https://journaledutech.com/index.php/great
Global Research and Innovation Journal (GREAT)
Volume 1, Nomor 3, 2025, Hal. 43-51
ISSN : 3090-3289

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP PENEGAKAN HUKUM DI INDONESIA TAHUN 2024 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

M.Rio Pratama Srg
Sistem Informasi, STMIK Kaputama, Binjai
E-mail: *riosiregar1929@gmail.com

ABSTRAK

Penegakan hukum merupakan aspek penting dalam menjaga keadilan dan ketertiban di Indonesia. Namun, praktiknya sering menimbulkan polemik di masyarakat, sehingga menjadi isu yang ramai diperbincangkan di media sosial, khususnya Twitter. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap penegakan hukum di Indonesia pada tahun 2024 dengan memanfaatkan data tweet berbahasa Indonesia. Data dikumpulkan menggunakan kata kunci yang relevan, kemudian melalui tahapan *preprocessing* teks seperti pembersihan data, tokenisasi, dan stopword removal. Fitur teks diekstraksi menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), sedangkan klasifikasi sentimen dilakukan dengan algoritma Support Vector Machine (SVM). Metodologi yang digunakan adalah pendekatan CRISP-DM untuk memandu alur penelitian mulai dari pemahaman data hingga evaluasi model. Hasil penelitian menunjukkan bahwa metode SVM mampu mengklasifikasikan sentimen positif, negatif, dan netral secara cukup akurat, sehingga dapat digunakan sebagai acuan dalam memahami persepsi publik terkait isu penegakan hukum. Temuan ini diharapkan dapat menjadi masukan bagi pihak terkait dalam mengevaluasi dan memperbaiki kualitas penegakan hukum di Indonesia.

Kata kunci

Analisis Sentimen, Twitter, Penegakan Hukum, Support Vector Machine (SVM), TF-IDF, CRISP-DM.

ABSTRACT

Law enforcement is a fundamental aspect in maintaining justice and social order in Indonesia. However, its implementation often raises public controversy, making it a widely discussed issue on social media, particularly Twitter. This study aims to analyze public sentiment toward law enforcement in Indonesia during 2024 using Indonesian-language tweets as the primary data source. The data were collected based on relevant keywords and then processed through several text preprocessing stages, including data cleaning, tokenization, and stopword removal. Text features were extracted using the Term Frequency–Inverse Document Frequency (TF-IDF) method, while sentiment classification was carried out using the Support Vector Machine (SVM) algorithm. The research employed the CRISP-DM framework to guide the process from data understanding to model evaluation. The results demonstrate that SVM is capable of classifying positive, negative, and neutral sentiments with satisfactory accuracy, providing valuable insights into public perceptions of law enforcement issues. These findings are expected to serve as input for policymakers and related stakeholders in evaluating and improving the quality of law enforcement in Indonesia.

Keywords

Sentiment Analysis, Twitter, Law Enforcement, Support Vector Machine (SVM), TF-IDF, CRISP-DM.

