

Implementasi Data Mining untuk Klasifikasi Komentar Hate Speech Menggunakan Algoritma Support Vector Machine (SVM)

Fathul Hakim Nesa^{*1}, Sucipto², Alda Cendekia Siregar³

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer,

Universitas Muhammadiyah Pontianak

Email: ^{*1}kimfathul@gmail.com, ²sucipto@unmuhpnk.ac.id, ³alda.siregar@unmuhpnk.ac.id

(Naskah masuk: 17 Desember 2024, diterima untuk diterbitkan: 10 Juli 2025)

Abstrak: Dalam era digital saat ini, media sosial telah menjadi sarana utama bagi individu untuk berkomunikasi, mengekspresikan pendapat, serta berbagi informasi secara cepat dan luas. Namun, kemudahan ini juga membawa tantangan serius, salah satunya adalah maraknya penyebaran ujaran kebencian (hate speech) yang dapat memicu konflik sosial, memengaruhi kesehatan mental, serta memperkuat diskriminasi terhadap kelompok tertentu. Penelitian ini bertujuan untuk mengimplementasikan teknik data mining guna mengklasifikasikan komentar yang mengandung hate speech dengan memanfaatkan algoritma Support Vector Machine (SVM). Metode yang digunakan meliputi beberapa tahapan, yaitu: perumusan masalah, pengumpulan dan pembersihan data, pembuatan model, evaluasi performa model, deployment, dan penarikan kesimpulan. Berdasarkan hasil pengujian, algoritma SVM mampu memberikan performa yang baik dengan nilai F1-score sebesar 89%, precision 89%, dan recall 89%. Aplikasi klasifikasi ini diharapkan dapat digunakan sebagai alat bantu dalam mengurangi penyebaran ujaran kebencian di platform media sosial dan turut menciptakan ekosistem digital yang lebih sehat, inklusif, dan positif bagi semua pengguna.

Kata Kunci – Ujaran Kebencian; Support Vector Machine; Data Mining; Machine Learning; Confusion Matrix

Implementation of Data Mining for Hate Speech Comment Classification Using Support Vector Machine (SVM) Algorithm

Abstract: In today's digital era, social media has become a primary platform for individuals to communicate, express opinions, and share information quickly and widely. However, this convenience also brings serious challenges, one of which is the widespread dissemination of hate speech, which can trigger social conflict, affect mental health, and reinforce discrimination against certain groups. This study aims to implement data mining techniques to classify comments containing hate speech by utilizing the Support Vector Machine (SVM) algorithm. The research method includes several stages: problem formulation, data collection and cleaning, model development, model performance evaluation, deployment, and conclusion drawing. Based on the testing results, the SVM algorithm demonstrates strong performance, achieving an F1-score of 89%, precision of 89%, and recall of 89%. The developed classification application is expected to serve as a tool to help reduce the spread of hate speech on social media platforms and contribute to the creation of a healthier, more inclusive, and positive digital ecosystem for all users.

Keywords – Hate Speech; Support Vector Machine; Data Mining; Machine Learning; Confusion Matrix

1. PENDAHULUAN

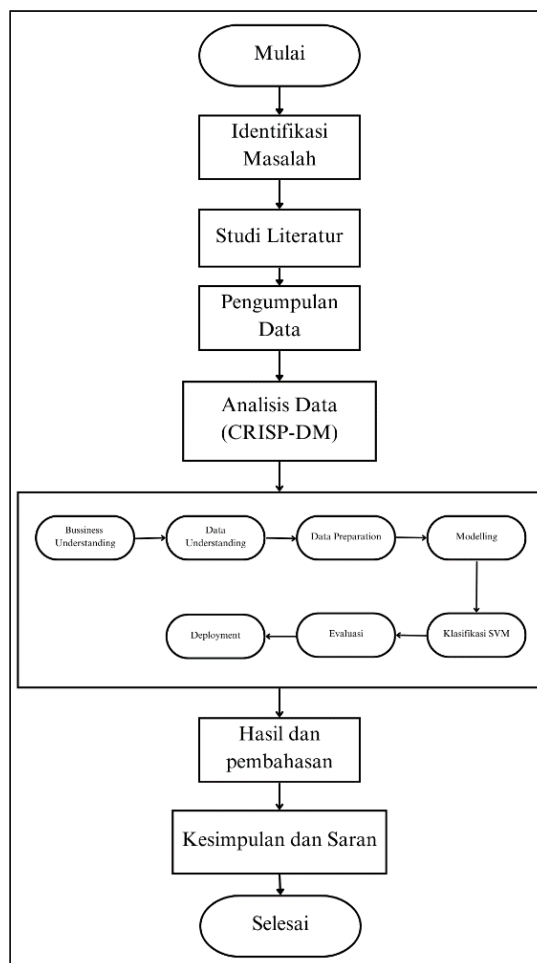
Dalam era digital, media sosial telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari. Platform seperti Twitter, Facebook, dan Instagram menyediakan ruang bagi jutaan pengguna untuk berkomunikasi, berbagi informasi, dan mengekspresikan pendapat. Namun, kebebasan dalam berinteraksi ini juga menimbulkan tantangan, salah satunya adalah penyebaran ujaran kebencian (hate speech). Menurut Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) serta Komisi Nasional Hak Asasi Manusia, ujaran kebencian adalah tindakan yang dilakukan secara sengaja untuk menyebarkan informasi dengan tujuan menimbulkan kebencian atau permusuhan terhadap individu atau kelompok tertentu berdasarkan suku, agama, dan ras[1].

