JIIFKOM (Jurnal Ilmiah Informatika & Komputer) STTR Cepu
E-Mail: jiifkom@sttrcepu.ac.id

Analisis Sentimen Opini Publik YouTube terhadap Pelantikan Menteri Keuangan Purbaya Menggunakan Labeling Kamus dan Decision Tree

Geovanda Kevin Dabinsa[a*], Adhika Pramita Widyassari[b]
siuuukevin72@gmail.com

Program Studi Informatika, STT Ronggolawe Cepu [a,b]

Kata kunci : Analisis Sentimen, YouTube, Pelabelan Kamus, Decision Tree, Pelantikan Menteri, Opini Publik

Intisari
Media sosial YouTube telah menjadi platform penting bagi masyarakat Indonesia untuk menyampaikan opini publik, termasuk respons terhadap peristiwa politik seperti pelantikan menteri kabinet. Penelitian ini bertujuan untuk menganalisis sentimen opini publik di YouTube terkait pelantikan Menteri Keuangan Purbaya pada 8 September 2025 dengan menggunakan pelabelan otomatis berbasis kamus sentimen bahasa Indonesia (InSet) dan menguji performa algoritma Decision Tree dibandingkan dengan algoritma klasifikasi lainnya. Metode yang digunakan meliputi pengumpulan 90 data komentar YouTube, preprocessing teks (cleaning, case folding, tokenizing, stopword removal, dan stemming), pelabelan otomatis berbasis kamus, ekstraksi fitur TF-IDF, serta klasifikasi menggunakan enam algoritma: Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes, Neural Network, dan Deep Learning. Hasil pelabelan menunjukkan distribusi sentimen: 82,2% netral, 13,3% positif, dan 4,4% negatif. Hasil evaluasi model menunjukkan bahwa Decision Tree mencapai akurasi sempurna (100%) pada data uji, mengungguli algoritma lain yang cenderung mengalami bias akibat ketidakseimbangan kelas. Penelitian ini menyimpulkan bahwa kombinasi pelabelan berbasis kamus dan algoritma Decision Tree efektif untuk analisis sentimen pada dataset terbatas di platform YouTube, serta memberikan wawasan bahwa respons publik terhadap pelantikan Menteri Keuangan Purbaya didominasi oleh sentimen netral.

Tentang naskah:
-diterima, 15 Januari 2026
-diterbitkan, 26 Januari 2026

Keywords : Sentiment Analysis, YouTube, Dictionary Labeling, Decision Tree, Ministerial Appointment, Public Opinion

Abstract
YouTube social media has become an important platform for Indonesians to express public opinion, including responses to political events such as cabinet minister inaugurations. This study aims to analyze public sentiment on YouTube regarding the inauguration of Finance Minister Purbaya on September 8, 2025, using automatic labeling based on the Indonesian sentiment dictionary (InSet) and testing the performance of the Decision Tree algorithm compared to other classification algorithms. The methods used include collecting 90 YouTube comment data, text preprocessing (cleaning, case folding, tokenizing, stopword removal, and stemming), dictionary-based automatic labeling, TF-IDF feature extraction, and classification using six algorithms: Decision Tree, Support Vector Machine (SVM), K -Nearest Neighbor (KNN), Naïve Bayes, Neural Network, and Deep Learning. The labeling results show the sentiment distribution: 82.2% neutral, 13.3% positive, and 4.4% negative. Model evaluation results show that Decision Tree achieved perfect accuracy (100%) on the test data, outperforming other algorithms that tended to experience bias due to class imbalance. This study concludes that the combination of dictionary-based labeling and the Decision Tree algorithm is effective for sentiment analysis on limited datasets on the YouTube platform, and provides insight that public response to the inauguration of Finance Minister Purbaya was dominated by neutral sentiment.

1. Pendahuluan

Media sosial YouTube telah berkembang menjadi salah satu platform utama bagi masyarakat Indonesia dalam menyampaikan opini publik, terutama terkait isu-isu politik dan kebijakan pemerintah. Setiap peristiwa penting, seperti pelantikan menteri kabinet, kerap memicu respons yang beragam dari publik yang tercermin dalam kolom komentar YouTube (Yasir & Suraji, 2023). Data yang dihasilkan dari interaksi ini merupakan sumber yang kaya untuk memahami kecenderungan sentimen masyarakat secara real-time (Aribowo et al., 2021). Analisis sentimen merupakan salah satu cabang Natural Language Processing (NLP) yang bertujuan mengklasifikasikan opini menjadi kategori positif, negatif, atau netral berdasarkan konten teks. Dalam konteks YouTube, analisis ini menghadapi tantangan tersendiri karena data teks yang

dihasilkan sering kali tidak terstruktur, mengandung bahasa informal, slang, dan emotikon (Armaeni et al., 2024). Oleh karena itu, diperlukan pendekatan preprocessing yang komprehensif dan pemilihan algoritma klasifikasi yang tepat.

Berbagai algoritma machine learning telah diuji untuk analisis sentimen di platform YouTube. Decision Tree menjadi salah satu algoritma yang sering digunakan karena kemampuannya dalam menangani data tidak terstruktur, mudah diinterpretasi, dan menunjukkan performa yang kompetitif. Penelitian oleh (Prasetyo et al., 2025) yang berjudul "Analisis Sentimen Pengguna Seputar Kendaraan Listrik Di Twitter Dengan Penerapan Algoritma Naïve Bayes, KNN, dan Decision Tree untuk Klasifikasi" menunjukkan bahwa Decision Tree mampu mencapai akurasi yang tinggi dalam klasifikasi sentimen topik terkait kebijakan publik di media sosial. Selain itu, penelitian lain juga mengonfirmasi efektivitas Decision Tree dalam analisis sentimen di YouTube, seperti pada studi mengenai penutupan TikTok Shop (Armaeni et al., 2024) dan pelaksanaan Pilkada 2024 (Safra & Zuliarso, 2025).