1. PENDAHULUAN

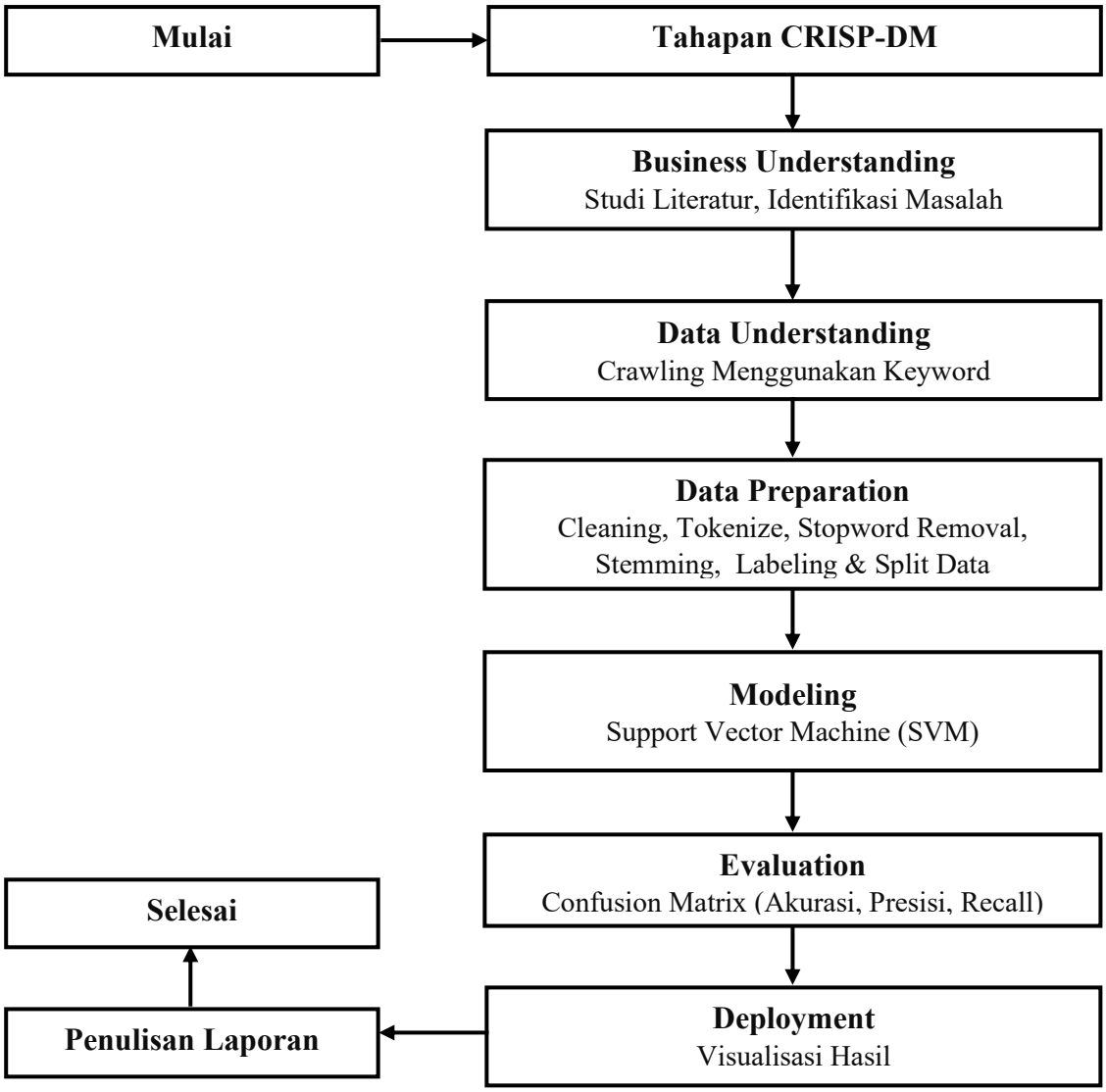
Indonesia adalah negara hukum, namun penegakan hukum di Indonesia seringkali dianggap lemah dan diskriminatif, seperti tercermin dari ungkapan "tumpul ke atas, tajam ke bawah" (I Gede Sujana and I Wayan Kandia, 2024). Krisis kepercayaan publik terhadap institusi hukum semakin diperparah oleh maraknya praktik korupsi yang bahkan melibatkan aparat penegak hukum (Daeng et al., 2024). Situasi ini mendorong masyarakat untuk mencari ruang alternatif untuk menyuarakan aspirasi mereka, dan media sosial seperti Twitter (X) menjadi platform utama untuk tujuan tersebut (Styawati et al., 2021).

Twitter menawarkan data yang aktual dan mencerminkan persepsi publik secara *real-time* mengenai isu penegakan hukum. Penelitian oleh (Khoirunnisa and Topiq, 2024) menunjukkan adanya kecenderungan sentimen negatif yang kuat dari masyarakat terhadap penegakan hukum di Indonesia. Analisis sentimen terhadap data dari media sosial menjadi penting untuk mengukur dan memahami opini publik ini secara sistematis.

