

KOMPARASI ALGORITMA C4.5 DAN NAIVE BAYES UNTUK ANALISIS SENTIMEN KOMENTAR VIDEO YOUTUBE MENTERI KEUANGAN

Egi Gusnandi Pratama¹, Ahmad Homaidi², Hermanto³,
Universitas Ibrahimy
egigusnandipratama@gmail.com

ABSTRAK

Fenomena masifnya penggunaan media sosial, khususnya YouTube, menjadikannya sebagai ruang diskusi digital paling aktif yang merepresentasikan opini masyarakat secara langsung. Setelah pelantikan Purbaya Yudhi Sadewa sebagai Menteri Keuangan menggantikan Sri Mulyani Indrawati, muncul beragam reaksi, kritik, maupun dukungan publik di kolom komentar YouTube terkait program kerja fiskalnya. Diperlukan penanganan sistematis untuk memetakan persepsi publik ini dari ribuan komentar teks bervolume besar secara objektif. Penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen guna membandingkan performa akurasi klasifikasi antara algoritma C4.5 dan Naïve Bayes. Metode penelitian kuantitatif ini mengacu pada model Knowledge Discovery in Database (KDD) dengan subjek data berupa teks komentar berbahasa Indonesia dari YouTube Comment API pada saluran Liputan6. Pengujian komputasi dari total dataset asli sebanyak 9.256 data mentah berhasil menyaring sampel data optimal sebanyak 832 data (396 negatif, 291 netral, dan 145 positif). Evaluasi model menggunakan algoritma C4.5 berhasil mencapai nilai performa yang lebih unggul dibandingkan algoritma Naïve Bayes, dengan nilai akurasi sebesar 82,54%, precision 82,57%, dan recall 82,54%. Sementara itu, algoritma Naïve Bayes menghasilkan nilai akurasi yang lebih rendah yaitu 78,45%, precision 78,94%, dan recall 78,45%. Sistem analisis sentimen dengan algoritma C4.5 terbukti bekerja secara efektif dan memiliki performa yang lebih unggul daripada Naïve Bayes dalam menangani pengolahan teks opini publik multi-kelas pada media sosial YouTube.

Kata kunci : Analisis Sentimen, Menteri Keuangan, Algoritma C4.5, Naive Bayes, Youtube

ABSTRACT

The massive phenomenon of social media usage, particularly YouTube, has turned it into the most active digital discussion space that directly represents public opinion. Following the inauguration of Purbaya Yudhi Sadewa as the Minister of Finance, replacing Sri Mulyani Indrawati, various public reactions, criticisms, and support emerged in the YouTube comment section regarding his fiscal work program. A systematic approach is required to objectively map this public perception from thousands of large-volume text comments. This study aims to develop a sentiment analysis system to compare the classification accuracy performance between the C4.5 and Naïve Bayes algorithms. This quantitative research method refers to the Knowledge Discovery in Database (KDD) model, using Indonesian comment text data crawled via the YouTube Comment API from the Liputan6 channel. Computational testing on the original dataset of 9,256 raw data successfully filtered an optimal sample of 832 data (396 negative, 291 neutral, and 145 positive). Model evaluation using the C4.5 algorithm achieved superior performance compared to the Naïve Bayes algorithm, with an accuracy of 82,54%, precision of 82,57%, and recall of 82,54%. Meanwhile, the Naïve Bayes algorithm produced lower values: accuracy of 78,45%, precision of 78,94%, and recall of 78,45%. The sentiment analysis system using the C4.5 algorithm is proven to work effectively and has superior performance compared to Naïve Bayes in handling multi-class public opinion text processing on YouTube social media.

