

Persepsi Publik terhadap Pendidikan Tinggi dalam Komentar YouTube Mata Najwa: Analisis Sentimen dengan Algoritma Support Vector Machine

Muhammad Hilmy Maulana¹, Reyner Archi Sanchia², Dwi Arman Prasetya³

¹²³Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur

¹²³Surabaya-Indonesia

Email: ¹ 23083010106@student.upnjatim.ac.id, ²

23083010085@student.upnjatim.ac.id, ³ arman.prasetya.sada@upnjatim.ac.id

Abstract

Kolom komentar YouTube merupakan cerminan opini publik terhadap berbagai isu sosial dan pendidikan. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap pendidikan tinggi berdasarkan komentar pada video YouTube Mata Najwa berjudul “Kenapa Perlu Kuliah ala Maudy, Andhika, & Amanda”. Metode yang digunakan adalah Support Vector Machine (SVM) untuk mengklasifikasikan opini masyarakat ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Data dikumpulkan menggunakan YouTube Data API dan melalui beberapa tahap pra-pemrosesan, meliputi case folding, tokenization, stopword removal, serta stemming. Pelabelan sentimen dilakukan secara otomatis menggunakan TextBlob sebagai data latih untuk model SVM. Hasil evaluasi menunjukkan bahwa model SVM memperoleh tingkat akurasi sebesar 57%, yang menunjukkan performa sedang dalam mengklasifikasikan teks informal berbahasa Indonesia. Hasil analisis memperlihatkan bahwa komentar dengan sentimen positif mendominasi, mencerminkan dukungan masyarakat terhadap pentingnya pendidikan tinggi dalam pengembangan diri dan karier. Sementara itu, komentar negatif lebih banyak menyoroti persoalan relevansi antara pendidikan formal dan kebutuhan dunia kerja. Penelitian ini menyimpulkan bahwa algoritma SVM efektif dalam memetakan opini publik terhadap isu pendidikan tinggi di media sosial.

Keywords: Analisis Sentimen, Support Vector Machine, Komentar, YouTube, Pendidikan Tinggi

Abstraksi

YouTube comment sections serve as a reflection of public opinions regarding various social and educational issues. This study aims to analyze public sentiment toward higher education, as reflected in comments on the Mata Najwa YouTube video titled “Kenapa Perlu Kuliah ala Maudy, Andhika, & Amanda”. The research implements the Support Vector Machine (SVM) method to classify public opinions into three sentiment categories: positive, negative, and neutral. The dataset was collected using the YouTube Data API and underwent several pre-processing steps, including case folding, tokenization, stopword removal, and stemming. Sentiment labeling was performed automatically using TextBlob to generate the target classes for model training. The evaluation results show that the SVM model achieved an accuracy of 57%, indicating a moderate level of performance in classifying informal Indonesian text. The analysis also revealed that positive sentiment dominates the comment section, showing that many users support higher education as an

essential part of self-development and opportunity expansion. Meanwhile, negative comments mostly highlight concerns about the practical relevance of formal education. This research concludes that the SVM algorithm is capable of mapping the polarization of public opinion in online discussions, particularly regarding higher education in Indonesia.

Kata Kunci: *Sentiment Analysis, Support Vector Machine, YouTube Comments, Higher Education, Maudy Ayunda.*

1. PENDAHULUAN

Pendidikan tinggi merupakan salah satu pilar penting dalam pembangunan sumber daya manusia yang berkualitas [1]. Namun, opini masyarakat terkait Pendidikan tinggi seringkali beragam, terbukti dari berbagai komentar yang muncul di media sosial, salah satunya pada platform YouTube. Salah satu kanal yang memberikan ruang diskusi bagi masyarakat adalah video wawancara dan diskusi yang diselenggarakan oleh Mata Najwa, yang dalam salah satu episodenya mengundang Maudy Ayunda sebagai narasumber. Opini yang diberikan oleh Maudy Ayunda dapat digunakan sebagai penilaian dan menganalisa bagaimana sentimen yang muncul terhadap narasumber serta topik yang diperbincangkan [2].

Analisis Sentimen merupakan metode yang efektif untuk mengidentifikasi dan mengklasifikasi opini dalam teks secara otomatis menjadi kategori Positif, Negatif dan Netral [3]. Dalam analisis sentimen ini, metode Support Vector Machine digunakan sebagai model klasifikasi untuk mengukur sentimen dalam komentar netizen pada video Mata Najwa tersebut.