Fenomena ujaran kebencian berdampak signifikan, termasuk meningkatnya ketegangan sosial, diskriminasi, dan dalam kasus ekstrem, memicu tindak kekerasan[1][2]. Pemerintah telah berupaya menangani isu ini melalui regulasi hukum, seperti Pasal 156 dan Pasal 157 KUHP serta Pasal 45 ayat (2) UU No. 19 Tahun 2016[2]. Meski demikian, tingginya volume dan sifat dinamis dari konten media sosial membuat upaya manual dalam mendeteksi ujaran kebencian menjadi tidak efisien. Pendekatan berbasis teknologi, seperti pembelajaran mesin, menawarkan solusi potensial. Support Vector Machine (SVM) menjadi salah satu algoritma yang menonjol dalam klasifikasi data teks karena kemampuannya dalam menangani data tidak terstruktur.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi berbasis SVM untuk mendeteksi komentar berisi ujaran kebencian. Dengan memanfaatkan data cuitan (tweets) berbahasa Indonesia, penelitian ini berfokus pada penerapan data mining untuk mengidentifikasi komentar yang tergolong hate speech atau non-hate speech. Evaluasi kinerja model dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall, dan f-measure. Penelitian ini diharapkan dapat memberikan kontribusi dalam menciptakan lingkungan media sosial yang lebih positif, mendukung penegakan hukum terkait ujaran kebencian, serta memperkaya wawasan dalam penerapan teknologi untuk menyelesaikan masalah sosial di era digital.

2. METODE PENELITIAN

Proses dalam penelitian ini dimulai dengan mengidentifikasi masalah dan berlanjut hingga tahap kesimpulan dan saran.



Gambar 1. Tahapan Penelitian

2.1 Identifikasi Masalah

Ini merupakan Langkah awal dalam merumuskan masalah, pada tahapan ini dilakukan pencarian suatu permasalahan dan dipelajari sehingga mendapatkan solusi. Permasalahan yang diambil yaitu bagaimana implementasi data mining dalam klasifikasi komentar *hate speech* menggunakan algoritma SVM.

2.2 Studi Literatur

Tahapan ini adalah melakukan studi literatur yang relevan tentang penelitian sebelumnya yang menggunakan algoritma *Support Vector Machine*. Tujuan dari penelitian ini adalah untuk memperkuat teori yang sudah ada dengan melihat bagaimana algoritma *Support Vector Machine* berfungsi untuk mengklasifikasikan komentar *hate speech*. Dalam contoh ini, peneliti bermaksud untuk mempelajari pola dan cara klasifikasi dilakukan dalam data mining.