Keywords: Sentiment Analysis, Minister Of Finance, C4.5 Algorithm, Naïve Bayes, Youtube

Pendahuluan

Kementerian negara merupakan perangkat pemerintah yang membidangi urusan tertentu dalam pemerintahan, sedangkan Menteri Keuangan merupakan pembantu Presiden yang memimpin Kementerian dalam birokrasi pemerintahan [1]. Kementerian Keuangan memiliki peran besar dalam memobilisasi sumber pembiayaan domestik maupun internasional, sehingga figurnya selalu menjadi pusat perhatian masyarakat luas [2], [3]. Sosok Menteri Keuangan Purbaya Yudhi Sadewa yang resmi dilantik sejak 8 September 2025 menggantikan Sri Mulyani Indrawati memicu intensitas perbincangan yang masif di ruang digital [4]. Seiring dengan pergeseran pola konsumsi informasi masyarakat ke media sosial [5], YouTube menyediakan fitur kolom komentar sebagai wadah penyalur respons nyata [6]. Namun, akumulasi data komentar teks yang berjumlah ratusan hingga ribuan memunculkan tantangan baru dalam melakukan ekstraksi makna secara manual [7]. Teknik text mining melalui analisis sentimen hadir sebagai solusi untuk mendeteksi karakteristik polaritas teks secara otomatis [8]. Algoritma C4.5 yang mengukur tingkat ketidakpastian dataset berdasarkan nilai entropi dan information gain dipilih karena menghasilkan pohon keputusan yang sangat interpretatif [9], [10]. Kinerja C4.5 ini perlu dibandingkan dengan Naïve Bayes, sebuah algoritma probabilistik yang sangat efisien dalam pemrosesan data teks media sosial [11], [12], [13]. Melalui komparasi performa kedua metode ini, diharapkan sistem mampu menyajikan pemetaan opini publik secara lebih akurat dan objektif [14].

Penelitian ini dilaksanakan dengan tujuan untuk menganalisis dan membandingkan kinerja algoritma C4.5 dan Naïve Bayes dalam mengklasifikasikan sentimen komentar YouTube terhadap Menteri Keuangan. Terkait dengan studi literatur tersebut, dilakukan eksperimen komputasi terstruktur guna memetakan persepsi publik dari data bervolume besar. Cara pengumpulan data dan rancangan tahapan eksperimen akan dijelaskan secara rinci dalam bagian Materi dan Metode. Selanjutnya, hasil evaluasi metrik performa

serta visualisasi model klasifikasi dari kedua algoritma akan dipaparkan dan dibahas pada bagian Hasil dan Pembahasan sebelum ditarik kesimpulan akhir dari penelitian ini.

Materi dan Metode

Penelitian ini termasuk ke dalam jenis penelitian kuantitatif dengan pendekatan deskriptif eksperimental melalui pemrosesan data mining pada bahasa pemrograman Python [9], [15]. Populasi data dalam penelitian ini meliputi seluruh komentar publik pada saluran YouTube Liputan6 yang berjudul "[Full] Pernyataan Purbaya Yudhi Sadewa Setelah Dilantik Menjadi Menteri Keuangan | Liputan 6". Data mentah diekstraksi secara otomatis menggunakan teknik crawling memanfaatkan pustaka YouTube Comment API hingga berhasil mengumpulkan dataset asli sebanyak 9.256 teks komentar. Dari populasi besar tersebut, dilakukan teknik purposive sampling melalui tahapan seleksi data (data cleaning) untuk mengeliminasi spam, duplikasi, dan komentar di luar topik ekonomi, sehingga diperoleh sampel representatif bersih sebanyak 832 baris yang divalidasi oleh sistem.