Untuk menganalisis volume data teks yang besar dari media sosial, dibutuhkan metode *machine learning* yang andal. Dalam penelitian ini, metode *Support Vector Machine* (SVM) dipilih karena terbukti efektif dalam mengklasifikasikan teks pendek seperti *tweet* dengan akurasi tinggi (Handayani, 2021); (Saputra and Parjito, 2025). Guna memastikan proses analisis berjalan secara terstruktur, penelitian ini menggunakan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang telah terbukti efisien dalam proyek *data mining* (Kannengiesser and Gero, 2023). Kombinasi SVM dan CRISP-DM diharapkan dapat memberikan landasan metodologis yang kuat untuk menghasilkan analisis yang akurat dan dapat dipertanggungjawabkan secara akademis.

2. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan *data mining* dengan menggunakan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*). Metodologi ini dipilih karena alurnya yang sistematis, fleksibel, dan terstruktur (Hasanah et al., 2021). Alur CRISP-DM terdiri dari enam fase yang saling terhubung: pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan penyebaran (*deployment*).



Gambar 1. Alur Penerapan Metode CRISP-DM

2.1 *Support Vector Machine* (SVM)

Menurut (Permata Aulia et al., 2021), *Support Vector Machine* (SVM) adalah algoritma klasifikasi dengan generalisasi tinggi yang menggunakan fungsi kernel untuk memisahkan data non-linear, dan penelitian ini membandingkan kinerja empat kernel umum.

1) Linear Karnel
   $K = (X_i, X_j) = (X_i . X_j)$
2) Polynomial Kernel
   $K = (X_i, X_j) = (X_i . X_j + 1)^p$
3) RBF Kernel
   $K = (X_i, X_j) = e^{-y(X_i . X_j)^2}$
4) Sigmoid Kernel
   $K = (X_i, X_j) = tanh(nX_i + X_j + v)$

2.2 *Term Frequency – Inverse Document Frequency* (TF-IDF)

(Septiani and Isabela, 2022)*Term Frequency – Inverse Document Frequency* (TF-IDF) adalah salah satu metode yang digunakan untuk mengevaluasi tingkat kepentingan sebuah kata (term) dalam suatu dokumen dibandingkan dengan kumpulan dokumen lain. Kata yang sering muncul pada satu dokumen namun jarang dijumpai pada dokumen lain akan diberikan bobot lebih besar. Dengan demikian, TF-IDF mampu mengidentifikasi kata-kata yang lebih relevan dalam merepresentasikan isi dokumen.

Secara teknis, TF-IDF dihitung melalui tiga tahap utama, yaitu *Term Frequency* (TF), *Inverse Document Frequency* (IDF), dan pembobotan TF-IDF.

a. Term Frequency (TF)

(Darwis et al., 2020)Term Frequency menunjukkan frekuensi kemunculan sebuah kata dalam dokumen. Untuk mengurangi dominasi kata yang sering muncul, digunakan pendekatan *augmented term frequency*. Rumusnya adalah:

$$TF = 0{,}5 + 0{,}5\left(\frac{tf}{\max\ (tf)}\right)$$

Keterangan:

- tf = jumlah kemunculan kata dalam sebuah dokumen
- max (tf) = jumlah kemunculan kata terbanyak pada dokumen tersebut

Dengan pendekatan ini, nilai TF dibatasi antara 0,5 hingga 1, sehingga pembobotan kata lebih proporsional.

b. Inverse Document Frequency (IDF)

(Doewes et al., 2022)IDF digunakan untuk mengukur tingkat kelangkaan suatu kata di seluruh dokumen. Kata yang muncul pada banyak dokumen akan memperoleh bobot lebih rendah, sedangkan kata yang jarang muncul akan diberi bobot lebih tinggi. Rumus yang digunakan adalah:

$$IDF = log\frac{N+1}{DF(t)+1} + 1$$

Keterangan:

- N = jumlah keseluruhan dokumen dalam koleksi
- DF(t) = jumlah dokumen yang mengandung kata t
- penambahan konstanta "1" pada pembilang dan penyebut digunakan untuk mencegah pembagian dengan nol serta memastikan semua kata tetap mendapatkan bobot.

