ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP POLEMIK PENGADAAN CHROMEBOOK MENGGUNAKAN ALGORITMA NAIVE BAYES

Johannes Wikri Lumban Gaol[1], Imam Budiawan[2], Yumi Novita Dewi[3]

[1,2,3]*Universitas Bina Sarana Informatika*
E-mail : wikrigaol@gmail.com[1], imam.budiawan@bsi.ac.id[2], yumi.novita@bsi.ac.id[3]

ABSTRAK

Program pengadaan Chromebook oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) telah memicu beragam opini publik di media sosial, khususnya YouTube. Studi ini bertujuan untuk menganalisis tren sentimen publik dan mengevaluasi kinerja algoritma Naive Bayes Classifier dalam mengklasifikasikan opini tersebut. Sebanyak 789 baris ulasan dikumpulkan menggunakan YouTube Data API v3. Tahapan penelitian meliputi pra-pemrosesan data (case folding, tokenisasi, penyaringan, dan stemming), pelabelan otomatis menggunakan kamus leksikon, dan pembagian dataset dengan rasio 80:20 (631 data pelatihan dan 158 data pengujian). Hasil pelabelan menunjukkan bahwa opini publik didominasi oleh sentimen Netral sebesar 60,2%, diikuti oleh sentimen Positif sebesar 31,7%, dan sentimen Negatif sebesar 8,1%. Berdasarkan hasil pengujian sistem, model Naive Bayes berhasil mencapai Akurasi Global sebesar 72,15%. Model tersebut terbukti sangat kuat dalam mengenali sentimen Positif, mencapai tingkat presisi 95%, dan sangat sensitif dalam mendeteksi kelas Netral dengan tingkat recall 99%. Namun, metrik evaluasi untuk kelas Negatif tercatat 0,00 karena ketidakseimbangan data dalam sampel pengujian minoritas. Secara keseluruhan, studi ini menyimpulkan bahwa mayoritas publik memiliki sikap objektif, dan model yang dikembangkan dianggap representatif untuk memetakan opini publik.

Kata Kunci : *Analisis Sentimen, Kemendikbudristek Chromebook, YouTube, Pengklasifikasi Naive Bayes, Leksikon.*

ABSTRACT

The Chromebook procurement program by the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) has sparked diverse public opinions on social media, particularly YouTube. This study aims to analyze public sentiment trends and evaluate the performance of the Naive Bayes Classifier algorithm in classifying these opinions. A total of 789 review rows were collected using the YouTube Data API v3. The research stages involved data preprocessing (case folding, tokenizing, filtering, and stemming), automated labeling using a lexicon dictionary, and dataset splitting with an 80:20 ratio (631 training data and 158 testing data). The labeling results revealed that public opinion was dominated by Neutral sentiment at 60.2%, followed by Positive sentiment at 31.7%, and Negative sentiment at 8.1%. Based on the system testing results, the Naive Bayes model successfully achieved a Global Accuracy of 72.15%. The model proved to be highly robust in recognizing Positive sentiment, reaching a precision rate of 95%, and highly sensitive in detecting the Neutral class with a recall rate of 99%. However, the evaluation metric for the Negative class was recorded at 0.00 due to data imbalance within the minority testing samples. Overall, this study concludes that the majority of the public holds an objective stance, and the developed model is considered representative for mapping public opinion.

Keywords : *Sentiment Analysis, Kemendikbudristek Chromebook, YouTube, Naive Bayes Classifier, Lexicon.*

PENDAHULUAN

Latar Belakang

Perkembangan teknologi informasi telah memberikan pengaruh besar terhadap sektor pendidikan nasional di Indonesia. Era digital mengubah lanskap sosialisasi kebijakan publik dari media konvensional ke platform berbasis audio-visual seperti YouTube. YouTube kini menjadi ruang terbuka bagi masyarakat untuk berinteraksi, memberikan ulasan, hingga menuangkan kritik secara langsung melalui kolom komentar.

Sebagai inovasi layanan publik, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) meluncurkan kebijakan digitalisasi sekolah melalui pengadaan perangkat laptop dalam negeri bermerek Chromebook. Kebijakan ini ditujukan untuk mendukung administrasi sekolah dan pelaksanaan Asesmen Nasional Berbasis Komputer (ANBK). Namun, program ini memicu polemik hangat di tengah masyarakat. Berbagai narasi kritis muncul terkait efisiensi anggaran belanja negara yang dinilai terlalu tinggi dan keterbatasan fungsionalitas sistem operasi ChromeOS.