Proses pengolahan data mengacu pada kerangka Knowledge Discovery in Database (KDD). Tahapan pertama adalah Seleksi Data untuk menentukan kumpulan 832 baris komentar ke dalam satu dataset utama. Tahap kedua adalah Praproses Teks (Text Preprocessing) melalui proses cleaning, tokenizing, stopword removal, dan stemming menggunakan Python [16]. Tahap ketiga adalah Transformasi Data untuk mengubah teks bersih menjadi bentuk fitur numerik menggunakan metode pembobotan Term Frequency-Inverse Document Frequency (TF-IDF). Tahap keempat adalah Penerapan Algoritma Klasifikasi di mana dataset diuji menggunakan skrip komputasi algoritma C4.5 [10] dan Naïve Bayes Classifier [17]. Tahap terakhir adalah Evaluasi Model dengan menguji kualitas hasil klasifikasi model sentimen menggunakan parameter Confusion Matrix untuk mengukur akurasi, precision, dan recall [18].

Hasil dan Pembahasan

Eksperimen klasifikasi data teks opini komentar YouTube dijalankan melalui pemrograman Python menggunakan editor Cursor AI dan diintegrasikan ke dalam sistem aplikasi berbasis website dengan kerangka kerja Streamlit. Dari total populasi mentah sebanyak 9.256 baris data, tahapan seleksi dan preprocessing teks berhasil menyaring sampel data optimal yang merepresentasikan polaritas sentimen publik secara riil. Karakteristik distribusi sebaran label data sentimen publik hasil tahapan penyaringan sampel dipaparkan pada Tabel 1..

Tabel 1. Distribusi Klasifikasi Sampel Data Hasil Preprocessing Komentar Video YouTube

Kategori Sentimen	N	%
Negatif	396	47,6
Netral	291	35,0
Positif	145	17,4
Total Sampel Valid	832	100,0
Dataset Asli (<i>Raw Data</i>)	9,256	-

Berdasarkan data pada Tabel 1, opini masyarakat didominasi oleh sentimen negatif yang mencapai 396 komentar (47,6%), disusul respons netral sebanyak 291 komentar (35,0%), dan sentimen positif sebanyak 145 komentar (17,4%). Pemodelan dari sebaran data tersebut menghasilkan perbandingan performa metrik evaluasi yang disajikan secara teratur pada Tabel 2.

Tabel 2. Perbandingan Metrik Evaluasi Pemodelan

Metrik Evaluasi	Algoritma C4.5	Algoritma Naïve Bayes
Accuracy	82,54	78,45
Precision	82,57	78,94
Recall	82,54	78,45
F1-Score	82,55	77,65

Berdasarkan hasil pengujian pada Tabel 2, performa klasifikasi algoritma C4.5 menunjukkan keunggulan signifikan dengan

nilai akurasi sebesar 82,54% dan F1-score sebesar 82,55%. Tingginya tingkat keandalan pemodelan ini didukung secara kuat oleh teori pembentukan pohon keputusan (decision tree) melalui perhitungan matematis nilai indeks entropy dan rasio keuntungan (gain ratio) [9]. Kekuatan utama C4.5 terletak pada kemampuannya memilih atribut pembagi terbaik yang memiliki tingkat ketidakpastian paling rendah di dalam dataset [10]. Fakta empiris luaran sistem membuktikan bahwa fitur kata kunci seperti "baik", "pajak", "mudah", dan "sesuai" terpilih sebagai simpul akar (root node) dengan bobot gain ratio tertinggi. Kehadiran kata "pajak" sebagai batas keputusan krusial di puncak hierarki pohon merepresentasikan korelasi teoretis yang kuat dengan realitas objek penelitian, di mana opini publik secara spontan mengaitkan pelantikan Menteri Keuangan Purbaya Yudhi Sadewa dengan sektor kebijakan fiskal negara. Aturan keputusan kondisional berbentuk if-then yang dihasilkan C4.5 terbukti efektif mengisolasi variasi ekspresi teks informal [8]. Hal ini membuat model mampu memetakan polaritas multi-kelas secara presisi (precision 82,57%) dan menekan angka misclassification pada kelas minoritas positif di tengah dominasi sentimen negatif.