Di sisi lain, pendekatan pelabelan otomatis berbasis kamus sentimen (lexicon-based) telah banyak digunakan sebagai solusi efisien untuk mengurangi ketergantungan pada pelabelan manual yang memakan waktu dan subjektif. Pendekatan ini terbukti efektif dalam memberikan label awal pada data teks sebelum diklasifikasikan lebih lanjut dengan algoritma machine learning (Aribowo et al., 2021). Namun, penelitian yang secara khusus mengombinasikan pelabelan otomatis berbasis kamus dengan algoritma Decision Tree untuk menganalisis opini publik di YouTube terkait pelantikan menteri kabinet masih sangat terbatas.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan dengan tujuan untuk menganalisis sentimen opini publik di YouTube terkait pelantikan Menteri Keuangan Purbaya pada 8 September 2025 dengan menggunakan pelabelan otomatis berbasis kamus sentimen bahasa Indonesia dan menguji performa beberapa algoritma klasifikasi, dengan fokus pada efektivitas Decision Tree dibandingkan dengan algoritma lain seperti SVM, KNN, Naïve Bayes, Neural Network, dan Deep Learning. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan metode analisis sentimen yang efisien untuk platform YouTube serta memberikan wawasan mengenai respons masyarakat terhadap suatu peristiwa politik penting.

2. Kerangka Teori

2.1 Analisis Sentimen dan Opini Publik di Media Sosial

Analisis sentimen merupakan bidang studi dalam Natural Language Processing (NLP) yang berfokus pada pengidentifikasian, ekstraksi, dan kategorisasi opini, emosi, serta sikap yang terekspresikan dalam data teks (Liu, 2020). Dalam konteks media sosial, analisis ini menjadi alat penting untuk memahami opini publik secara real-time, mengingat platform seperti Twitter, YouTube, dan Facebook telah menjadi ruang diskusi utama masyarakat untuk menanggapi berbagai isu sosial, politik, dan ekonomi (Wibowo, 2024).

Opini publik di media sosial sering kali muncul dalam bentuk komentar, atau post yang tidak terstruktur, sehingga memerlukan pendekatan komputasional untuk mengolah dan menganalisisnya. Menurut (Saputra et al., 2025), analisis sentimen memungkinkan pengklasifikasian opini menjadi kategori seperti positif, negatif, atau netral, yang dapat digunakan untuk memetakan kecenderungan umum masyarakat terhadap suatu topik. Hal ini sejalan dengan penelitian (Wibowo, 2024) yang membandingkan algoritma SVM, KNN, dan Naïve Bayes untuk menganalisis sentimen terkait popularitas calon presiden di Twitter, menunjukkan bahwa analisis sentimen dapat memberikan gambaran objektif mengenai opini publik di ranah digital.

Selain itu, (Aribowo et al., 2021) menekankan bahwa analisis sentimen tidak hanya berguna untuk kepentingan akademis, tetapi juga bagi pemerintah dan pihak terkait dalam mengevaluasi respons masyarakat terhadap suatu kebijakan. Dengan menganalisis sentimen dari media sosial, pemerintah dapat memperoleh umpan balik yang cepat dan representatif, sehingga dapat digunakan sebagai bahan pertimbangan dalam perumusan atau evaluasi kebijakan.

Secara umum, analisis sentimen telah menjadi pendekatan yang semakin relevan seiring dengan meningkatnya volume dan pengaruh opini digital dalam membentuk wacana publik.

2.2 Platform YouTube sebagai Sumber Data Opini Publik

YouTube telah berkembang menjadi salah satu platform media sosial dengan pengguna aktif terbesar di Indonesia, sekaligus menjadi sumber data opini publik yang kaya dan real-time (Aribowo et al., 2021). Platform ini memungkinkan pengguna tidak hanya menonton konten video, tetapi juga berinteraksi melalui kolom komentar, yang sering kali berisi ekspresi pendapat, kritik, dukungan, atau tanggapan emosional terhadap suatu isu (Armaeni et al., 2024).

Karakteristik data komentar YouTube memiliki beberapa kekhususan, di antaranya: teks yang tidak terstruktur, penggunaan bahasa informal, slang, singkatan, emotikon, serta variasi ejaan yang tidak (Yasir & Suraji, 2023). Hal ini menjadi tantangan tersendiri dalam tahap preprocessing, namun juga memberikan cerminan yang lebih autentik terhadap opini masyarakat dibandingkan dengan platform berbasis teks formal.

Menurut (Aribowo et al., 2021), sekitar 60% komentar YouTube dalam bahasa Indonesia mengandung slang words, dan sekitar 27-28% mengandung stopwords. Oleh karena itu, penelitian yang berfokus pada analisis sentimen di YouTube perlu menerapkan teknik preprocessing yang komprehensif, termasuk pembersihan stopwords dan konversi slang words, untuk meningkatkan akurasi klasifikasi.

Beberapa penelitian telah memanfaatkan YouTube sebagai sumber data utama dalam analisis sentimen kebijakan publik di Indonesia. Misalnya, (Yasir & Suraji, 2023) menganalisis sentimen terkait kenaikan biaya haji 2023, sementara (Armaeni et al., 2024) meneliti respons masyarakat terhadap penutupan TikTok Shop. Keduanya menunjukkan bahwa YouTube dapat menjadi sumber data yang valid untuk memahami sentimen publik secara kuantitatif.

Dengan demikian, YouTube tidak hanya berperan sebagai platform hiburan, tetapi juga sebagai ruang publik digital yang merekam dan merefleksikan opini masyarakat Indonesia secara langsung dan spontan.

2.3 Metode Preprocessing untuk Teks Bahasa Indonesia

Preprocessing teks merupakan tahapan kritis dalam analisis sentimen, terutama untuk data teks tidak terstruktur seperti komentar YouTube. Tujuan utamanya adalah untuk membersihkan, menyeragamkan, dan mengubah data teks mentah menjadi format yang siap untuk diekstraksi fitur dan diklasifikasi (Aribowo et al., 2021). Pada teks berbahasa Indonesia, preprocessing menjadi semakin kompleks karena adanya karakteristik seperti kata berimbuhan, slang, singkatan, dan variasi dialek (Armaeni et al., 2024).

Beberapa tahapan preprocessing yang umum diterapkan meliputi:

- Cleaning: menghapus karakter tidak relevan seperti URL, emotikon, tagar, dan tanda baca yang tidak diperlukan (Safra & Zuliarso, 2025).
- Case Folding: mengubah seluruh teks menjadi huruf kecil untuk menghindari perbedaan akibat kapitalisasi.
- Tokenizing: memecah kalimat menjadi unit kata (token) (Yasir & Suraji, 2023).
- Stopword Removal: menghapus kata-kata umum yang tidak memiliki nilai informasi untuk analisis sentimen (contoh: "yang", "dan", "di", "pada").
- Stemming: mengembalikan kata berimbuhan ke bentuk dasarnya menggunakan algoritma seperti Sastrawi untuk bahasa Indonesia (Safra & Zuliarso, 2025).
- Normalization dan Slang Word Conversion: mengubah kata tidak baku atau slang menjadi bentuk standar sesuai Kamus Besar Bahasa Indonesia (KBBI).

