

ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PENJARAHAN RUMAH ANGGOTA DPR MENGGUNAKAN METODE KLASIFIKASI ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE

SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON THE LOOTING OF A HOUSE OF A MEMBER OF THE HOUSE OF REPRESENTATIVES (DPR) USING NAIVE BAYES AND SUPPORT VECTOR MACHINE CLASSIFICATION ALGORITHMS

Siti Magfiroh^{1*}, Abd Ghofur², Nur Asize³

Email: *03siti24@gmail.com, apunkbwi@gmail.com, nuriza3010@gmail.com

^{1,2,3} Teknologi Informasi, Universitas Ibrahimy

Abstrak

YouTube menjadi ruang publik digital yang memungkinkan masyarakat menyampaikan opini terhadap isu sosial dan politik. Salah satu isu yang menarik perhatian adalah penjarahan rumah anggota DPR yang diberitakan melalui kanal iNews Official. Banyaknya komentar dengan sentimen positif, negatif, dan netral mendorong perlunya analisis sentimen untuk memahami persepsi publik secara objektif. Penelitian ini bertujuan menganalisis sentimen komentar YouTube terkait kasus tersebut menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM), serta membandingkan performa keduanya. Metode yang digunakan adalah *Knowledge Discovery in Database (KDD)* yang meliputi seleksi data, *preprocessing*, *transformasi*, *TF-IDF*, *klasifikasi*, dan evaluasi. Data dikumpulkan melalui scraping menggunakan YouTube API, kemudian diproses melalui *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. Evaluasi dilakukan menggunakan confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian diharapkan dapat menunjukkan algoritma yang lebih efektif dalam mengklasifikasikan sentimen serta memberikan gambaran objektif mengenai opini publik di ruang digital.

Kata kunci: Analisis Sentimen, YouTube, Naive Bayes, SVM, TF-IDF.

Abstract

YouTube serves as a digital public space where users express opinions on social and political issues. One widely discussed issue is the looting of a member of the House of Representatives' residence, reported on the iNews Official channel. The large number of positive, negative, and neutral comments highlights the need for sentiment analysis to objectively understand public perception. This study aims to analyze YouTube comment sentiments using Naive Bayes and Support Vector Machine (SVM) algorithms and compare their performance. This research applies the Knowledge Discovery in Database (KDD) method, including data selection, preprocessing, TF-IDF transformation, classification, and evaluation. Data were collected through YouTube API scraping and processed using cleaning, tokenizing, stopwords removal, and stemming. Model performance was evaluated using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results are expected to determine the most effective algorithm and provide an objective overview of public opinion in the digital space.

Keywords: Sentiment Analysis, YouTube, Naive Bayes, SVM, TF-IDF.

1. PENDAHULUAN

Di era digital saat ini, media sosial telah menjadi sasaran utama bagi masyarakat dalam menyampaikan opini dan berinteraksi terhadap berbagai isu sosial maupun politik. Salah satu platform yang paling banyak digunakan adalah YouTube, yang hanya berfungsi sebagai media berbagai video, tetapi juga menyediakan kolom komentar sebagai ruang diskusi publik yang terbuka. Tingginya jumlah pengguna YouTube di Indonesia menjadikan platform ini sebagai sumber data yang potensial untuk menggali opini publik secara luas dan beragam [1].

Salah satu isu yang menarik perhatian publik adalah peristiwa penjarahan rumah anggota DPR yang diberitakan melalui kanal iNews Official. Peristiwa tersebut memicu berbagai reaksi dari masyarakat, yang tercermin melalui komentar-komentar pengguna di YouTube [2]. Komentar tersebut mengandung beragam sentimen, baik positif, negatif, maupun netral, yang mencerminkan persepsi, sikap, dan emosi masyarakat terhadap peristiwa tersebut [3]. Namun, banyaknya jumlah komentar serta sifat data yang tidak terstruktur menyebabkan analisis secara manual menjadi kurang efektif dan rentan terhadap subjektivitas [4].