Data komentar diperoleh melalui proses crawling dari komentar video YouTube Mata Najwa dengan Maudy Ayunda, kemudian dilakukan pra-pemrosesan teks mulai dari penghilangan karakter khusus, tokenisasi, dan normalisasi. Hasil analisis diharapkan dapat memberikan gambaran yang jelas mengenai opini publik terhadap pendidikan tinggi sekaligus memberikan kontribusi pada pemahaman dinamika sosial di ranah digital. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kualitas pendidikan di Indonesia berdasarkan komentar YouTube dari video “Kenapa Perlu Kuliah ala Maudy, Andhika, & Amanda | Mata Najwa” Hasil analisis diharapkan mampu memberikan gambaran umum tentang opini masyarakat serta menjadi masukan bagi pemerintah dalam merumuskan kebijakan pendidikan yang lebih responsif.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah proses yang menggunakan teknik pemrosesan bahasa alami (NLP), analisis teks, dan linguistik komputasional untuk mengidentifikasi, mengekstrak, dan mengukur informasi subjektif dan keadaan emosional yang terkandung dalam teks tertulis. Sentimen ini bisa berupa opini, emosi, atau sikap terhadap suatu objek atau topik tertentu. Biasanya, analisis sentimen mengkategorikan teks menjadi sentimen

positif, negatif, atau netral, namun juga dapat dilakukan secara lebih rinci dengan mengelompokkan sentimen ke dalam berbagai tingkat seperti sangat positif, positif, netral, negatif, dan sangat negatif. Ada juga analisis sentimen yang fokus pada aspek tertentu dari suatu produk atau layanan, serta deteksi emosi spesifik seperti kebahagiaan, kemarahan, atau kesedihan.[4]

Analisis sentimen banyak diterapkan pada ulasan pelanggan, tanggapan survei, media sosial, dan materi lain untuk membantu bisnis dan organisasi memantau opini publik, meningkatkan layanan pelanggan, melakukan analisis produk, serta memahami posisi pelanggan dalam siklus pembelian. Teknik ini juga digunakan untuk memberikan respons personal dalam layanan pelanggan dan memantau reputasi merek secara real-time.

2.2. Natural Language Processing (NLP)

NLP adalah bidang dalam kecerdasan buatan yang bertujuan agar komputer dapat memahami dan menghasilkan bahasa manusia. NLP mencakup beberapa tahapan seperti *tokenization*, *stemming*, *stopword removal*, dan *text classification*. Pada penelitian berbasis Bahasa Indonesia, tantangan utama terletak pada keragaman ekspresi informal dan penggunaan singkatan yang sering muncul di platform media sosial.[5]

2.3. Pra-Pemrosesan Teks Bahasa Indonesia

Pra-pemrosesan data merupakan tahapan penting untuk memastikan teks berada dalam struktur yang dapat diolah oleh model *machine learning*. Tahapan tersebut mencakup *case folding*, pembersihan simbol, *tokenization*, *stopword removal*, serta *stemming* untuk mengembalikan kata ke bentuk dasar. Dalam Bahasa Indonesia, *stemming* umumnya dilakukan menggunakan library Sastrawi, karena model ini telah disesuaikan dengan morfologi Bahasa Indonesia.

2.4. TextBlob

TextBlob merupakan library Python yang menyediakan fungsi analisis sentimen berbasis pendekatan *lexicon-based*, dengan mengukur polaritas dan subjektivitas teks [6]. Walaupun metode ini praktis dan cepat digunakan, performanya cenderung kurang optimal untuk Bahasa Indonesia karena keterbatasan kamus sentiment yang digunakan.

2.5. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma supervised learning yang bekerja dengan membangun *hyperplane* untuk memisahkan kelas data secara optimal [6]. SVM dikenal memiliki performa tinggi pada klasifikasi teks karena mampu bekerja baik pada dataset kecil hingga menengah dan mendukung penggunaan *kernel* seperti RBF untuk kasus non-linear.