Menurut (Aribowo et al., 2021), penghapusan stopword dan konversi slang word dapat meningkatkan akurasi model hingga 3–5% dalam konteks analisis sentimen lintas domain. Hal ini juga didukung oleh (Armaeni et al., 2024) yang menyatakan bahwa pembersihan data yang baik dapat mengurangi noise dan

meningkatkan konsistensi fitur yang dihasilkan.

Dalam penelitian (Safra & Zuliarso, 2025), preprocessing yang teliti terbukti membantu algoritma Decision Tree dalam menangani data komentar YouTube yang beragam, sehingga menghasilkan akurasi klasifikasi yang lebih stabil.

Dengan demikian, pemilihan dan penerapan metode preprocessing yang tepat sangat berpengaruh terhadap kualitas hasil analisis sentimen, terutama untuk data berbahasa Indonesia yang kaya akan variasi linguistik.

2.4 Pelabelan Otomatis Berbasis Kamus Sentimen (Lexicon-Based)

Pelabelan otomatis berbasis kamus sentimen (lexicon-based sentiment labeling) merupakan pendekatan yang digunakan untuk memberikan label sentimen (seperti positif, negatif, atau netral) pada data teks secara otomatis dengan memanfaatkan daftar kata yang telah diberi skor polaritas (Kardian & Gustiana, 2021). Pendekatan ini menjadi alternatif efisien untuk menggantikan atau melengkapi pelabelan manual, yang seringkali memakan waktu, subjektif, dan tidak konsisten, terutama ketika menghadapi data dalam jumlah besar (Gede Bagus Resa Pratama Putra et al., 2025).

Kamus sentimen (sentiment lexicon) untuk bahasa Indonesia umumnya berisi kata-kata beserta bobot sentimennya, misalnya kata "bagus" memiliki skor positif, sedangkan "buruk" memiliki skor negatif. Beberapa kamus yang umum digunakan antara lain InSet (Indonesian Sentiment Lexicon) dan kamus kustom yang dikembangkan berdasarkan domain penelitian tertentu (Aribowo et al., 2021). Proses pelabelan dilakukan dengan menghitung total skor sentimen dalam sebuah kalimat atau dokumen berdasarkan kata-kata yang muncul dalam kamus.

(Kardian & Gustiana, 2021) dalam penelitiannya mengenai analisis sentimen terhadap BPJS di Twitter, berhasil mengombinasikan metode lexicon-based dengan algoritma Naïve Bayes. Hasilnya menunjukkan bahwa penggunaan kamus sentimen dapat meningkatkan akurasi dan konsistensi pelabelan data latih. Studi serupa oleh (Gede Bagus Resa Pratama Putra et al., 2025) yang menganalisis sentimen terkait RUU TNI di media sosial X juga membuktikan efektivitas lexicon sebagai alat pra-pelabelan sebelum klasifikasi lebih lanjut dengan metode machine learning.

Namun, pendekatan lexicon-based juga memiliki keterbatasan, terutama dalam menangani konteks kalimat, ironi, sarkasme, dan kata-kata yang memiliki makna ganda (ambiguous). Oleh karena itu, seringkali lexicon-based dikombinasikan dengan metode lain, seperti machine learning atau deep learning, untuk mendapatkan hasil yang lebih akurat (Aribowo et al., 2021).

Dalam konteks penelitian ini, pelabelan otomatis berbasis kamus digunakan sebagai tahap awal untuk memberikan label pada data komentar YouTube sebelum diklasifikasikan dengan algoritma Decision Tree dan algoritma pembanding lainnya. Pendekatan ini diharapkan dapat mempercepat proses pelabelan dan menjaga objektivitas dalam penentuan sentimen.

2.5 Decision Tree

Decision Tree adalah salah satu algoritma supervised learning yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan membangun struktur pohon keputusan (tree-like model) yang terdiri dari root node, internal node (decision node), dan leaf node (Safra & Zuliarso, 2025). Setiap node merepresentasikan pengujian terhadap suatu fitur, setiap cabang merepresentasikan hasil pengujian, dan setiap leaf node merepresentasikan kelas atau keputusan akhir (seperti sentimen positif, negatif, atau netral).

Kelebihan utama Decision Tree adalah mudah diinterpretasi, tidak memerlukan normalisasi data, serta mampu menangani data numerik dan kategorikal (Prasetyo et al., 2025). Selain itu, Decision Tree relatif robust terhadap missing value dan cocok untuk dataset yang tidak terlalu besar (Armaeni et al., 2024). Namun, algoritma ini rentan terhadap overfitting, terutama jika kedalaman pohon (max depth) tidak dibatasi dengan tepat.

Dalam konteks analisis sentimen di media sosial, Decision Tree telah terbukti memberikan performa yang kompetitif.

(Prasetyo et al., 2025) dalam penelitiannya mengenai analisis sentimen topik kendaraan listrik di Twitter, membandingkan Decision Tree dengan Naïve Bayes dan KNN. Hasil penelitian menunjukkan bahwa Decision Tree mampu mencapai akurasi yang tinggi dan konsisten, bahkan pada dataset yang tidak seimbang (imbalanced data).

Studi lain oleh (Armaeni et al., 2024) yang membandingkan Decision Tree dengan Naïve Bayes pada komentar YouTube terkait penutupan TikTok Shop juga menemukan bahwa Decision Tree menghasilkan akurasi yang lebih stabil pada berbagai skenario pembagian data (80:20, 70:30, 60:40). Sementara itu, (Safra & Zuliarso, 2025) berhasil mengimplementasikan Decision Tree untuk analisis sentimen Pilkada 2024 di YouTube dengan akurasi 79,72%, meskipun ditemukan keterbatasan dalam memprediksi kelas netral.

Berdasarkan temuan-temuan tersebut, Decision Tree dipilih dalam penelitian ini sebagai salah satu algoritma utama, mengingat kemampuannya dalam menangani data komentar YouTube yang berukuran terbatas (90 data) serta kemudahan dalam interpretasi hasil klasifikasi.

2.6 Algoritma Pembanding dalam Analisis Sentimen

Selain Decision Tree, terdapat beberapa algoritma klasifikasi lain yang umum digunakan dalam analisis sentimen, seperti Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes, Neural Network, dan Deep Learning. Setiap algoritma memiliki keunggulan dan keterbatasan tersendiri, sehingga sering dibandingkan untuk menentukan metode terbaik sesuai dengan karakteristik data.