Analisis sentimen merupakan salah satu teknik dalam text mining yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam suatu teks [5]. Dengan memanfaatkan metode machine learning, proses analisis sentimen dapat dilakukan secara otomatis, cepat, dan objektif. Dalam penelitian ini, digunakan dua algoritma klasifikasi yang umum diterapkan dalam analisis teks, yaitu Naive Bayes dan Support Vector Machine (SVM). Naive Bayes dikenal sebagai metode probabilistik yang sederhana dan efisien dalam mengolah data teks, sedangkan SVM memiliki kemampuan yang baik dalam memisahkan data dengan tingkat akurasi yang tinggi melalui pembentukan hyperplane optimal [6].

Beberapa penelitian sebelumnya menunjukkan bahwa kedua algoritma tersebut memiliki performa yang baik dalam klasifikasi sentimen, dengan SVM cenderung menghasilkan akurasi yang lebih tinggi, sementara Naive Bayes unggul dalam kecepatan komputasi [7]. Oleh karena itu, perbandingan antara kedua metode ini menjadi penting untuk mengetahui algoritma yang paling efektif dalam mengklasifikasikan sentimen komentar berbahasa Indonesia, khususnya pada konteks isu sosial dan politik.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen komentar YouTube terkait peristiwa penjarahan rumah anggota DPR menggunakan metode klasifikasi Naive Bayes dan Support Vector Machine (SVM), serta membandingkan performa keduanya. Penelitian ini diharapkan dapat memberikan gambaran objektif mengenai opini publik di ruang digital, serta berkontribusi dalam pengembangan metode analisis sentimen yang lebih akurat dan relevan dalam memahami dinamika sosial di masyarakat.

2. METODOLOGI

Penelitian ini menggunakan kuantitatif dengan metode Knowledge Discovery in Database (KDD) untuk menganalisis sentimen komentar YouTube terkait penjarahan rumah anggota DPR. Metode ini dipilih karena mampu mengolah data dalam jumlah besar dan tidak terstruktur menjadi informasi yang bermakna melalui tahapan sistematis. Knowledge Discovery in Database (KDD) merupakan kegiatan data mining berbasis pengetahuan yang melibatkan pengumpulan data, pemrosesan data, transformasi data, penerapan algoritma dan evaluasi [8]. Tahapan dalam metode KDD meliputi selection (pemilihan data), preprocessing (pembersihan data), transformation (transformasi data), data mining (pemodelan), dan evaluasi. Setiap dirancang secara terstruktur agar proses analisis dapat dilakukan secara sistematis dan hasil yang diperoleh dapat direplikasi oleh peneliti lain [9].

2.1. Alur Penelitian

Alur penelitian ini ditunjukkan pada gambar 1 yang menggambarkan tahapan proses analisis

sentimen menggunakan metode KDD.



Gambar.1 Alur metode Knowledge Discovery in Database (KDD)

2.2. Tahapan Metode KDD

2.2.1 Selection (Pemilihan Data)

Tahap selection merupakan proses awal dalam penelitian yang bertujuan untuk mengumpulkan data yang relevan dengan topik penelitian. Data yang digunakan dalam penelitian ini berupa komentar pengguna pada video YouTube yang membahas penjarahan rumah anggota DPR dari kanal iNews Official [10].

Pengumpulan data dilakukan menggunakan teknik web scraping dengan bantuan YouTube Data API. Proses ini dilakukan dengan memanfaatkan bahasa pemrograman Python untuk mengambil data secara otomatis. Data yang dikumpulkan meliputi : Isi komentar (text); Waktu komentar (timestamp); Nama pengguna (username). Setelah data terkumpul, dilakukan proses seleksi untuk memastikan hanya komentar berbahasa indonesia yang digunakan. Hal ini bertujuan agar analisis sentimen yang dilakukan lebih relevan dan akurat sesuai konteks penelitian.