2.6. Penelitian Terdahulu

Peneliti	Metode	Dataset	Hasil	Gap
Larasakti (2023) [2]	KNN	YouTube random	Akurasi 61%	Tidak fokus pada pendidikan
Pramudita (2024) [5]	Naive Bayes	Twitter KIP Kuliah	>80%	Fokus kebijakan, bukan persepsi publik umum
Nurrohman (2025) [7]	SVM + TextBlob	YouTube	Akurasi 63%	Tidak memasukkan perspektif tokoh publik dalam analisis

2.7. Algoritma Naïve Bayes

Pendekatan *machine learning* dengan algoritma *Naive Bayes* banyak digunakan dalam analisis sentimen karena kemampuannya mengklasifikasikan teks berdasarkan probabilitas kemunculan kata. Metode ini bekerja dengan menghitung peluang suatu teks termasuk ke dalam kelas sentimen tertentu berdasarkan frekuensi kata yang muncul. Pendekatan ini dianggap efisien karena mampu memberikan hasil klasifikasi yang baik meskipun menggunakan data berukuran besar dan tidak memerlukan waktu komputasi yang lama [4].

Model *Naive Bayes* dipadukan dengan ekstraksi fitur *Term Frequency–Inverse Document Frequency (TF-IDF)* untuk menentukan bobot kata pada setiap dokumen. Hasil penelitian menunjukkan tingkat akurasi di atas 80% dengan dominasi sentimen positif terhadap kebijakan tersebut, yang membuktikan efektivitas *Naive Bayes* dalam mengklasifikasikan opini publik terhadap isu pendidikan nasional [7].

2.8. Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah algoritma supervised learning yang umum digunakan untuk tugas klasifikasi, termasuk klasifikasi teks pada analisis sentimen. Secara dasar, SVM mencari sebuah *hyperplane* di ruang fitur yang memisahkan dua kelas dengan margin maksimal sehingga model memiliki kemampuan generalisasi yang baik terhadap data baru [8].

3. METODE PENELITIAN

3.1. Pengumpulan Dataset (Crawling Data)

Data dikumpulkan dari komentar publik pada video YouTube yang berjudul “Kenapa Perlu Kuliah ala Maudy, andhika, & Amanda | Mata Najwa” Pengambilan data dilakukan menggunakan *YouTube Data API v3* dengan bantuan pustaka *googleapiclient* di *Python Comment Downloader* dengan script Python yang menyimpan data ke format CSV. Data yang didapat memuat 1 kolom yaitu “Full Text Coment”.

3.2. Pra-Pemrosesan Data

Pra-pemrosesan merupakan tahap krusial dalam analisis sentimen. Langkah pertama yang dilakukan pada data teks adalah **Case Folding**, yakni proses konversi

seluruh karakter ke dalam bentuk huruf kecil sebagai upaya standarisasi data sebelum tahap analisis. Langkah kedua adalah **Cleansing**, yaitu penghapusan tanda baca, angka, serta karakter khusus. Langkah ketiga adalah **Tokenisasi**, yaitu memecah teks ulasan menjadi satuan kata atau token. Langkah keempat adalah **Stopword**, yaitu bertujuan untuk menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis data. Tahap yang terakhir adalah Stemming, yaitu mengembalikan setiap kata ke bentuk dasarnya. [9]

3.3. Labeling Data

Data komentar YouTube diberi label Positif, Netral dan Negatif menggunakan **TextBlob** dilakukan dengan mengukur skor polaritas (**polarity score**) yang dihasilkan dari analisis sentimen teks.

- A. Jika polaritas >0, label komentar sebagai positif
- B. Jika polaritas =0, label komentar sebagai netral
- C. Jika polaritas <0, label komentar sebagai negatif

Dalam analisis sentimen komentar YouTube menunjukkan bahwa TextBlob efektif untuk memberikan label sentimen awal secara otomatis yang selanjutnya bisa digunakan untuk pelatihan model lain seperti Support Vector Machine (SVM) guna meningkatkan akurasi klasifikasi[6].