Support Vector Machine (SVM) dikenal karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan hyperplane yang optimal untuk pemisahan kelas. (Saputra et al., 2025) menunjukkan bahwa SVM mampu memberikan performa yang konsisten dalam klasifikasi sentimen di berbagai platform media sosial. Namun, SVM cenderung memerlukan waktu komputasi yang lebih lama pada dataset besar dan kurang efektif jika data tidak terpisah secara linear.

K-Nearest Neighbor (KNN) merupakan algoritma sederhana yang bekerja berdasarkan kedekatan jarak antar data. (Halimi & Rudyanto Arief, 2021) berhasil mengombinasikan KNN dengan lexicon dalam analisis sentimen pembelajaran online di Twitter. Meskipun mudah diimplementasikan, KNN rentan terhadap ketidakseimbangan kelas dan kinerjanya sangat bergantung pada pemilihan nilai k.

Naïve Bayes adalah algoritma probabilistik yang mengasumsikan independensi antar fitur. Algoritma ini sering digunakan karena kecepatannya dan performa yang baik pada data teks. (Mario Conroy Paridy Man et al., 2025) menggunakan Naïve Bayes untuk menganalisis sentimen terkait pembangunan IKN di media sosial X, dan mendapatkan hasil yang cukup akurat. Namun, asumsi independensi fitur sering tidak terpenuhi pada data dunia nyata, sehingga dapat mempengaruhi akurasi.

Neural Network dan Deep Learning merupakan pendekatan yang semakin populer karena kemampuannya menangkap pola kompleks dan hubungan kontekstual dalam teks. Namun, metode ini umumnya memerlukan data dalam jumlah besar dan sumber daya komputasi yang tinggi. Pada dataset yang terbatas, performa deep learning sering kali tidak optimal (Armaeni et al., 2024).

Berdasarkan perbandingan tersebut, pemilihan algoritma sangat bergantung pada ukuran dataset, karakteristik data, dan tujuan interpretasi. Dalam penelitian ini, beberapa algoritma tersebut akan dibandingkan dengan Decision Tree untuk mengevaluasi performa masing-masing pada data komentar YouTube yang berjumlah terbatas.

2.7 Studi Terkait Analisis Sentimen Kebijakan Publik di Indonesia

Analisis sentimen terhadap kebijakan publik di Indonesia telah banyak dilakukan dengan memanfaatkan data dari media sosial seperti Twitter, Facebook, dan YouTube. Studi-studi ini umumnya bertujuan untuk memahami respons masyarakat terhadap suatu kebijakan, mengidentifikasi isu-isu dominan, dan memberikan rekomendasi berdasarkan temuan sentimen.

Salah satu contoh adalah penelitian (Yasir & Suraji, 2023) yang menganalisis sentimen masyarakat terkait kenaikan

biaya haji 2023 di YouTube. Penelitian ini membandingkan tiga algoritma—Naïve Bayes, Decision Tree, dan Random Forest—dan menemukan bahwa Naïve Bayes memberikan akurasi tertinggi (90%), diikuti Random Forest (87%) dan Decision Tree (83%). Temuan ini menunjukkan bahwa meskipun Decision Tree tidak selalu menjadi yang terbaik, algoritma ini tetap memberikan performa yang kompetitif dan mudah diinterpretasi.

Studi lain oleh (Safra & Zuliarso, 2025) berfokus pada pelaksanaan Pilkada 2024 di YouTube menggunakan Decision Tree. Hasil penelitian menunjukkan akurasi sebesar 79,72% dengan sentimen dominan negatif, yang mencerminkan ketidakpuasan publik terhadap dinamika politik yang terjadi. Penelitian ini juga mengonfirmasi bahwa Decision Tree cocok digunakan untuk dataset dengan jumlah data yang tidak terlalu besar.

Selain itu, (Prasetyo et al., 2025) meneliti sentimen masyarakat terkait kendaraan listrik di Twitter dengan membandingkan Naïve Bayes, KNN, dan Decision Tree. Decision Tree menunjukkan performa yang baik dan stabil, terutama dalam menangani variasi fitur teks yang diekstraksi menggunakan TF-IDF.

Penelitian-penelitian di atas menunjukkan bahwa analisis sentimen berbasis media sosial telah menjadi pendekatan yang valid dan banyak digunakan untuk mengevaluasi respons publik terhadap kebijakan di Indonesia. Namun, masih terdapat celah penelitian, terutama dalam mengombinasikan pelabelan otomatis berbasis kamus dengan Decision Tree untuk analisis sentimen di YouTube terkait pelantikan menteri kabinet.

2.8 Gap Penelitian

Berdasarkan tinjauan pustaka yang telah diuraikan, dapat diidentifikasi beberapa celah (gap) penelitian yang menjadi landasan bagi studi ini:

2.8.1 Kombinasi Lexicon-Based Labeling dengan Decision Tree di Platform YouTube

Meskipun penggunaan lexicon-based labeling dan Decision Tree masing-masing telah banyak diteliti secara terpisah, penelitian yang secara spesifik menggabungkan keduanya untuk analisis sentimen di platform YouTube masih terbatas. Kebanyakan studi menggunakan lexicon hanya sebagai alat bantu pra-pemrosesan atau pelabelan awal, tanpa mengevaluasi efektivitasnya secara eksplisit dalam kerangka klasifikasi Decision Tree (Aribowo et al., 2021; Kardian & Gustiana, 2021).

2.8.2 Fokus pada Topik Pelantikan Menteri Kabinet

Studi-studi terdahulu banyak berfokus pada isu kebijakan publik seperti biaya haji (Yasir & Suraji, 2023), penutupan TikTok Shop (Armaeni et al., 2024), atau Pilkada (Safra & Zullarso, 2025). Namun, analisis sentimen terhadap pelantikan menteri kabinet—khususnya Menteri Keuangan—masih jarang dieksplorasi, padahal momen ini sering kali mendapatkan respons publik yang signifikan di media sosial.

2.8.3 Dataset Kecil dengan Pendekatan Decision Tree

Sebagian besar penelitian menggunakan dataset dalam jumlah besar (ratusan hingga ribuan data). Namun, efektivitas Decision Tree pada dataset kecil (seperti 90 data) belum banyak dibahas secara mendalam, terutama dalam konteks analisis sentimen YouTube yang cenderung memiliki noise linguistik yang tinggi.

2.8.4 Perbandingan Multi-Algoritma pada Dataset Terbatas

Studi komparatif algoritma klasifikasi seperti SVM, KNN, Naïve Bayes, Neural Network, dan Deep Learning umumnya dilakukan pada dataset besar. Evaluasi performa algoritma-algoritma tersebut pada dataset terbatas—seperti dalam penelitian ini—masih jarang dilakukan, terutama dalam konteks labeling otomatis berbasis kamus.