2.2.2 Labelling Data

Tahap labeling merupakan proses pemberian label sentimen pada data komentar yang telah dikumpulkan. Dalam penelitian ini, pelabelan dilakukan menggunakan pendekatan lexicon-based labeling yang diimplementasikan menggunakan bahasa pemrograman Python. Metode lexicon-based labeling merupakan pendekatan berbasis kamus sentimen (sentiment lexicon) yang digunakan untuk menentukan polaritas suatu teks berdasarkan kata-kata yang terkandung di dalamnya. Setiap komentar dianalisis dengan mencocokkan kata-kata yang terdapat dalam teks dengan daftar kata yang memiliki nilai sentimen positif maupun negatif [11]. Label yang digunakan dalam penelitian ini terdiri dari tiga kategori, yaitu:

- **Positif**, yaitu komentar yang mengandung kata-kata bernilai positif seperti “bagus”, “setuju”, dan “mantap”.
- **Negatif**, yaitu komentar yang mengandung kata-kata bernilai negatif seperti “buruk”, “jelek”, dan “marah”.
- **Netral**, yaitu komentar yang tidak menunjukkan kecenderungan sentimen tertentu.

2.2.3 Preprocessing (Pembersihan Data)

Tahap preprocessing merupakan proses pembersihan dan normalisasi data teks sebelum dilakukan analisis lebih lanjut. Data komentar yang diperoleh dari YouTube umumnya masih mengandung noise seperti simbol, tanda baca, dan kata tidak baku sehingga perlu dilakukan proses pembersihan agar kualitas data meningkat [12]. Tahapan preprocessing dalam penelitian ini meliputi beberapa langkah sebagai berikut:

1. **Cleaning**

Menghapus elemen yang tidak diperlukan seperti URL, emoji, angka, tanda baca, dan karakter khusus lainnya yang tidak berkontribusi terhadap analisis sentimen.

2. **CaseFolding**

Mengubah seluruh teks menjadi huruf kecil (lowercase) untuk menyeragamkan bentuk kata dan menghindari perbedaan akibat penggunaan huruf besar dan kecil.

3. **Tokenizing**

Memecah kalimat menjadi kata-kata individu (token) sehingga memudahkan proses analisis teks.

4. **Stopword Removal**

Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis sentimen seperti “dan”, “yang”, “di”, dan sejenisnya.

5. **Stemming**

Mengubah kata berimbuhan menjadi bentuk dasar menggunakan algoritma stemming, sehingga kata seperti “berjalan”, “berlari” diubah menjadi “jalan”, “lari”.

2.2.4 Transformation (TF-IDF)

Tahap transformasi bertujuan untuk mengubah data teks menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning. Data teks pada awalnya tidak dapat langsung digunakan dalam proses komputasi, sehingga perlu direpresentasikan dalam bentuk vektor numerik menggunakan teknik seperti Term Frequency-Inverse Document Frequency (TF-IDF) [13].

- **Term Frequency (TF)**: mengukur seberapa sering suatu kata muncul dalam satu dokumen
- **Inverse Document Frequency (IDF)**: mengukur seberapa penting kata tersebut berdasarkan frekuensi kemunculannya di seluruh dokumen

Kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain akan memiliki bobot yang tinggi. Hasil dari proses TF-IDF berupa matriks fitur yang merepresentasikan setiap komentar sebagai vektor numerik, yang kemudian digunakan sebagai input dalam proses klasifikasi.

2.2.5 Modelling (Naive Bayes & SVM)

Tahap modeling merupakan proses pembangunan model klasifikasi sentimen menggunakan algoritma machine learning [14]. Dalam penelitian ini digunakan dua algoritma yaitu Naive Bayes dan Support Vector Machine (SVM) untuk membandingkan performa keduanya.

Sebelum dilakukan pemodelan, dataset dibagi menjadi dua bagian, yaitu:

- Data training (80%) untuk melatih model
- Data testing (20%) untuk menguji model

2.2.6 Evaluation (Evaluasi Model)

Tahap evaluasi dilakukan untuk mengukur kinerja model klasifikasi yang telah dibangun. Evaluasi dilakukan menggunakan metode confusion matrix yang membandingkan hasil prediksi model dengan label sebenarnya [15].

Dari confusion matrix, diperoleh beberapa metrik evaluasi, yaitu:

1. **Akurasi**

Mengukur tingkat ketepatan model dalam mengklasifikasikan data secara keseluruhan.

2. **Presisi**

Mengukur seberapa tepat model dalam memprediksi kelas tertentu.