2.3 Pengumpulan Data

Pada tahapan ini penulis mengumpulkan data untuk keperluan penelitian. Data dikumpulkan melalui *web scrapping* menggunakan *tweet harvest* pada aplikasi twitter dengan berbagai macam topik. Topik dipilih berdasarkan kasus kasus yang sedang viral di Indonesia pada saat pengumpulan data. Data yang didapatkan berjumlah 2.549 data. Data tersebut disimpan dalam format csv dengan 15 atribut/fitur yaitu *conversation_id_str*, *created_at*, *favorite_count*, *full_text*, *id_str*, *image_url*, *in_reply_to_screen_name*, *lang*, *location*, *quote_count*, *reply_count*, *retweet_count*, *tweet_url*, *user_id_str*, dan *username*. Berikut adalah data komentar atau cuitan yang didapatkan penulis:

Tabel 1. Data Komentar atau cuitan Twitter

conversation_id_str	180090768631 9636709	18013225926217 07311	1800718896208 724112	
created_at	Thu Jun 13 18:36:22 +0000 2024	Thu Jun 13 18:35:47 +0000 2024	Thu Jun 13 18:34:17 +0000 2024	
favorite_count	1	0	0	
full_text	@zoelfick Yg aneh emang panitianya dah serius. Tp mau gmn lagi netizen ga kenal panitianya cuma kenal anang ashanty jadi bulan ² an netizen deh mereka. Wkwk	Kasian dah anang panitia e guoblok cok pasti gapernah ikut organisasi wkwk bercandaa	SiaranBolaLive Ni orang apa gak nonton Anang nyanyi kali yak...	
id_str	180132273891 9072107	18013225926217 07311	1801322214207 406151	
image_url	NaN	NaN	NaN	
in_reply_to_screen_name	zoelfick	NaN	SiaranBolaLive	
lang	in	in	in	
location	Panceng, Indonesia	NaN	NaN	

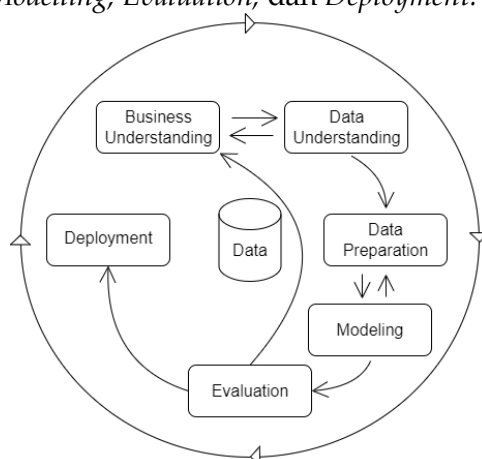
quote_count	0	0	0	
reply_count	0	0	0	
retweet_count,	0	0	0	
tweet_url	https://x.com/ZakkyAbdil/status/1801322738919072107	https://x.com/BizDaily/status/1801322592621707311	https://x.com/pelawananarus/status/1801322214207406151	
user_id_str	3191199303	3655651754	1601574118834720776	
username	ZakkyAbdil	BizDaily	pelawananarus	

Tabel 2. Penjelasan Fitur Pada Data

FITUR	KETERANGAN
conversation_id_str	ID percakapan untuk mengelompokkan tweet dalam satu thread
created_at	Tanggal dan waktu tweet dibuat
favorite_count	Jumlah 'likes' pada tweet.
full_text	Teks lengkap dari tweet
id_str	ID unik tweet
image_url	URL gambar dalam tweet (jika ada)
in_reply_to_screen_name	Username yang dituju jika tweet adalah balasan
lang	Kode bahasa tweet
location	Lokasi pengguna di profil mereka
quote_count	Jumlah tweet yang mengutip tweet ini
reply_count	Jumlah balasan yang diterima
retweet_count,	Jumlah retweet yang diterima
tweet_url	URL menuju tweet ini
user_id_str	ID unik pengguna yang memposting tweet
username	Handle Twitter (@username) dari pengguna yang memposting

2.4 Analisis Data

Penelitian ini menggunakan tahapan model *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk analisis data. Tahapan-tahapan ini meliputi *Busssiness Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*.



Gambar 1. Tahapan CRISP-DM

2.4.1 Pemahaman Bisnis (*Business Understanding*)

Tahap pemahaman bisnis adalah tahap di mana peneliti menganalisis untuk memahami kebutuhan dan tujuan bisnis yang akan diterapkan. Tahap ini merupakan langkah awal yang mengarahkan penelitian ke tujuannya. Penelitian akan melakukan klasifikasi komentar *hate speech* menggunakan algoritma SVM dan mengetahui tingkat akurasi yang dihasilkan dari algoritma tersebut.

2.4.2 Pemahaman Data (*Data Understanding*)

Pada tahap pemahaman data ini hal yang dilakukan antara lain memahami data yang akan digunakan, mengumpulkan data yang diperlukan, dan melakukan analisis terhadap data yang digunakan. Data yang digunakan yaitu data yang diperoleh dari hasil scraping pada sosial media twitter menggunakan *tweet harvest*.