c. Bobot TF-IDF

Bobot TF-IDF diperoleh dari hasil perkalian nilai TF dan IDF untuk setiap kata dalam dokumen. Rumusnya adalah:

$$W_{\text{d.t}} = TF_{\text{d.t}} \times IDF_{\text{d.t}}$$

Keterangan:

- Wd,t = bobot kata ke-t pada dokumen ke-d
- TFd,t = nilai *term frequency* kata ke-t pada dokumen ke-d
- IDFd,t = nilai *inverse document frequency* kata ke-t pada dokumen ke-d

Nilai bobot ini menunjukkan seberapa penting suatu kata dalam dokumen tertentu dibandingkan dengan keseluruhan koleksi dokumen. Kata dengan skor TF-IDF tinggi dianggap lebih relevan dalam menjelaskan isi dokumen.

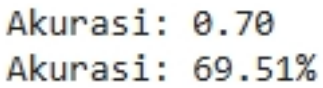
2.3 Tahapan Penelitian

a. Pengumpulan Data (Data Understanding): Data dikumpulkan dari media sosial Twitter (X) menggunakan teknik crawling. Kata kunci yang digunakan mencakup frasa-frasa seperti "penegakan hukum", "korupsi", "keadilan", "aparat hukum", dan "pengadilan". Pengambilan data dilakukan dalam rentang waktu Januari hingga Desember 2024 untuk mendapatkan data yang komprehensif.
b. Pra-pemrosesan Data (Data Preparation): Data teks mentah dibersihkan dan disiapkan melalui beberapa tahapan (Ridwansyah, 2022); (Pratama, 2024):
   1) *Case Folding*: Mengubah seluruh teks menjadi huruf kecil.
   2) *Cleaning*: Menghapus karakter-karakter yang tidak relevan seperti URL, *mention*, *hashtag*, angka, emoji, dan tanda baca.
   3) *Tokenizing*: Memecah teks menjadi kata-kata individu (*token*).
   4) *Stopword Removal*: Menghilangkan kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen (Septiani & Isabela, 2022).
   5) *Stemming*: Mengubah kata-kata menjadi bentuk dasarnya menggunakan algoritma Sastrawi.
c. Transformasi Data (Data Preparation): Data yang telah dibersihkan kemudian dilabeli secara otomatis sebagai positif, negatif, atau netral menggunakan kamus leksikon (*lexicon*) InSet. Setelah itu, data teks diubah menjadi representasi numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*) agar dapat diproses oleh algoritma *machine learning*.
d. Klasifikasi (*Modeling*): Data yang sudah ditransformasi dibagi menjadi data latih (70%) dan data uji (30%). Algoritma *Support Vector Machine* (SVM) dengan *kernel linear* kemudian digunakan untuk membangun model klasifikasi sentimen (Permata Aulia, Arifin & Mayasari, 2021).
e. Evaluasi (*Evaluation*): Performa model diukur menggunakan *confusion matrix* dan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil evaluasi ini digunakan untuk menilai seberapa efektif model SVM dalam mengklasifikasikan sentimen masyarakat.

3. HASIL DAN PEMBAHASAN

Berdasarkan analisis sentimen terhadap *tweet* mengenai penegakan hukum di Indonesia sepanjang tahun 2024, diperoleh hasil sebagai berikut:

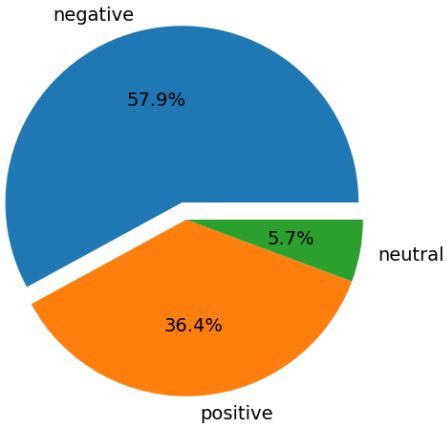
a. Akurasi Model: Model klasifikasi SVM menunjukkan akurasi sebesar 69,51% pada data uji. Akurasi ini lebih rendah dari beberapa penelitian terdahulu seperti yang dilakukan oleh (Handayani, 2021) dan (Saputra and Parjito, 2025) yang mencapai akurasi di atas 90%. Keterbatasan pada penelitian ini, khususnya jumlah data uji yang sangat terbatas (hanya 3 dokumen), kemungkinan menjadi penyebab utama akurasi yang rendah.