3. Recall

Mengukur kemampuan model dalam menemukan seluruh data yang termasuk dalam kelas tertentu.

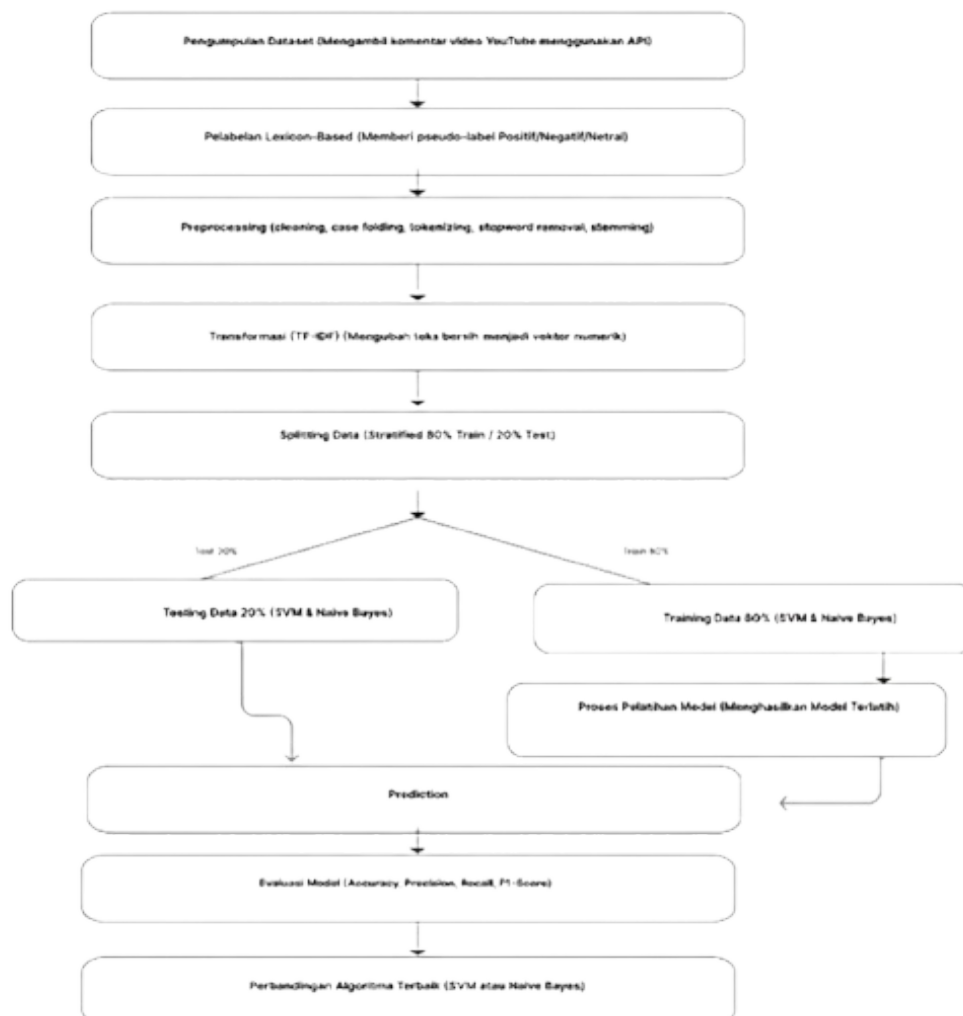
4. F1-Score

Merupakan rata-rata harmonis dari presisi dan recall yang digunakan untuk menilai keseimbangan performa model.

Hasil evaluasi ini digunakan untuk membandingkan performa antara algoritma Naive Bayes dan Support Vector Machine (SVM), sehingga dapat diketahui metode mana yang lebih efektif dalam mengklasifikasikan sentimen komentar YouTube.