2.4.3 Persiapan Data (*Data Preparation*)

Tahap persiapan data dilakukan untuk memperbaiki kondisi data yang tidak ideal sebelum data masuk ke tahap pemodelan, sehingga masalah yang ada dapat diatasi dan model yang dihasilkan menjadi lebih baik. Data harus dilakukan proses *preprocessing/cleaning* dan seleksi fitur sebelum digunakan agar data tetap valid dan akan berpengaruh terhadap hasil yang diperoleh. Adapun tahapan yang dilakukan yaitu:

- a. *Cleaning*
Proses ini adalah membersihkan data yang tidak memiliki pengaruh pada proses data mining.
- b. *Case Folding*
Proses merubah semua huruf menjadi huruf kecil dan pada proses ini juga menghapus karakter, simbol, URL, dan angka yang tidak digunakan dalam dataset.
- c. *Tokenizing*
Proses ini memecah kalimat menjadi kata, sehingga terbentuklah unit unit terkecil atau yang disebut token. Contohnya seperti ["Kami", "tidak", "datang", "hari", "ini,"].
- d. *Word Normalization*
Proses merubah kata yang tidak baku menjadi baku contohnya seperti penyingkatan kata "yg" menjadi yang dan "ok" menjadi oke.
- e. *Stopword removal*
Proses menghapus *stopword* atau kata hubung contohnya seperti "dan", "di", dan lain-lain.
- f. *Stemming*
Proses ini mengubah kata menjadi bentuk dasar contohnya seperti "ikuti" menjadi ikut.
- g. *Data Labelling*
Pada tahap ini yaitu pemberian label atau kategori pada data. Label ini membantu algoritma SVM memahami dan mengkategorikan data dengan benar. Data akan diberi label HS (*hate speech*) dan Non HS (*non hate speech*).
- h. *Encoding Target Value*
Proses ini digunakan untuk mengubah data dengan label kategorikal menjadi numerikal. Hal ini dikarenakan mesin hanya bisa membaca data angka.
- i. *Feature Extraction (TF-IDF)*
Tahapan ini adalah proses memberikan bobot pada setiap kata agar proses klasifikasi dapat diterapkan. Metode yang digunakan untuk pembobotan kata adalah *Term Frequency – Inverse Document Frequency* (TF-IDF). TF digunakan untuk menghitung nilai kata yang muncul dalam suatu dokumen, sedangkan IDF digunakan untuk menghitung nilai kata yang muncul dalam keseluruhan dokumen. Nilai TF berbanding terbalik dengan IDF, semakin sering kata muncul, semakin kecil nilai IDF. Perhitungan bobot kata dalam sebuah dokumen menggunakan TF-IDF dilakukan dengan menghitung nilai TF dan IDF untuk setiap kata.

2.4.4 Pemodelan (*Modelling*)

Di tahap pemodelan ini, langkah yang akan dilakukan adalah penerapan algoritma SVM untuk mengklasifikasi data. Data yang telah melewati tahap *pre processing* dibagi menjadi data *training* 80% dan data *testing* 20%.

2.4.5 Evaluasi (Evaluation)

Selanjutnya, tahap evaluasi dilakukan dengan menguji dan menilai model yang telah dihasilkan untuk menentukan keakuratan dan generalitasnya. Performa klasifikasi dari algoritma SVM diukur menggunakan *confusion matrix* untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Keakuratan ditentukan dengan acuan sebagai berikut :

1. *True Positive* (TP) adalah data komentar *non hate speech* yang diidentifikasi dengan benar oleh sistem
2. *True Negative* (TN) adalah data komentar *hate speech* yang diidentifikasi dengan benar oleh sistem
3. *False Positive* (FP) adalah data komentar *non hate speech* yang diklasifikasi secara salah oleh sistem
4. *False Negative* (FN) adalah data komentar *hate speech* yang diklasifikasi secara salah oleh sistem.