Dengan demikian, penelitian ini bertujuan untuk mengisi celah-celah tersebut dengan:

- Mengimplementasikan pelabelan otomatis berbasis kamus sentimen bahasa Indonesia sebagai tahap awal.
- Menguji dan membandingkan performa Decision Tree dengan algoritma lain (SVM, KNN, Naïve

Bayes, Neural Network, Deep Learning) pada dataset komentar YouTube yang terbatas.

- Menganalisis sentimen publik terhadap pelantikan Menteri Keuangan Purbaya sebagai studi kasus yang kontekstual dan aktual.

Tinjauan pustaka ini telah menguraikan landasan teoritis dan empiris terkait analisis sentimen, platform YouTube, preprocessing teks bahasa Indonesia, pelabelan berbasis kamus, algoritma Decision Tree dan pembandingnya, serta studi-studi terkait kebijakan publik di Indonesia. Identifikasi gap penelitian menunjukkan bahwa masih ada ruang untuk pengembangan metode yang lebih efisien dan akurat, khususnya dalam konteks dataset terbatas dan topik yang spesifik seperti pelantikan menteri. Bab selanjutnya akan menjabarkan metodologi yang digunakan untuk menjawab tantangan penelitian ini.

3. Metodologi

3.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan tujuan menguji dan membandingkan performa berbagai algoritma klasifikasi dalam menganalisis sentimen opini publik di YouTube. Alur penelitian mengikuti tahapan sistematis yang mencakup pengumpulan data, preprocessing, pelabelan otomatis berbasis kamus, ekstraksi fitur, klasifikasi, dan evaluasi model.

3.2. Diagram Alir Penelitian (Flowchart)

Berikut adalah diagram alir yang menggambarkan tahapan penelitian secara visual:



Gambar 1. Diagram Flowchart

3.3. Tahapan Penelitian

3.3.1. Pengumpulan Data

Data diperoleh dari komentar YouTube pada video pelantikan Menteri Keuangan Purbaya tanggal 8 September 2025. Teknik web scraping menggunakan YouTube Data API v3 diterapkan dengan bantuan Python dan Google Colab. Sebanyak 90 komentar berhasil dikumpulkan dan disimpan dalam format CSV untuk proses lebih lanjut.

3.3.2. Preprocessing Teks

Tahapan preprocessing meliputi:

- Cleaning: menghapus karakter tidak relevan seperti URL, emotikon, tagar, dan tanda baca yang tidak diperlukan (Safra & Zuliarso, 2025).
- Case Folding: Mengubah semua huruf menjadi lowercase.
- Tokenizing: memecah kalimat menjadi unit kata (token) (Yasir & Suraji, 2023).
- Stopword Removal: Menghilangkan kata penghubung yang tidak bermakna.

- Stemming: mengembalikan kata berimbuhan ke bentuk dasarnya menggunakan algoritma seperti Sastrawi untuk bahasa Indonesia (Safra & Zuliarso, 2025).

3.3.3. Pelabelan Otomatis Berbasis Kamus Sentimen

Pelabelan sentimen dilakukan secara otomatis menggunakan kamus sentimen bahasa Indonesia (InSet). Setiap komentar diberi skor sentimen berdasarkan kata-kata yang terdapat dalam kamus, kemudian diklasifikasikan ke dalam tiga kategori:

- Positif jika skor > 0
- Netral jika skor = 0
- Negatif jika skor < 0

3.3.4. Ekstraksi Fitur dengan TF-IDF

Fitur teks diubah menjadi representasi numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan seluruh korpus.

3.3.5. Pembagian Data (Data Splitting)

Dataset dibagi menjadi dua bagian:

- Data latih (training): 80% (72 data)
- Data uji (testing): 20% (18 data)

Pembagian ini bertujuan untuk menguji kemampuan generalisasi model pada data yang belum pernah dilihat.

3.3.6. Pemodelan Klasifikasi

Ada 6 (Enam) algoritma klasifikasi digunakan untuk membandingkan performa:

Decision Tree : Decision Tree menjadi salah satu algoritma yang sering digunakan karena kemampuannya dalam menangani data tidak terstruktur, mudah diinterpretasi, dan menunjukkan performa yang kompetitif. Penelitian oleh (Prasetyo et al., 2025)

Support Vector Machine (SVM) : metode ini dikenal karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan hyperplane yang optimal untuk pemisahan kelas.

K-Nearest Neighbor (KNN) : K-Nearest Neighbor (KNN) merupakan algoritma sederhana yang bekerja berdasarkan kedekatan jarak antar data.

Naïve Bayes : Algoritma Naïve Bayes merupakan salah satu metode machine learning yang sangat populer untuk analisis sentimen, terutama karena kesederhanaannya. Disebut “naive” (naif) karena algoritma ini mengasumsikan bahwa setiap kata dalam teks adalah independen atau tidak saling bergantung satu sama lain, meskipun dalam kenyataannya kata-kata sering kali memiliki korelasi (Pramita Widyassari et al., 2025)

Neural Network dan Deep Learning : Neural Network dan Deep Learning merupakan pendekatan yang semakin populer karena kemampuannya menangkap pola kompleks dan hubungan kontekstual dalam teks. Namun, metode ini umumnya memerlukan data dalam jumlah besar dan sumber daya komputasi yang tinggi. Pada dataset yang terbatas, performa deep learning sering kali tidak optimal (Armaeni et al., 2024).

Setiap model dilatih menggunakan data latih yang sama.

3.3.7. Evaluasi Model

Evaluasi dilakukan menggunakan confusion matrix dengan metrik:

- Akurasi (Accuracy)
- Presisi (Precision)
- Recall (Sensitivity)
- F1-Score

Perbandingan performa antar algoritma akan dianalisis untuk menentukan model terbaik.

3.4. Alat dan Teknologi yang Digunakan

Ada 5 Alat dan Teknologi yang Digunakan yaitu :

- Bahasa Pemrograman: Python 3.9
- Libraries: Pandas, NumPy, Scikit-learn, NLTK, Sastrawi, TensorFlow/Keras
- Platform: Google Colab
- Kamus Sentimen: InSet (Indonesian Sentiment Lexicon)
- Visualisasi: Matplotlib, Seaborn

4. Hasil dan Pembahasan

4.1 Crawling Data

Crawling data menggunakan API youtube dengan tools "google cloud console", dilantik pada tanggal dengan 8 September 2025 dengan tagar: #PelantikanPurbaya, #menterikeungan, #Purbaya, diperoleh data sejumlah 90 komentar. **Gambar 2** berikut ini tampilan data teks komentar yang diperoleh atau data mentah hasil dari crawling data.