2.2.7. Alur Proses

Penelitian analisis sentimen terhadap komentar YouTube terkait peristiwa penjarahan rumah anggota DPR dilakukan melalui serangkaian tahapan yang sistematis dan terstruktur. Alur proses penelitian yang telah dirancang tersebut kemudian divisualisasikan dalam bentuk block diagram untuk memberikan gambaran yang lebih jelas mengenai langkah-langkah yang ditempuh dalam penelitian ini. Seperti gambar 2 dibawah ini :



Gambar . 2 Alur Proses Penelitian

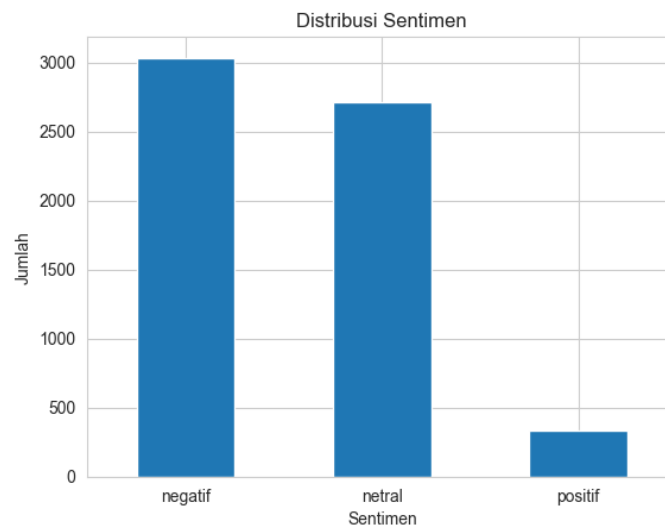
3 HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan komentar dari video YouTube yang diambil menggunakan API. Setelah proses pengambilan, didapatkan total 6086 data mentah yang kemudian disimpan dalam format dataset.

3.2 Pelabelan Otomatis (*Lexicon-based*)

Pada tahap ini, data mentah diberikan label sentimen secara otomatis menggunakan metode *Lexicon-based*. Proses ini dilakukan sebelum tahap preprocessing agar bobot kata pada kamus leksikon dapat terdeteksi secara optimal pada teks asli. Hasil pelabelan membagi data ke dalam tiga kategori: Positif, Negatif, dan Netral. Seperti gambar. 3 dibawah ini



Gambar. 3 Hasil Pelabelan Data

3.3 Preprocessing Data

Setelah data memiliki label, dilakukan pembersihan data (preprocessing) untuk meningkatkan kualitas data sebelum masuk ke model machine learning. Tahapan yang dilakukan meliputi:

- **Cleaning:** Menghapus URL, simbol, dan angka.
- **Case Folding:** Menyeragamkan semua huruf menjadi huruf kecil.
- **Tokenizing:** Memecah kalimat menjadi potongan kata.
- **Stopword Removal:** Menghapus kata umum yang tidak memiliki makna penting.
- **Stemming:** Mengubah kata menjadi kata dasar.

3.4 Transformasi Data dan Splitting

Data teks yang telah bersih ditransformasikan menjadi bentuk numerik menggunakan teknik TF-IDF. Selanjutnya, dataset dibagi menggunakan metode *Stratified Split* dengan proporsi 80% data latih (training) dan 20% data uji (testing) untuk menjaga keseimbangan distribusi kelas pada masing-masing subset.