2.4.6 Penyebaran (Deployment)

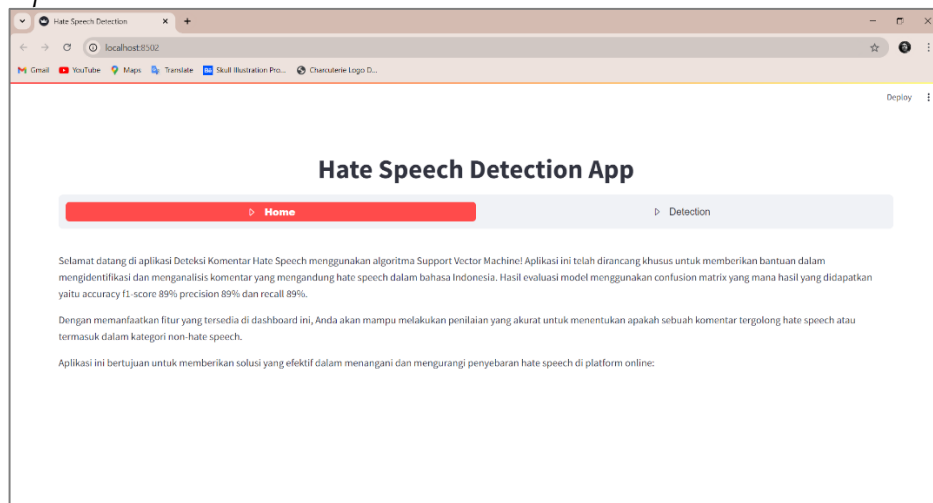
Deployment adalah tahap implementasi. Setelah keseluruhan model selesai dibangun pada visual studio code, kemudian sistem akan diimplementasikan dalam bentuk situs web. Hasil pemodelan akan disimpan kedalam bentuk file *.pkl* menggunakan modul *Pickle*. *Framework* yang digunakan untuk membuat aplikasi web ini adalah *Streamlit*. Implementasi ini akan dibuat sedemikian rupa sehingga pengguna dapat dengan mudah mengakses dan menggunakan model klasifikasi komentar *hate speech*.

3. HASIL DAN PEMBAHASAN

Pada aplikasi yang dibangun bahasa yang digunakan yaitu *python* dan *framework* yang digunakan yaitu *streamlit*.

3.1 Hasil Halaman Beranda

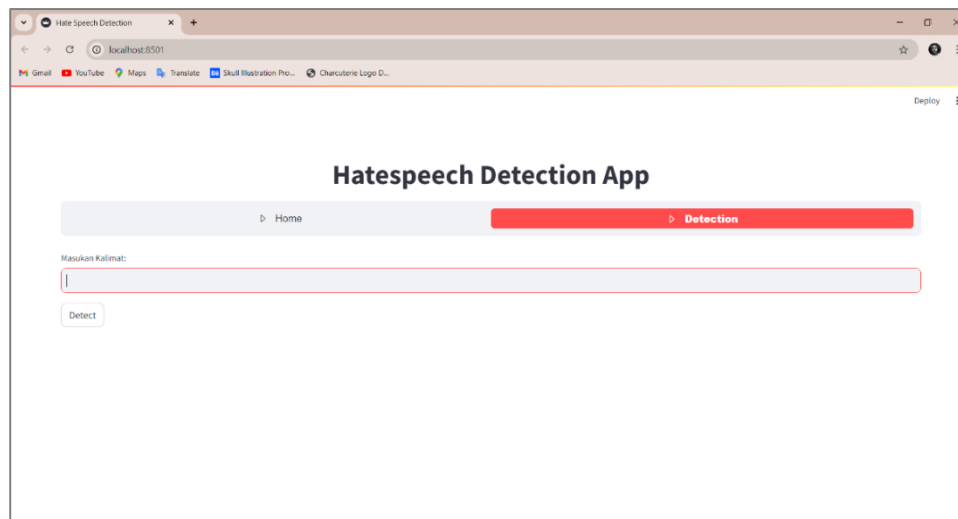
Halaman Home merupakan halaman yang berisikan informasi mengenai aplikasi deteksi komentar *hate speech*.