3.4. Model Klasifikasi

Dua Algoritma klasifikasi yang digunakan untuk menganalisis sentimen Netizen pada kolom komentar YouTube Najwa Shihab antara lain:

A. Model Support Vector Machine(SVM)

SVM berfungsi untuk membangun hyperplane pemisah yang optimal, dengan margin selebarmungkin antara dua kelompok data yang berbeda kelas. Fungsi objektifnya dinyatakan sebagai:

$$\min \frac{1}{2} ||w||^2 \text{ dengan syarat } y_i(w \cdot x_i + b) \geq 1, \forall i$$

Dalam kasus non-linier, digunakan kernel Radial Basis Function(RBF)

$$K(x_1, x_j) = \exp(-\gamma ||x_i - x_j||^2)$$

B. Random Forest (RF)

Random Forest adalah metode endemle yang terdiri dari sejumlah pohon keputusan, dimana hasil prediksi ditentukan berdasarkan voting mayoritas dari seluruh pohon yang dihasilkan. Proses pembentukan pohon dilakukan secara acak baik dari sisi fitur maupun sampel data, sehingga mengurangi overfitting.

Prediksi akhir diperoleh dari voting mayoritas:[10]

$$\hat{y} = \text{mode}(\{h_1(x), h_2(x), \dots, h_n(x)\})$$

di mana $h_i(x)$ adalah prediksi dari pohon ke- i .

3.5. Evaluasi Model

Evaluasi model dilakukan untuk mengukur sejauh mana sistem klasifikasi sentimen yang dibangun mampu menghasilkan prediksi yang akurat. Dalam penelitian ini, proses evaluasi dilakukan menggunakan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik tersebut digunakan untuk menilai performa model *Support Vector Machine (SVM)* dalam membedakan tiga kategori sentimen, yakni positif, negatif, dan netral.

Secara matematis, metrik evaluasi dirumuskan sebagai berikut:

$$\begin{aligned} Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1 - Score &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned}$$

dengan TP (True Positive) menunjukkan jumlah prediksi benar pada kelas positif, TN (True Negative) adalah jumlah prediksi benar pada kelas negatif, FP (False Positive) adalah jumlah prediksi salah pada kelas positif, dan FN (False Negative) adalah jumlah prediksi salah pada kelas negatif.

4. HASIL DAN PEMBAHASAN

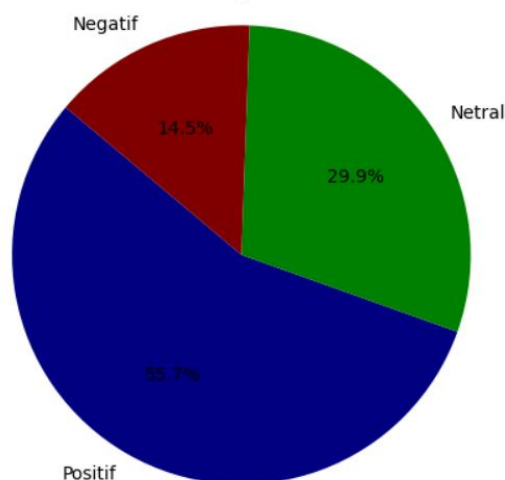
4.1. Statistik Data Ulasan

Data yang diperoleh dari kolom komentar YouTube melalui proses crawling data yang mencakup 415 komentar sentimen Netizen terhadap video “Kenapa Perlu Kuliah ala Maudy, andhika, & Amanda | Mata Najwa”. Komentar dilabeli menggunakan *TextBlob* berdasarkan skor yang dihasilkan oleh *Polaritas*. [10]

Tabel 1. Distribusi Hasil Labeling

Kategori Sentimen	Jumlah Komentar
Positif	231
Netral	124
Negatif	60
Total	415

Distribusi Komentar Negatif, Netral, dan Positif



Gambar 2. Distribusi Sentimen Komentar YouTube

Dari piechart tersebut bisa dilihat bahwa komentar dengan sentimen positif mendominasi, sedangkan komentar negatif memiliki presentase paling sedikit. Artinya, beberapa masyarakat sudah memiliki kesadaran bahwa kuliah itu juga diperlukan

4.2. Hasil Klasifikasi Menggunakan Naïve Bayes

Setelah pelabelan, dilakukan proses klasifikasi menggunakan algoritma Multinomial Naïve Bayes (MNB). Data dibagi menjadi 80% data latih dan 20% data uji. Hasil pengujian model ditampilkan pada Tabel 2 berikut.