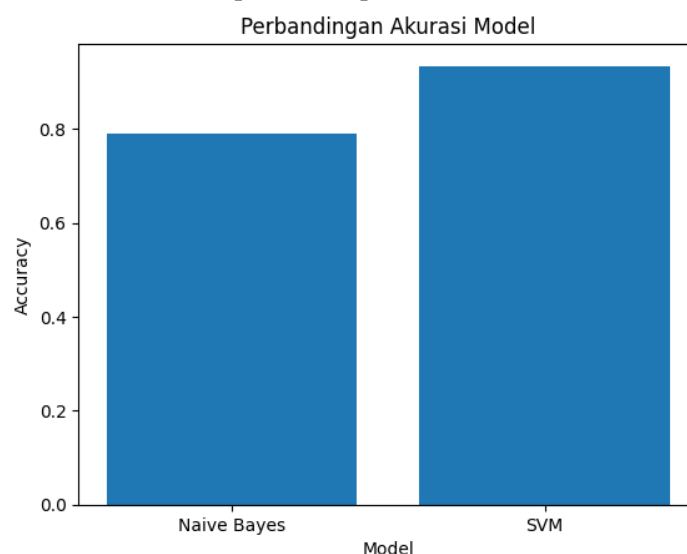
3.5 Pengujian dan Evaluasi Model

Tahap pengujian dilakukan dengan membandingkan performa dua algoritma klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Naive Bayes*. Pengujian dilakukan menggunakan data uji sebanyak 1.192 data yang telah diproses melalui tahapan KDD. Perbandingan metrik evaluasi dari kedua model tersebut disajikan pada Tabel 1:

Tabel . 1 Perbandingan Performa Algoritma

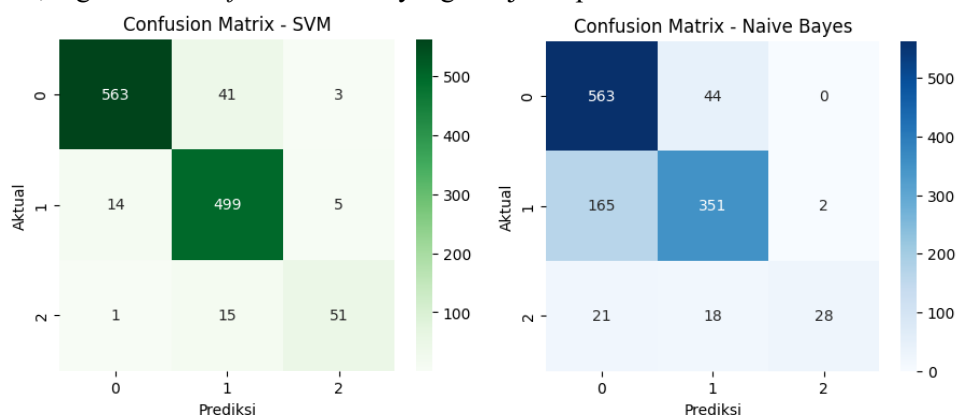
Algoritma	Accuracy	Precision	Recall	F1-Score
SVM	93,37%	91,00%	88,33%	89,67%
Naive Bayes	79,03%	84,33%	67,67%	72,00%

Berdasarkan Tabel . 1 terlihat secara signifikan bahwa algoritma SVM memiliki performa yang lebih unggul dibandingkan Naive Bayes di seluruh parameter evaluasi. Perbandingan akurasi secara visual dapat dilihat pada Gambar. 4 di bawah ini:



Gambar. 4 Perbandingan Akurasi Model

Untuk menganalisis sebaran klasifikasi pada masing-masing label sentimen secara lebih mendalam, digunakan *Confusion Matrix* yang disajikan pada Gambar . 5



Gambar . 5 Confusion Matrix Model SVM dan NB

Berdasarkan hasil pengujian, algoritma Support Vector Machine (SVM) menunjukkan efektivitas yang sangat tinggi dengan akurasi mencapai 93,37%. Hal ini membuktikan bahwa SVM mampu menentukan *hyperplane* yang optimal dalam memisahkan fitur-fitur teks yang

kompleks hasil dari transformasi TF-IDF. Kemampuan SVM dalam menangani data dimensi tinggi pada teks komentar YouTube terbukti lebih stabil dibandingkan Naive Bayes.

Pada Gambar Y (sisi kiri), terlihat bahwa Naive Bayes mengalami banyak kesalahan klasifikasi, terutama pada data yang memiliki label tertentu (terlihat dari angka 165 kesalahan prediksi). Sebaliknya, pada sisi kanan Gambar Y, SVM menunjukkan diagonal utama yang lebih solid (563, 499, dan 51), yang menandakan bahwa mayoritas data berhasil diprediksi dengan benar sesuai label aktualnya. Rendahnya nilai *recall* pada Naive Bayes (67,67%) dibandingkan SVM (88,33%) memperkuat kesimpulan bahwa SVM lebih handal dalam mengenali pola sentimen publik terkait isu penjarahan rumah anggota DPR ini secara objektif.