Gambar 2. Akurasi

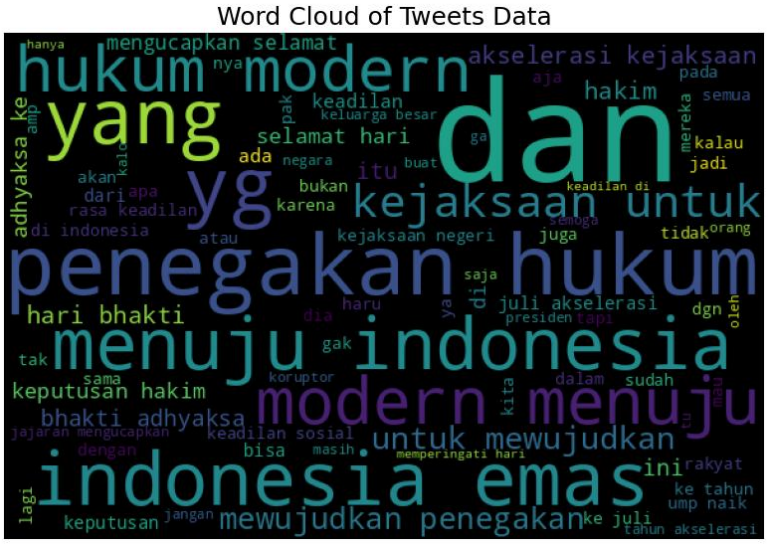
b. Distribusi Sentimen: Distribusi sentimen dari 4277 tweet yang dianalisis menunjukkan sentimen negatif mendominasi dengan proporsi 57,9%, diikuti oleh sentimen positif sebesar 36,4%, dan sentimen netral 5,7%. Hasil ini mengindikasikan

bahwa opini publik cenderung terbagi antara kritik dan apresiasi terhadap penegakan hukum, meskipun sentimen negatif sedikit lebih menonjol.

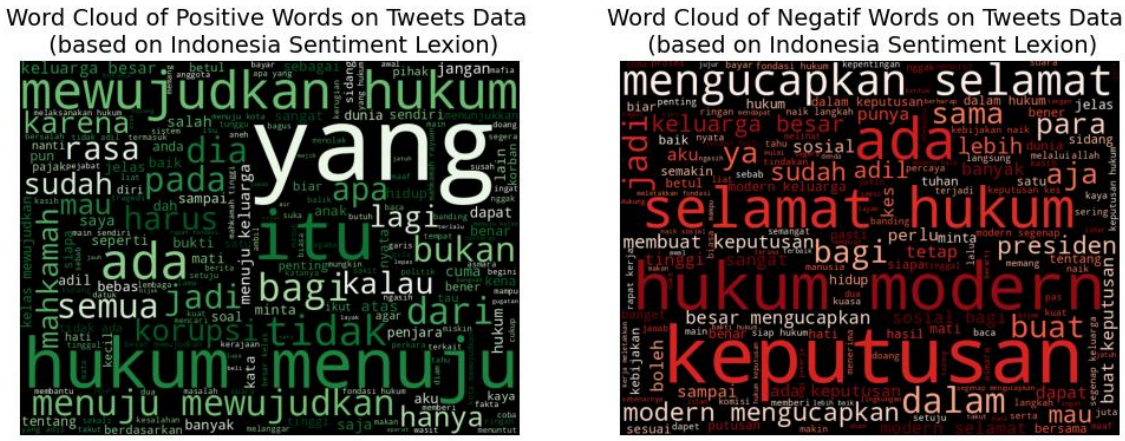


Gambar 3. Pie Chart

c. Kata Kunci Dominan: Visualisasi *Word Cloud* menunjukkan bahwa kata-kata yang paling sering muncul adalah "koruptor", "keputusan", "hakim", "bijak", dan "ringan". Kombinasi kata-kata ini secara jelas menggambarkan narasi publik yang berfokus pada kritik terhadap putusan hakim yang dianggap terlalu ringan terhadap kasus korupsi. Hal ini sejalan dengan temuan penelitian yang menunjukkan adanya fenomena "No Viral No Justice" yang mendorong aparat hukum merespons tekanan publik (Gussela et al., 2024).