	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location
0	Ugz_43XNz7hCbLQB_XZ4AaABAg	2025-09-09T10:46:55Z	0	Kami doakan semoga negeri Indonesia menjadi le...	Ugz_43XNz7hCbLQB_XZ4AaABAg	None	None	und	None
1	UgxWaYUhvkkbzlRJay94AaABAg	2025-09-09T10:45:49Z	0	Agak ragu aku dgn Mentri ke uangan kita yg ini...	UgxWaYUhvkkbzlRJay94AaABAg	None	None	und	None
2	UgzlCfv9YwobmbpjKe4AaABAg	2025-09-09T09:47:21Z	0	Hayu luhut ganti bahlul ganti zulkipli ganti e...	UgzlCfv9YwobmbpjKe4AaABAg	None	None	und	None
3	UgxQiJn5tNTAEfC1wCh4AaABAg	2025-09-09T80:17:09Z	0	Coba angkat kembali pak Mahfud MD, pak Dedi Mu...	UgxQiJn5tNTAEfC1wCh4AaABAg	None	None	und	None
4	Ugyg6AEXdC-n9Ia1qFB4AaABAg	2025-09-08T14:19:41Z	0	Gara-Gara Amarika Syarikat(USA) Kawal Ekonomi ...	Ugyg6AEXdC-n9Ia1qFB4AaABAg	None	None	und	None
5	UgxcnuZL5SPLGI0LfKN4AaABAg	2025-09-08T12:10:30Z	0	Bu' Sri sakit hati.. rumahnya kalian jarah_jad...	UgxcnuZL5SPLGI0LfKN4AaABAg	None	None	und	None
6	Ugwe6p8bmcQ_pN5V6Zf4AaABAg	2025-09-08T12:04:17Z	0	Sudah tepat Bu Sri diganti, karena sudah tidak...	Ugwe6p8bmcQ_pN5V6Zf4AaABAg	None	None	und	None

Gambar 2. Data teks komentar / data mentah hasil dari crawling data

4.2 Preprocessing

Pada bagian ini dilakukan beberapa tahapan preprocessing, antara lain: filtering/cleaning, stopword removal, tokenisasi, stemming.

4.2.1 Filtering/Cleaning

Berikut ini merupakan hasil *filtering/cleaning* yaitu menghapus elemen-elemen yang tidak relevan seperti yang tersaji pada gambar 3 berikut

```
def clean_text(text):
    # Hapus URL
    text = re.sub(r'http\S+|www.\S+', '', text)
    # Hapus angka
    text = re.sub(r'\d+', '', text)
    # Hapus emoji dan simbol
    text = re.sub(r'[^\w\s]', '', text)
    # Hapus karakter non-huruf
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    # Hapus spasi ganda
    text = re.sub(r'\s+', ' ', text).strip()
    return text
df['clean_comment'] = df['comment_lower'].apply(clean_text)
df[['comment_lower', 'clean_comment']].head(10)
```

	comment_lower	clean_comment
0	kami doakan semoga negeri indonesia menjadi le...	kami doakan semoga negeri indonesia menjadi le...
1	agak ragu aku dgn mentri ke uangan kita yg ini...	agak ragu aku dgn mentri ke uangan kita yg ini...
2	hayu luhut ganti bahlul ganti zulkipli ganti e...	hayu luhut ganti bahlul ganti zulkipli ganti e...
3	coba angkat kembali pak mahfud md, pak dedi mu...	coba angkat kembali pak mahfud md pak dedi mul...
4	gara-gara amarika syarikat(usa) kawal ekonomi ...	garagara amarika syarikatusa kawal ekonomi dun...
5	bu' sri sakit hati.. rumahnya kalian jarah...	bu sri sakit hati rumahnya kalian jarahjadi di...
6	sudah tepat bu sri diganti, karena sudah tidak...	sudah tepat bu sri diganti karena sudah tidak ...
7	harusnya ahok yg jdi menkeu.	harusnya ahok yg jdi menkeu
8	ada yang tau beliau dari kalangan apa ? par...	ada yang tau beliau dari kalangan apa brparpol...
9	the best menteri keuangan ri yg baru bpk purba...	the best menteri keuangan ri yg bpk purbaya yg...

Gambar 3. *Tampilan Data Teks setelah dilakukan Filtering/cleaning*

4.2.2 Stopwod removal

Berikut ini merupakan hasil penghapusan daftar *Stop words*, yang tidak memiliki kontribusi penting terhadap sentimen atau makna utama teks, seperti yang tersaji pada gambar 4 berikut.

```
import nltk
from nltk.corpus import stopwords

# Download stopwords (cukup sekali)
nltk.download('stopwords')

# Daftar stopword Bahasa Indonesia
stop_words = set(stopwords.words('indonesian'))

# Fungsi stopword removal
def stopword_removal(text):
    words = text.split()
    filtered_words = [word for word in words if word not in stop_words]
    return ' '.join(filtered_words)

# Terapkan stopword removal
df['stopword_removed'] = df['clean_comment'].apply(stopword_removal)

# Tampilkan hasil
df[['clean_comment', 'stopword_removed']].head(10)
```

	clean_comment	stopword_removed
0	kami doakan semoga negeri indonesia menjadi le...	doakan semoga negeri indonesia kedepannyabrdan...
1	agak ragu aku dgn mentri ke uangan kita yg ini...	ragu dgn mentri uangan yg inibsmengantur keuan...
2	hayu luhut ganti bahlul ganti zulkipli ganti e...	hayu luhut ganti bahlul ganti zulkipli ganti e...
3	coba angkat kembali pak mahfud md pak dedi mul...	coba angkat mahfud md dedi mulyadi anis
4	garagara amarika syarikatusa kawal ekonomi dun...	garagara amarika syarikatusa kawal ekonomi dun...
5	bu sri sakit hati rumahnya kalian jarahjadi di...	bu sri sakit hati rumahnya jarahjadi mundur
6	sudah tepat bu sri diganti karena sudah tidak ...	bu sri diganti berpihak rakyat menaikan pajak ...
7	harusnya ahok yg jdi menkeu	ahok yg jdi menkeu
8	ada yang tau beliau dari kalangan apa brparpol...	tau beliau kalangan brparpol kah
9	the best menteri keuangan ri yg baru bpk purba...	the best menteri keuangan ri yg bpk purbaya yg...