Di sisi lain, penerapan algoritma Naïve Bayes Classifier menghasilkan metrik evaluasi yang kurang optimal, dengan capaian nilai akurasi yang lebih rendah yaitu sebesar 78,45% dan F1-score di angka 77,65%. Keterbatasan performa ini berakar pada fondasi teoretis teorema Bayes yang mengandalkan asumsi independensi bersyarat yang sangat ketat (naive independence assumption), di mana algoritma memperlakukan setiap token kata sebagai variabel mandiri yang terisolasi tanpa memedulikan konteks struktural kalimat [14]. Karakteristik teks komentar YouTube Liputan6 yang kaya akan bahasa gaul, singkatan, serta struktur kalimat kompleks bermuara pada kelemahan fatal bagi algoritma ini. Teori probabilistik Naïve Bayes mendasarkan klasifikasinya pada akumulasi frekuensi perkalian probabilitas kata tunggal

[17]. Konsekuensinya, sistem menaruh preferensi bobot logistik tertinggi pada kata-kata umum yang tidak memiliki muatan polaritas emosi spesifik, seperti "pak", "yg", dan "rakyat". Ketidakmampuan menangkap ketergantungan semantik ini menyebabkan algoritma gagal mengidentifikasi batas esensi kalimat netral secara utuh, sehingga model mengalami bias klasifikasi (class bias) yang masif dengan melemparkan 63,3% data ke dalam kluster sentimen netral. Eksperimen ini menegaskan secara teoretis bahwa untuk pemrosesan teks opini publik multi-kelas yang tidak seimbang (imbalanced dataset) dan bersifat non-formal, algoritma C4.5 jauh lebih adaptif dibandingkan pendekatan berbasis probabilitas independen Naïve Bayes.

Kesimpulan

Penelitian komparasi ini berhasil menyimpulkan bahwa opini publik terhadap pelantikan Menteri Keuangan Purbaya Yudhi Sadewa didominasi oleh sentimen negatif sebesar 47,6%, diikuti sentimen netral sebesar 35,0%, dan sentimen positif sebesar 17,4% dari total 832 sampel bersih tervalidasi. Dalam pengujian performa klasifikasi multi-kelas, algoritma C4.5 (Decision Tree) terbukti bekerja secara lebih efektif dan memiliki ketahanan pemodelan yang jauh lebih unggul dibandingkan dengan algoritma Naïve Bayes Classifier. Keunggulan tersebut ditunjukkan oleh capaian metrik evaluasi C4.5 yang mendominasi di seluruh parameter, meliputi nilai akurasi sebesar 82,54%, presisi 82,57%, recall 82,54%, dan F1-score sebesar 82,55%, sedangkan algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 78,45%, presisi 78,94%, recall 78,45%, dan F1-score sebesar 77,65%. Melalui visualisasi pohon keputusan, algoritma C4.5 juga berhasil menyajikan interpretasi batas aturan keputusan yang kuat melalui fitur kata kunci utama seperti "baik", "pajak", "mudah", dan "sesuai" sebagai indikator utama orientasi opini publik.

Ucapan terima kasih

Penulis menyampaikan terima kasih kepada seluruh pihak di Universitas Ibrahimy yang

telah memberikan dukungan moral dan teknis selama proses komputasi data penelitian ini. Penulis juga mengapresiasi panitia penyelenggara Seminar Nasional atas kesempatan dan ruang diseminasi yang diberikan, sehingga hasil penelitian komparasi algoritma ini dapat dipublikasikan dalam prosiding ilmiah.