Tabel 2. Hasil Evaluasi Model Naïve Bayes

Metrik Evaluasi	Nilai
Akurasi (Accuracy)	0.571
Precision Rata-rata	0.59
Recall Rata-rata	0.55
F1-Score Rata-rata	0.54

Nilai akurasi sebesar 57% menunjukkan bahwa model hanya mampu mengenali pola sentimen dengan tingkat ketepatan menengah. Meskipun tidak terlalu tinggi, hasil ini masih dapat diterima mengingat jumlah data relatif sedikit dan teks komentar bersifat informal serta tidak baku.

4.3. Laporan Klasifikasi Detail

Tabel 4. Laporan Klasifikasi

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	1,00	0.43	0.60	14
Netral	0.00	0.00	0.00	26
Positif	0.56	1.00	0.72	43

bermasalah, misalnya terkait biaya pendidikan, tantangan memperoleh pekerjaan setelah lulus, atau relevansi ilmu yang dipelajari dengan kebutuhan dunia kerja.

1. WordCloud positif, kata “kerja”, “kuliah”, dan “ilmu” tetap menjadi kata yang paling sering muncul. Namun, Pada komentar positif, kata-kata itu muncul dalam konteks yang menegaskan bahwa kuliah masih dianggap membawa manfaat, terutama terkait pengembangan diri dan kesempatan berkarier.
2. Sementara WordCloud negatif, muncul kata “akabri”, “militer”, “polisi” menandakan kritik terhadap aparat yang kurangnya pendidikan di Indonesia.

5. KESIMPULAN

Penelitian yang berjudul “Mengungkap Persepsi Publik terhadap Pendidikan Tinggi: Studi Analisis Sentimen Komentar YouTube Mata Najwa dengan Metode Support Vector Machine” memberikan gambaran mengenai bagaimana opini publik terbentuk terhadap topik pendidikan tinggi di ruang digital. Berdasarkan hasil analisis terhadap 415 komentar yang dikumpulkan dari video “*Kenapa Perlu Kuliah ala Maudy, Andhika, & Amanda | Mata Najwa*”, ditemukan bahwa sentimen positif mendominasi, diikuti sentimen netral dan negatif. Dominasi sentimen positif menunjukkan bahwa sebagian besar masyarakat memiliki pandangan yang mendukung pentingnya pendidikan tinggi sebagai sarana pengembangan diri, peningkatan kapasitas intelektual, dan pembentukan karakter yang lebih matang.

Hasil klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) menunjukkan tingkat akurasi sebesar 57%, yang dikategorikan sebagai performa sedang dalam klasifikasi teks informal berbahasa Indonesia. Nilai ini menunjukkan bahwa model mampu mengenali kecenderungan umum opini publik, meskipun masih menghadapi tantangan dalam mengidentifikasi komentar dengan sentimen ambigu atau netral. Berdasarkan hasil visualisasi, komentar positif banyak menyoroti nilai-nilai pentingnya pendidikan formal, seperti perluasan wawasan, kesempatan karier, dan pengembangan karakter. Sementara itu, komentar negatif lebih banyak membahas isu relevansi pendidikan tinggi dengan realitas dunia kerja, seperti ketimpangan kesempatan, biaya kuliah yang tinggi, serta persepsi bahwa kesuksesan tidak semata-mata ditentukan oleh gelar akademik.

Dari hasil penelitian ini dapat disimpulkan bahwa diskursus publik mengenai pendidikan tinggi tidak hanya berfokus pada manfaat akademis semata, tetapi juga mencerminkan pandangan sosial yang lebih luas tentang peran pendidikan dalam kehidupan modern. Sebagian masyarakat menilai kuliah sebagai kebutuhan fundamental untuk meningkatkan daya saing, sementara sebagian lain menilai pengalaman praktis dan kemampuan non-akademis justru lebih penting dalam dunia kerja saat ini. Perbedaan pandangan ini menunjukkan adanya dinamika sosial yang menarik dalam memahami makna pendidikan di era digital.

Secara metodologis, penggunaan algoritma *Support Vector Machine* (SVM) terbukti mampu memproses dan mengklasifikasikan opini publik dari data komentar yang bersifat tidak terstruktur. Kelebihan dari pendekatan ini adalah kemampuannya dalam mengidentifikasi pola sentimen yang kompleks tanpa memerlukan dataset berukuran besar. Namun demikian, penelitian ini masih memiliki keterbatasan dalam hal ukuran data, penggunaan bahasa tidak baku, serta belum memperhitungkan konteks sarkasme dan makna tersirat dalam teks. Faktor-faktor tersebut memengaruhi kinerja model dalam mengenali sentimen dengan akurasi tinggi.