Gambar 4. *Tampilan Data Teks setelah dilakukan Stopword Removal*

4.2.3 Tokenisasi

Berikut ini merupakan hasil tokenisasi yaitu memecah teks menjadi unit-unit kata yang lebih kecil atau token, seperti yang tersaji pada gambar 5 berikut.

```
# Fungsi tokenisasi
def tokenize_text(text):
    return text.split()

# Terapkan tokenisasi
df['tokenized_comment'] = df['stopword_removed'].apply(tokenize_text)

# Tampilkan hasil
df[['stopword_removed', 'tokenized_comment']].head(10)
```

	stopword_removed	tokenized_comment
0	doakan semoga negeri indonesia kedepannyabrdan...	[doakan, semoga, negeri, indonesia, kedepannya...
1	ragu dgn mentri uangan yg inibsmengantur keuan...	[ragu, dgn, mentri, uangan, yg, inibsmengantur...
2	hayu luhut ganti bahlul ganti zulkipli ganti e...	[hayu, luhut, ganti, bahlul, ganti, zulkipli, ...
3	coba angkat mahfud md dedi mulyadi anis	[coba, angkat, mahfud, md, dedi, mulyadi, anis]
4	garagara amarika syarikatusa kawal ekonomi dun...	[garagara, amarika, syarikatusa, kawal, ekonom...
5	bu sri sakit hati rumahnya jarahjadi mundur	[bu, sri, sakit, hati, rumahnya, jarahjadi, mu...
6	bu sri diganti berpihak rakyat menaikan pajak ...	[bu, sri, diganti, berpihak, rakyat, menaikan,...
7	ahok yg jdi menkeu	[ahok, yg, jdi, menkeu]
8	tau beliau kalangan brparpol kah	[tau, beliau, kalangan, brparpol, kah]
9	the best menteri keuangan ri yg bpk purbaya yg...	[the, best, menteri, keuangan, ri, yg, bpk, pu...

Gambar 5. *Tampilan Data Teks setelah dilakukan Tokenisasi*

4.2.4 Stemming

Berikut ini merupakan tampilan hasil stemming yaitu menghilangkan imbuhan atau mengubah kata menjadi bentuk

dasarnya, seperti yang tersaji pada gambar 6 berikut.

```
!pip install Sastrawi

# Import stemmer
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

# Inisialisasi stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# Fungsi stemming (input: list token, output: list token hasil stemming)
def stemming_text(token_list):
    return [stemmer.stem(word) for word in token_list]

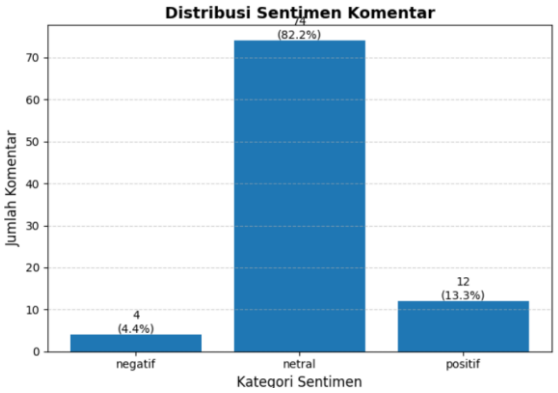
# Terapkan stemming
df['stemmed_comment'] = df['tokenized_comment'].apply(stemming_text)

# Tampilkan hasil perbandingan
df[['tokenized_comment', 'stemmed_comment']].head(10)
```

	tokenized_comment	stemmed_comment
0	[doakan, semoga, negeri, indonesia, kedepannya...	[doa, moga, negeri, indonesia, kedepannyabrdan...
1	[ragu, dgn, mentri, uangan, yg, inibsmengantur...	[ragu, dgn, tri, uang, yg, inibsmengantur, uan...
2	[hayu, luhut, ganti, bahlul, ganti, zulkipli, ...	[hayu, luhut, ganti, bahlul, ganti, zulkipli, ...
3	[coba, angkat, mahfud, md, dedi, mulyadi, anis]	[coba, angkat, mahfud, md, dedi, mulyadi, anis]
4	[garagara, amarika, syarikatusa, kawal, ekonom...	[garagara, amarika, syarikatusa, kawal, ekonom...
5	[bu, sri, sakit, hati, rumahnya, jarahjadi, mu...	[bu, sri, sakit, hati, rumah, jarahjadi, mundur]
6	[bu, sri, diganti, berpihak, rakyat, menaikan,...	[bu, sri, ganti, pihak, rakyat, naik, pajak, t...
7	[ahok, yg, jdi, menkeu]	[ahok, yg, jdi, menkeu]
8	[tau, beliau, kalangan, brparpol, kah]	[tau, beliau, kalang, brparpol, kah]
9	[the, best, menteri, keuangan, ri, yg, bpk, pu...	[the, best, menteri, uang, ri, yg, bpk, purbay...

Gambar 6. *Tampilan Data Teks setelah dilakukan Stemming*

4.3 Hasil Pelabelan Sentimen Berbasis Kamus



Gambar 7. Hasil Pelabelan Sentimen Berbasis Kamus

Pelabelan dilakukan menggunakan kamus sentimen bahasa Indonesia. Hasil distribusi sentimen dari 90 data adalah sebagai berikut:

- Sentimen Positif: 12 komentar (13.3%)
- Sentimen Negatif: 4 komentar (4.4%)
- Sentimen Netral: 74 komentar (82.2%)

Gambar diatas menunjukkan diagram batang Hasil Pelabelan Sentimen Berbasis Kamus. Tiga kategori sentimen yang dianalisis adalah positif, negatif, dan netral. Dari hasil yang ditampilkan, terlihat bahwa mayoritas sentimen publik cenderung netral, dengan 82.2% mengungkapkan sentimen netral. Sementara itu, sentimen positif memperoleh 13.3%, dan hanya 4.4% yang menunjukkan sentimen negatif terhadap Menteri Purbaya.

4.4 Hasil Klasifikasi dan Evaluasi Model

4.4.1Decision Tree

```
Accuracy: 1.0

Classification Report:
              precision    recall  f1-score   support

     Negatif       1.00      1.00      1.00         2
      Netral       1.00      1.00      1.00        15
     Positif       1.00      1.00      1.00         1

    accuracy                           1.00        18
   macro avg       1.00      1.00      1.00        18
weighted avg       1.00      1.00      1.00        18
```

Gambar 8. Model Decision Tree

Model Decision Tree adalah model terbaik dalam hal akurasi pada dataset ini dibandingkan SVM, K-NN, dan Naive Bayes. Hasil evaluasi menunjukkan bahwa model Decision Tree mampu mengklasifikasikan seluruh sampel uji dengan tepat, sehingga memperoleh skor akurasi, presisi, recall, dan F1-score sebesar 100%. Semua metrik untuk kelas Negatif, Netral, dan Positif bernilai sempurna.