Gambar 3. Hasil Halaman Home

3.2 Hasil Halaman Deteksi

Pada halaman ini adalah halaman untuk mendeteksi komentar. Dengan menginput komentar kemudian hasil inputan akan diproses oleh model yang telah kita bangun yang akan menghasilkan output hasil deteksi. Hasil dari deteksi tersebut yaitu *hate speech* atau *non hate speech*



Gambar 4. Hasil Halaman Deteksi

4. KESIMPULAN

Dari hasil penelitian diatas dapat di ambil kesimpulan berdasarkan rumusan masalah sebagai berikut, berdasarkan hasil yang dilakukan dalam penelitian ini dapat disimpulkan bahwa dapat dilakukan untuk mengimplementasikan data mining untuk klasifikasi komentar hate speech menggunakan algoritma Support Vector Machine (SVM) dan aplikasi yang dibangun dalam mendeteksi komentar hate speech juga dapat digunakan dengan baik. Dari hasil penelitian ini juga disimpulkan bahwa implementasi data mining untuk klasifikasi komentar hate speech menggunakan algoritma Support Vector Machine didapatkan hasil yaitu accuracy f1-score 89% precision 89% dan recall 89%.

UCAPAN TERIMA KASIH

Terimakasih kepada Allah swt, kedua orangtua, keluarga, dan teman-teman yang sudah memberikan support kepada saya untuk menyelesaikan skripsi dan jurnal.

DAFTAR PUSTAKA

- [1] Y. Septiawan and C. Chairani, "Perbandingan Akurasi Metode Deteksi Ujaran Kebencian dalam Postingan Twitter Menggunakan Metode SVM dan Decision Trees yang Dioptimalkan dengan Adaboost," *Teknika*, vol. 17, no. 2, pp. 287 – 299–287 – 299, 2023.
- [2] M. A. Jibrán, A. Eviyanti, and Y. Findawati, "Deteksi Ujaran Kebencian Menggunakan Metode Support Vector Machine (SVM)," *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 4, no. 4, pp. 890–900, 2023.
- [3] R. Triyana, O. Virgantara Putra, and F. R. Pradhana, "Deteksi Cyberbullying Pada Tweet Berbahasa Inggris Dengan Metode Support Vector Machine," *Seminar Nasional Hasil Penelitian & Pengabdian Masyarakat Bidang Ilmu Komputer*, pp. 98–103, 2022.
- [4] F. Abdusyukur, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter," *Komputa: Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 73–82, 2023, doi: [10.34010/komputa.v12i1.9418](https://doi.org/10.34010/komputa.v12i1.9418).
- [5] L. Mutawalli, M. T. A. Zaen, and W. Bagye, "KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 2, no. 2, p. 43, 2019, doi: [10.36595/jire.v2i2.117](https://doi.org/10.36595/jire.v2i2.117).

- [6] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 11, pp. 4704–4713, 2018.
- [7] M. S. Mujib Ridwan, Hadi Suyono, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, vol. 4, no. 2, p. 151, 2018, doi: [10.24076/citec.2017v4i2.106](https://doi.org/10.24076/citec.2017v4i2.106).
- [8] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Informatika*, vol. 2, no. 2, pp. 213–219, 2017, doi: [10.22202/ei.2016.v2i2.1465](https://doi.org/10.22202/ei.2016.v2i2.1465).
- [9] B. Domas Hidayat *et al.*, "UJARAN KEBENCIAN PADA MEDIA SOSIAL PADA SAAT PANDEMI COVID-19 STUDI KASUS PUTUSAN No.72/PID.SUS/2020/PT.DPS", [Online]. Available: www.icjr.co.id
- [10] A. Fikriani, I. Asror, and Y. R. Murti, "Klasifikasi Kepribadian Berdasarkan Data Twitter dengan Menggunakan Metode Support Vector Machine," *e-Proceeding of Engineering*, vol. 6, no. 3, pp. 10436–10450, 2019.
- [11] Widi Hastomo, Nur Aini, Adhitio Satyo Bayangkari Karno, and L.M. Rasdi Rere, "Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 11, no. 2, pp. 131–139, 2022, doi: [10.22146/jnteti.v11i2.2586](https://doi.org/10.22146/jnteti.v11i2.2586).
- [12] N. A. Ramdhan and D. A. Nufriana, "Rancang Bangun Dan Implementasi Sistem Informasi Skripsi Oline Berbasis WEB," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 1, no. 02, pp. 1–12, 2019, doi: [10.46772/intech.v1i02.75](https://doi.org/10.46772/intech.v1i02.75).
- [13] M. K. Rahmadhika and A. M. Thantawi, "Rancang Bangun Aplikasi Face Recognition Pada Pendekatan CRM Menggunakan Opencv Dan Algoritma Haarcascade," *IKRA-ITH INFORMATIKA: Jurnal Komputer dan Informatika*, vol. 5, no. 1, pp. 109–118, 2021.