Daftar Pustaka

- [1] O. Christin, N. Liu, H. Pondaag, and K. Y. Umboh, "KEDUDUKAN KEMENTERIAN NEGARA DALAM SISTEM PEMERINTAHAN NEGARA REPUBLIK INDONESIA," 2022.
- [2] M. B. Basri, "OPTIMASI PENGELOLAAN BARANG MILIK NEGARA SEBAGAI SUMBER PENERIMAAN NEGARA YANG HANDAL," Kementrian Keuangan. Accessed: Nov. 17, 2025. [Online]. Available: <https://www.djkn.kemenkeu.go.id/kpkn-l-kendari/baca-artikel/15602/OPTIMASI-PENGELOLAAN-BARANG-MILIK-NEGARA-SEBAGAI-SUMBER-PENERIMAAN-NEGARA-YANG-HANDAL.html>
- [3] K. Pembiayaan, *Suatu pengantar*. Bogor: IPN Press, 2019.
- [4] A. Khaidar, "ANALISIS SENTIMEN DI INSTAGRAM TERHADAP MENTERI KEUANGAN PURBAYA YUDHI SADEWA MENGGUNAKAN METODE LOGISTIC REGRESSION," vol. 13, no. 3, 2025.
- [5] N. P. Panjaitan, S. Z. Harahap, and R. M. Ah, "Analisis Minat Masyarakat Menggunakan Media Sosial Menggunakan Algoritma C4.5 dan Metode Naïve Bayes," vol. 12, no. 3, pp. 546–555, 2024.
- [6] K. S. Putri, I. R. Setiawan, A. Pambudi, A. Sentimen, and N. B. Classifier, "'Technologia' Vol 14, No. 3, Juli 2023 227 ANALISIS SENTIMEN TERHADAP BRAND SKINCARE LOKAL MENGGUNAKAN NAÏVE

- BAYES CLASSIFIER,” vol. 14, no. 3, pp. 227–232, 2023.
- [7] D. Rahmadden, “Analisis Sentimen Pengguna Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45,” vol. 13, no. 5, pp. 8153–8164, 2024.
- [8] A. D. P. M. Betty, “Analisis Sentimen pada Ulasan Penyedia Layanan Sentiment Analysis on Service Provider Reviews Using the C4 . 5 Algorithm,” vol. 6, no. 2, pp. 1–5, 2023.
- [9] D. Hari Suparwito, *PENGANTAR PEMBELAJARAN MESIN MENGGUNAKAN BAHASA PEMOGRAMAN PYTHON*. Yogyakarta: SANATA DHARMA UNIVERSITY PRESS, 2023.
- [10] I. Vusuvangat and S. K. Gusti, “Penerapan Algoritma C4 . 5 untuk Klasifikasi Sentimen Masyarakat terhadap # RUUKUHP pada Twitter,” vol. 6, no. 4, 2023, doi: 10.32493/jtsi.v6i3.33814.
- [11] M. Adriansa, L. Yulianti, L. Elfianty, and U. D. Bengkulu, “Analisis Kepuasan Pelanggan Menggunakan Algoritma,” vol. 07, no. 21, pp. 115–121, 2022.
- [12] I. I. Hajar, M. Khandava M., and U. Enri, “Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita,” vol. 8, no. July, pp. 116–125, 2022.
- [13] R. Hidayat, A. B. Hakim, and R. Nugraha, “Perbandingan Metode Naïve Bayes Dan Decision Tree C4 . 5 untuk Analisis Sentimen Produk Es Teh Indonesia di Media Sosial Twitter,” 2024.
- [14] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [15] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” vol. 15, no. 1, pp. 131–145, 2021.
- [16] S. Junaidi, M. Devegi, and H. Kurniawan, “Pelatihan Pengolahan dan Visualisasi Data Penduduk Menggunakan Python,” vol. 4, no. 1, pp. 151–162, 2023, doi: 10.30812/adma.v4i1.2963.
- [17] A. Safira, F. N. Hasan, K. J. Timur, and N. B. Classifier, “PAYLATER MENGGUNAKAN METODE NAIVE BAYES,” vol. 5, no. 1, pp. 59–70, 2023.
- [18] P. A. Permatasari and L. Jasa, “Survei Tentang Analisis Sentimen Pada Media Sosial,” vol. 20, no. 2, 2021.