4 KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa metode *machine learning* efektif dalam memetakan opini publik di YouTube, di mana algoritma *Support Vector Machine* (SVM) terbukti jauh lebih unggul dengan akurasi mencapai 93,37% dibandingkan Naive Bayes yang hanya sebesar 79,03%. Keunggulan SVM ini dipengaruhi oleh penggunaan TF-IDF dan pelabelan *lexicon-based* yang mampu menangani karakteristik teks komentar secara optimal. Untuk pengembangan ke depan, disarankan untuk menambah volume dataset dan menerapkan algoritma *Deep Learning* seperti LSTM atau IndoBERT guna menangani kompleksitas bahasa informal serta meningkatkan keseimbangan performa pada setiap kelas sentimen.

5 DAFTAR RUJUKAN

- [1] M. R. Kety Nandini, "Sentiment Analysis of Economic Policy Comments on YouTube Using," *Journal of Applied Informatics and Computing (JAIC)*, 2025.
- [2] A. T. Sekar De Putri, Analisis percakapan netizen pada channel YouTube @TotalPolitik sebagai media ruang publik komunikasi., *Jurnal Ilmiah Kajian Ilmu Sosial dan Budaya.*, 2024.
- [3] H. O. d. G. A. Dirgahayu Marganingsih, "Analisis sentimen komentar YouTube MasterChef Indonesia menggunakan algoritma Support Vector Machine dan Gaussian Naive Bayes.," *Jurnal Informatika dan Teknologi Pendidikan.*, 2025.
- [4] I. F. H. Ahmad Rijal Hermawan, "Sentiment Classification of Public Perception on LHKPN Using SVM and Naive Bayes.," *Jurnal SISFOKOM (Sistem Informasi dan Komputer)*, 2025.
- [5] A. F. R. K. Wibowo, "Analisis Sentimen Ulasan Aplikasi Menggunakan Metode Naive Bayes Classifier," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, p. 233–240, 2022.
- [6] O. M. H. S. S. N. F. & H. B. Nurfauzi, "Analisis Sentimen Grab Indonesia pada Ulasan Google Play Store menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *SMARTICS Journal*, vol. 11, no. 1, p. 8–13., 2025.
- [7] J. Widyadhana, "Perbandingan Algoritma Naive Bayes dan Support Vector Machine (SVM) pada Analisis Sentimen.," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 8, no. 5, pp. 937-944, 2021.
- [8] T. D. P. Y. B. W. Z. A. E. & P. A. Permana, "Perbandingan Performa Algoritma Naive Bayes dan Support Vector Machine (SVM) untuk Analisis Sentimen Komentar YouTube terhadap Industri Esports di Indonesia.," *JINU (Jurnal Informatika Nusantara)*, vol. 2, no. 6, p. 1390–1399, 2025.
- [9] D. P. & W. A. Sari, "Penerapan Metode Knowledge Discovery in Database (KDD) untuk Analisis Data Pendidikan," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 9, no. 3, pp. 577-584, 2022.

- [10] A. & S. M. Kurniawan, "Analisis Sentimen Komentar YouTube Menggunakan Metode Machine Learning," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 234-240, 2022.
- [11] M. N. S. H. A. Haq, "Analisis Sentimen Media Sosial Menggunakan Pendekatan Lexicon-Based.," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 3, pp. 489-495, 2022.
- [12] S. & A. A. S. Khomsah, "model preprocessing komentar YouTube dalam bahasa Indonesia," *Jurnal RESTI*, vol. 4, no. 4, p. 648–654, 2020.
- [13] K. M. K. J. H. M. M. S. B. L. & B. D. Kowsari, "A survey on feature extraction techniques for text classification," *Information*, vol. 10, no. 4, p. 142, 2019.
- [14] Y. S. R. Azmi Fauziah Nur, " Naive Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (KNN)," *LINIER: Literatur Informatika dan Komputer*, vol. 1, no. 2, 2024.
- [15] A. P. D. N. A. Saputra, "Analisis Sentimen Menggunakan Naive Bayes dengan Confusion Matrix," *Jurnal Informatika*, vol. 8, no. 2, p. 120–126, 2020.