Gambar 4. Word Cloud



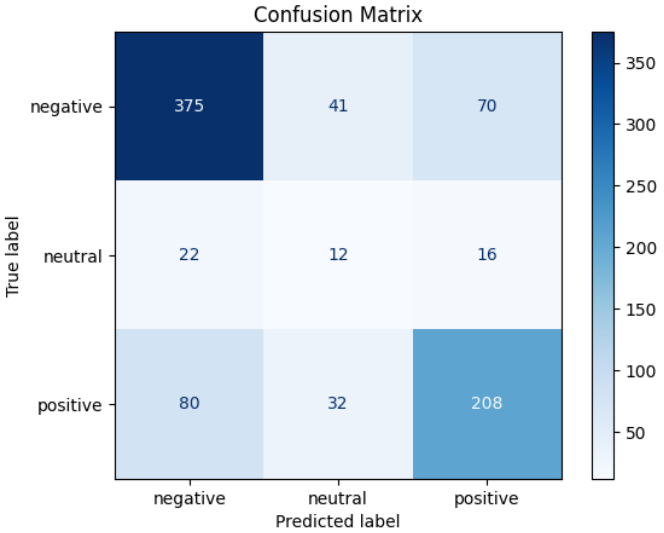
Gambar 5. Word Cloud Positive dan Negative

d. Performa Model per Kelas: Berdasarkan *confusion matrix* dan metrik evaluasi, model memiliki performa yang baik dalam memprediksi sentimen negatif (*precision* 1) dan sentimen positif (*recall* 1). Namun, performanya pada sentimen netral dan saat membedakan sentimen negatif dan positif masih kurang optimal. Akurasi yang rendah (69,51%) dan ketidakseimbangan performa ini dapat diperbaiki dengan jumlah data yang lebih besar dan seimbang.

```
Akurasi: 0.70
Akurasi: 69.51%

Laporan Klasifikasi:
              precision    recall  f1-score   support

    negative       0.79      0.77      0.78       486
     neutral       0.14      0.24      0.18        50
    positive       0.71      0.65      0.68       320

    accuracy                           0.70       856
   macro avg       0.54      0.55      0.54       856
weighted avg       0.72      0.70      0.71       856
```

Gambar 6. Laporan Klasifikasi

e. Confusion matrix: Confusion matrix menunjukkan model berhasil mengklasifikasikan sebagian besar tweet dengan benar, yaitu 375 negative, 12 neutral, dan 208 positive. Namun, masih terdapat kesalahan, terutama antara kelas yang berdekatan seperti neutral dan positive. Secara keseluruhan, performa model tergolong cukup baik.



Gambar 7. Confusion Matrix

4. KESIMPULAN

Analisis sentimen terhadap tweet terkait penegakan hukum di Indonesia pada tahun 2024 menunjukkan bahwa sentimen negatif lebih mendominasi dibandingkan dengan sentimen positif dan netral. Hal ini terlihat dari hasil analisis polaritas sentimen dan juga word cloud yang menampilkan kata-kata dengan sentimen negatif yang lebih menonjol. Model SVM yang dibangun untuk mengklasifikasikan sentimen tweet menunjukkan akurasi yang cukup baik. Namun, masih terdapat beberapa kesalahan klasifikasi, terutama pada tweet dengan sentimen netral. Hal ini kemungkinan disebabkan oleh jumlah data yang tidak seimbang, di mana jumlah tweet dengan sentimen netral jauh lebih sedikit dibandingkan dengan sentimen positif dan negatif. Secara keseluruhan, analisis ini memberikan gambaran umum mengenai sentimen masyarakat terhadap penegakan hukum di Indonesia pada tahun 2024. Untuk penelitian selanjutnya, disarankan untuk menggunakan dataset yang lebih besar dan seimbang untuk meningkatkan akurasi model dan mendapatkan hasil yang lebih representatif.

