatif    1223
Positif     1223
Netral      1223
Name: count, dtype: int64
```

Gambar 5. Setelah Menerapkan SMOTE

Assess

Tahap assess mengevaluasi performa keenam metode *machine learning* setelah penerapan SMOTE, menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* yang dihitung dari *confusion matrix*.

Klasifikasi Sesudah SMOTE

- *Logistic Regression*, Akurasi mencapai 0.77%. Dari *confusion matrix*, model berhasil mengklasifikasikan 191 data negatif dengan benar, 182 data netral dengan benar, dan 199 data positif dengan benar.
- *Decision Tree*, Akurasi sebesar 0.64%. Model berhasil mengklasifikasikan 146 data negatif dengan benar, 157 data netral dengan benar, dan 171 data positif dengan benar.
- *Random Forest*, Akurasi sebesar 0.75%. Meskipun akurasi yang dilaporkan di teks 0.66%, visualisasi setelah menerapkan SMOTE menunjukkan 0.7561. Model berhasil mengklasifikasikan 171 data negatif dengan benar, 194 data netral dengan benar, dan 190 data positif dengan benar.
- *Support Vector Machine*, Akurasi tertinggi, mencapai 0.80%. Model berhasil mengklasifikasikan 208 data negatif dengan benar, 185 data netral dengan benar, dan 197 data positif dengan benar.
- *K-Nearest Neighbors*: Akurasi menurun menjadi 0.41% setelah

menerapkan *SMOTE*. *Confusion matrix* menunjukkan model sangat mendominasi dalam mengklasifikasikan kelas netral (239 data benar), tetapi sangat buruk dalam membedakan kelas negatif (hanya 13 data benar) dan positif (hanya 56 data benar), yang mengindikasikan bias terhadap kelas netral.

- Naive Bayes, Akurasi mencapai 0.73%. Model berhasil mengklasifikasikan 186 data negatif dengan benar, 145 data netral dengan benar, dan 206 data positif dengan benar.

Perbandingan akurasi secara keseluruhan setelah *SMOTE* menunjukkan bahwa *Support Vector Machine* memiliki performa terbaik (0.80%), diikuti *Logistic Regression* (0.77%), dan *Random Forest* (0.75%). Peningkatan akurasi setelah menerapkan *SMOTE* terlihat pada *Logistic Regression*, *Decision Tree*, *Random Forest*, *Support Vector Machine*, dan *Naive Bayes*, sementara *KNN* mengalami penurunan.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai Komparasi Beberapa Algoritma *Machine Learning* untuk Klasifikasi Sentimen Pengguna X (Twitter) Terkait Isu Kabur Aja Dulu, didapatkan beberapa kesimpulan sebagai berikut:

- Hasil mengumpulkan sebanyak 3.081 data terkait isu "Kabur Aja Dulu". Setelah melakukan analisis sentimen

menggunakan pendekatan *lexicon-based*, distribusi sentimen menunjukkan 1.223 data bersifat negatif, 982 data bersifat netral, dan 870 data bersifat positif. Dengan demikian, dapat disimpulkan bahwa sentimen pengguna X mayoritas tidak setuju dengan opini "Kabur Aja Dulu".

- Penerapan enam algoritma *machine learning* *Support Vector Machine*, *Logistic Regression*, *Naive Bayes*, *Random Forest*, *Decision Tree*, dan *K-Nearest Neighbors* untuk klasifikasi sentimen menunjukkan performa yang bervariasi setelah proses *SMOTE*. Algoritma dengan performa terbaik adalah *Support Vector Machine*, mencapai tingkat akurasi 80%. *Logistic Regression* menjadi algoritma terbaik kedua dengan akurasi 77% , diikuti oleh *Random Forest* dengan akurasi 75%. Sementara itu, *K-Nearest Neighbors* (*KNN*) menunjukkan akurasi terendah sebesar 40% setelah *SMOTE*, karena cenderung mendominasi dalam mengklasifikasikan kelas netral sehingga kurang efektif dalam membedakan kelas negatif dan positif.

DAFTAR RUJUKAN

- Adinda Salsabila, N. (2023). Analisis sentimen pada media sosial twitter terhadap tokoh gus dur menggunakan metode naïve bayes dan support vector machine (svm). *Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Gus*

- Dur Menggunakan Metode Naïve Bayes Dan Support Vector Machine (Svm). <https://repository.uinjkt.ac.id/dspace/handle/123456789/66253>
- Algifari Rismawan, S., & Syahidin, Y. (2023). Implementasi Website Berita Online Menggunakan Metode Crawling Data dengan Bahasa Pemrograman Python. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 10(3), 167–178. <https://doi.org/10.35957/JATISI.V10I3.4902>
- Asri, Y., Suliyanti, W. N., Kuswardani, D., & Fajri, M. (2022). Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile. *Petir*, 15(2), 264–275. <https://doi.org/10.33322/PETIR.V15I2.1733>
- Aulia Girnanfa, F., & Susilo, A. (2022). Studi Dramaturgi Pengelolaan Kesan Melalui Twitter Sebagai Sarana Eksistensi Diri Mahasiswa di Jakarta. *Journal of New Media and Communication*, 1(1), 58–73. <https://doi.org/10.55985/JNMC.V1I1.2>
- Fathur Rahman, I., Nur Hasanah, A., & Heryana, N. (2024). Analisis sentimen ulasan pengguna aplikasi samsat digital nasional (signal) dengan menggunakan metode naïve bayes classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(2), 2830–7062. <https://doi.org/10.23960/JITET.V12I2.4073>
- Fikri, M. I., & Sabrila, T. S. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL: STIKI Informatika Jurnal*, 10(02), 71–76. <https://doi.org/10.32664/SMATIKA.V10I02.455>
- Herjanto, M. F. Y., & Carudin, C. (2024). Analisis sentimen ulasan pengguna aplikasi sirekap pada play store menggunakan algoritma random forest classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(2). <https://doi.org/10.23960/JITET.V12I2.4192>
- Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi*

- Informasi Dan Ilmu Komputer*, 11(4), 887–894. <https://doi.org/10.25126/JTIK.1148315>
- Koto, F., & Rahmaningtyas, G. Y. (2017). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *Proceedings of the 2017 International Conference on Asian Language Processing, IALP* 2017, 2018-January, 391–394. <https://doi.org/10.1109/IALP.2017.8300625>
- Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). Analisis sentimen pengguna gopay menggunakan metode lexicon based dan support vector machine. *KOMPUTEK*, 3(2), 52–63. <https://doi.org/10.24269/JKT.V3I2.270>