Untuk penelitian selanjutnya, disarankan agar jumlah data diperluas dengan mengambil komentar dari beberapa video dengan topik serupa, sehingga hasil analisis menjadi lebih representatif dan general. Selain itu, penerapan model berbasis transformer, seperti *IndoBERT* atau *IndoBERTweet*, diharapkan dapat meningkatkan kemampuan klasifikasi karena model ini lebih mampu membaca makna kalimat serta nuansa emosional yang terkandung di dalamnya. Penambahan tahapan pra-pemrosesan yang lebih kompleks—seperti deteksi sarkasme, emoji, dan frase tidak baku—juga dapat meningkatkan ketepatan dalam analisis sentimen.

Secara keseluruhan, penelitian ini memberikan kontribusi terhadap pemahaman mengenai persepsi publik terhadap urgensi pendidikan tinggi di Indonesia. Temuan ini dapat menjadi dasar bagi lembaga pendidikan, pemerintah, dan media dalam merancang kebijakan serta komunikasi publik yang lebih efektif dan sesuai dengan kebutuhan masyarakat. Penerapan analisis sentimen berbasis *Support Vector Machine* (SVM). ini membuktikan bahwa data media sosial dapat dimanfaatkan sebagai indikator opini publik terhadap isu-isu pendidikan, sekaligus membuka peluang bagi pengembangan riset lanjutan di bidang analisis teks dan kebijakan pendidikan.

DAFTAR PUSTAKA

- [1] R. Nugrahani, "MANAJEMEN SUMBER DAYA MANUSIA PERPUSTAKAAN PERGURUAN TINGGI UNTUK MEWUJUDKAN SUMBER DAYA MANUSIA YANG BERKUALITAS," vol. 10, no. 2, Oct. 2017, Accessed: Nov. 26, 2025. [Online]. Available: <https://ejournal2.undip.ac.id/index.php/wp/article/view/1776>
- [2] D. N. Larasakti, A. Aziz, and D. Aditya, "Analisis Sentimen Komentar Video Youtube Dengan Metode K-Nearest Neighbor," *Jurnal Ilmiah Wahana Pendidikan*, vol. 2023, no. 5, pp. 132–142, doi: 10.5281/zenodo.7728573.
- [3] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *Jurnal Ilmiah Computer Science*, vol. 2, no. 2, pp. 89–98, Jan. 2024, doi: 10.58602/jics.v2i2.24.
- [4] D. N. Larasakti, A. Aziz, and D. Aditya, "Analisis Sentimen Komentar Video Youtube Dengan Metode K-Nearest Neighbor," *Jurnal Ilmiah Wahana Pendidikan*, vol. 2023, no. 5, pp. 132–142, doi: 10.5281/zenodo.7728573.
- [5] A. S. Rizkia, W. Wufron, and F. F. Roji, "Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan Lexicon-Based pada Performa

- IndoBERT,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, Jul. 2025, doi: 10.57152/malcom.v5i3.2151.
- [6] F. Wijaya, R. Aswi Ramadhani, and A. Bagus Setiawan, “Analisis Sentimen Komentar Video Youtube Dengan Leksikon Textblob Menggunakan Algoritma Svm Berdasarkan Rasio Data Uji,” *Jurnal Ilmiah Inotek*, vol. 7, no. 1, 2025, Accessed: Nov. 26, 2025. [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/>
- [7] D. Pramudita, Y. Akbar, and T. Wahyudi, “Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah pada Media Sosial X Menggunakan Algoritma Naive Bayes,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1420–1430, Aug. 2024, doi: 10.57152/malcom.v4i4.1565.
- [8] S. O. Aisyah, N. Fauziah, F. Sains, and D. Teknologi, “ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES CLASSIFIER DAN INSET LEXICON PADA TWITTER (Studi Kasus: Mie Gacoan) PROGRAM STUDI TEKNIK INFORMATIKA.”
- [9] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 2, Aug. 2021, doi: 10.30813/jbase.v4i2.3000.
- [10] R. Westerdam *et al.*, “Analisis Sentimen Aplikasi KA Bandara Menggunakan Metode Support Vector Machine dan Random Forest,” *Seminar Nasional Sains Data*, vol. 2025, Jun. 2025.