5. DAFTAR PUSTAKA

Daeng, Y., Putri, D., S F, B., Rahmat, K., 2024. Keterbatasan Aparat Penegak Hukum Sebagai Hambatan Dalam Penegakan Hukum di Indonesia. J. Multidisiplin Teknol. dan Arsit. 2, 671–676. https://doi.org/10.57235/motekar.v2i2.3791

Darwis, D., Pratiwi, E.S., Pasaribu, A.F.O., 2020. Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. Edutic - Sci. J. Informatics Educ. 7, 1–11. https://doi.org/10.21107/edutic.v7i1.8779

Doewes, A., Saxena, A., Pei, Y., Pechenizkiy, M., 2022. Individual Fairness Evaluation for Automated Essay Scoring System. Proc. 15th Int. Conf. Educ. Data Mining, EDM 2022. https://doi.org/10.5281/zenodo.6853151

Gussela, M.D., Kurniawati, M., N, J.S., Hermanto, D., Fauziansah, S., Saebani, B.A., 2024. Fenomena "No Viral No Justice" Perspektif Teori Penegakkan Hukum. Ranah Res. J. Multidiscip. Res. Dev. 7, 792–800. https://doi.org/10.38035/rrj.v7i2.1326

Handayani, R.N., 2021. Optimasi Algoritma Support Vector Machine untuk Analisis Sentimen pada Ulasan Produk Tokopedia Menggunakan PSO. Media Inform. 20, 97–108. https://doi.org/10.37595/mediainfo.v20i2.59

Hasanah, M.A., Soim, S., Handayani, A.S., 2021. Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. J. Appl. Informatics Comput. 5, 103–108. https://doi.org/10.30871/jaic.v5i2.3200

I Gede Sujana, I Wayan Kandia, 2024. Indikator Lemahnya Penegakan Hukum di Indonesia. IJOLARES Indones. J. Law Res. 2, 56–62. https://doi.org/10.60153/ijolares.v2i2.67

Kannengiesser, U., Gero, J.S., 2023. Modelling the Design of Models: an Example Using Crisp-Dm. Proc. Des. Soc. 3, 2705–2714. https://doi.org/10.1017/pds.2023.271

Khoirunnisa, F., Topiq, S., 2024. Analisis Sentimen Terhdap Kepercayaan Masyarakat Pada Proses Penegak Hukum Di Indonesia Dengan Menggunakan Algoritma Naïve Bayes. J. Inform. dan Tek. Elektro Terap. 12, 2128–2139. https://doi.org/10.23960/jitet.v12i3.4683

Permata Aulia, T.M., Arifin, N., Mayasari, R., 2021. Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. SINTECH

(Science Inf. Technol. J. 4, 139–145. https://doi.org/10.31598/sintechjournal.v4i2.762

Pratama, R.A., 2024. Analisis Sentimen Konsumen Dengan Teknik Text Mining. J. Dunia Data 1, 1–17.

Ridwansyah, T., 2022. Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. KLIK Kaji. Ilm. Inform. dan Komput. 2, 178–185. https://doi.org/10.30865/klik.v2i5.362

Saputra, M.R., Parjito, P., 2025. Analisis Sentimen Twitter Terhadap Konflik Di Papua Menggunakan Perbandingan Naive Bayes Dan Svm. JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform. 10, 1197–1208. https://doi.org/10.29100/jipi.v10i2.6180

Septiani, D., Isabela, I., 2022. Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks. SINTESIA J. Sist. dan Teknol. Inf. Indones. 1, 81–88.

Styawati, S., Hendrastuty, N., Isnain, A.R., 2021. Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. J. Inform. J. Pengemb. IT 6, 150–155. https://doi.org/10.30591/jpit.v6i3.2870