Berdasarkan hasil pengujian awal, model Deep Learning menunjukkan performa yang tidak optimal. Hal ini disebabkan oleh keterbatasan jumlah data uji yang hanya berjumlah 18 sampel, sehingga model tidak memiliki referensi yang cukup untuk membentuk pola kompleks atau membedakan sentimen yang sangat spesifik. Sebaliknya, algoritma Decision Tree terbukti lebih stabil dan efektif dalam menangani dataset berukuran terbatas (90 data total). Oleh karena itu, penelitian ini memfokuskan analisis mendalam pada algoritma Decision Tree yang memberikan hasil lebih representatif terhadap karakteristik data yang digunakan

5. Simpulan

Pelabelan otomatis berbasis kamus sentimen InSet berhasil mengklasifikasikan sentimen komentar YouTube terkait pelantikan Menteri Keuangan Purbaya dengan distribusi

sentimen netral (82,2%), positif (13,3%), dan negatif (4,4%). Algoritma Decision Tree mencapai akurasi sempurna (100%) pada data uji, mengungguli SVM, KNN, Naïve Bayes, Neural Network, dan Deep Learning. Namun, penelitian ini memiliki keterbatasan berupa jumlah data terbatas (90 komentar), ketidakmampuan kamus InSet menangani ironi dan sarkasme, potensi overfitting pada Decision Tree akibat dataset kecil dan ketidakseimbangan kelas, serta fokus hanya pada satu platform dan satu momen waktu.

Penelitian lanjutan disarankan untuk mengumpulkan dataset yang lebih besar dan seimbang, menggabungkan lexicon-based dengan metode context-aware embedding, menguji model Transformer seperti BERT Indonesia, memperluas cakupan ke platform lain (Twitter, Instagram, TikTok), dan menambahkan analisis temporal untuk melihat perubahan sentimen dari waktu ke waktu. Penelitian ini telah mengisi celah metodologis dengan menggabungkan pelabelan berbasis kamus dan Decision Tree yang terbukti efektif untuk analisis real-time dan dapat diadopsi sebagai alat pendukung keputusan berbasis opini publik digital.

Ucapan Terima Kasih

Saya mengucapkan terima kasih kepada pembimbing, dosen, serta rekan-rekan yang memberikan arahan, bantuan, dan masukan berharga dalam proses penyusunan jurnal ini. Dukungan mereka sangat berarti dalam tercapainya penelitian ini.

Daftar Pustaka

Artikel jurnal:

Yasir, M., & Suraji, R. (2023). Perbandingan Metode Klasifikasi Naïve Bayes, Decision Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube. *Jurnal Cahaya Mandalika (JCM)*, *3*(2), 180–192.

Aribowo, A. S., Basiron, H., Yusof, N. F. A., & Khomsah, S. (2021). Cross-domain sentiment analysis model on indonesian youtube comment. *International Journal of Advances in Intelligent Informatics*, *7*(1), 12–25. https://doi.org/10.26555/ijain.v7i1.554

Armaeni, P. P., Wiguna, I. K. A. G., & Parwita, W. G. S. (2024). Sentiment Analysis of YouTube Comments on the Closure of TikTok Shop Using Naïve Bayes and Decision Tree Method Comparison. *Jurnal Galaksi*, *1*(2), 70–80. https://doi.org/10.70103/galaksi.v1i2.15

Prasetyo, T., Adhyaksa Waskita, A., Taryo, T., Raya Puspitek, J., & Pamulang Kota Tangerang Selatan, K. (2025). Analisis Sentimen Pengguna Seputar Kendaraan Listrik Di Twitter Dengan Penerapan Algoritma Naïve Bayes, KNN, dan Decision Tree untuk Klasifikasi. *Jurnal Sistem Komputer Dan Kecerdasan Buatan*, *8*(2), 108–117. www.dataindonesia.com,

Safra, K. K., & Zuliarso, E. (2025). Analisis Sentimen Terhadap Pelaksanaan Pilkada 2024 Pada Media Sosial Youtube Menggunakan Metode Decision Tree. *Jurnal Informatika Teknologi Dan Sains*, *7*(1), 117–126.

Liu, B. (2020). Sentiment analysis: Mining opinions, sentiments, and emotions (2nd ed.). Cambridge University Press.

Wibowo, A. (2024). *Analisis Sentimen Popularitas Capres dan Pilpres pada Media Sosial Twitter: Perbandingan Algoritma SVM, KNN, dan Naïve Bayes*. *23*(2), 362–374.

Saputra, J., Maryani, L., Wulandari, D., & Eka, W. (2025). Analisis Performa Naive Bayes dan SVM terhadap Sentimen Teks Media Sosial dengan Word2Vec dan SMOTE. *JURNAL INSTEK INFORMATIKA SAINS DAN TEKNOLOGI*, *10*(1), 143–155. https://journal.uin-alauddin.ac.id/index.php/instek

Gede Bagus Resa Pratama Putra, I., Abd Latif, K., Widyawati, L., & Innudin, M. (2025). *Analisis Sentimen Berbasis Lexicon Dan Social Network Analysis Terhadap Topik Ruu Tni Pada Media Sosial X*. 346–452.

Halimi, A., & Rudyanto Arief, M. (2021). Analisis Sentimen Masyarakat Indonesia Terhadap Pembelajaran Online Dari Di Media Sosial Twitter Menggunakan Lexicon Dan K-Nearest Neighbor. *Jurnal Kecerdasan Buatan, Komputasi Dan Teknologi Informasi*, *2*(1), 18–28. https://ejournal.unuja.ac.id/index.php/core

Mario Conroy Paridy Man, G., Aristo Jansen Sinlae, A., & Ngaga, E. (2025). JIP (Jurnal Informatika Polinema) Analisis Sentimen di Media Sosial X tentang IKN dengan Naïve Bayes. *JIP (Jurnal Informatika Polinema)*, *11*, 417–425.

Pramita Widyassari, A., Salsabilla, D., & Ali Amrozi, M. (2025). Analisis sentimen publik di twitter terhadap pelantikan presiden Prabowo menggunakan algoritma Naïve Bayes. Jurnal Ilmiah NERO, 10(1), 2025.