Besarnya volume opini publik yang acak dan tidak terstruktur membuat evaluasi manual menjadi tidak efisien. Oleh karena itu, diperlukan pendekatan komputasional otomatis berupa analisis sentimen (*sentiment analysis*) berbasis *text mining* untuk mengelompokkan opini ke dalam polaritas positif, negatif, dan netral. Algoritma Naïve Bayes Classifier dipilih karena keunggulannya yang memiliki komputasi cepat, pemodelan sederhana, serta efisiensi tinggi dalam menangani dokumen teks berdimensi besar.

Rumusan Masalah

1. Bagaimana kecenderungan sebaran sentimen pengguna YouTube di Indonesia merespon polemik kebijakan pengadaan Chromebook?
2. Bagaimana tahapan implementasi teknik *text preprocessing* dan ekstraksi fitur menggunakan metode TF-IDF untuk mempersiapkan data komentar YouTube?
3. Bagaimana performa algoritma Naïve Bayes Classifier yang diukur melalui metrik *Confusion Matrix*?

METODE

Secara keseluruhan, alur rekayasa data teks dalam penelitian ini terdiri dari beberapa tahapan terstruktur:

a. Pengumpulan Data: Menggunakan teknik *web scraping* via YouTube Data API v3, berhasil mengumpulkan 789 komentar dari video populer terkait topik.
b. Pelabelan Sentimen: Menggunakan kamus *lexicon* untuk mengelompokkan teks ke kelas Positif (dukungan), Netral (pernyataan objektif/pertanyaan), dan Negatif (kritik anggaran/spesifikasi).
c. Mekanisme Preprocessing:
   Case Folding: Menjadikan huruf kecil dan menghapus simbol non-alfabet.
   Tokenizing: Memecah kalimat menjadi token kata.
   Slang Word Normalization: Mengubah kata tidak baku menjadi baku (misal: "spek" → "spesifikasi").
   Stopword Removal: Menghapus kata umum tak bermakna.

Stemming: Menggunakan algoritma Nazief-Adriani melalui pustaka Sastrawi.

d. Pembagian Dataset: Memisahkan data dengan rasio 80% data latih (631 baris) dan 20% data uji (158 baris).

HASIL PENELITIAN DAN PEMBAHASAN

Distribusi Sentimen

Proses *crawling* menghasilkan 789 data valid. Berdasarkan visualisasi data, sebaran kategori opini publik menunjukkan dominasi sentimen objektif, dengan rincian:

Netral: 475 data (60,2%)
Positif: 250 data (31,7%)
Negatif: 64 data (8,1%)

Melalui analisis *Word Cloud*, kata-kata dominan pada sentimen positif melingkupi kata seperti *"jujur"*, *"bagus"*, *"indonesia"*, dan *"semoga"*.



Gambar 1. ***Word Cloud Positif***

Sedangkan pada kelas netral dipenuhi kata *"saja"*, *"ini"*, *"ada"*, dan *"pak"*.



Gambar 2. ***Word Cloud Netral***

Kelas negatif memuat istilah kritis seperti *"korupsi"*, *"mahal"*, *"hancur"*, dan *"kecewa"*.



Gambar 3. ***Word Cloud Negatif***

Hasil Pengujian dan Evaluasi Model Klasifikasi

Pengujian dilakukan dengan membagi dataset menggunakan metode *Hold-Out Validation* dengan rasio 80% untuk data latih (*training data*) sebesar 631 baris, dan 20% untuk data uji (*testing data*) sebesar 158 baris. Model dievaluasi menggunakan metrik *confusion matrix* pelaporan multi-kelas (3x3) untuk menghitung nilai *Precision*, *Recall*, dan *F1-Score*. Hasil kalkulasi akhir disajikan secara terperinci pada tabel di bawah ini:

Tabel 1. Matriks Evaluasi Performa Algoritma Naive Bayes

Kelas Klasifikasi	Presisi (Precision)	Sensitivitas (Recall)	F1-Score	Jumlah Sampel Uji (Support)
Sentimen Negatif	00.00	00.00	00.00	13
Sentimen Netral	0,047916667	0,06875	0,05625	95
Sentimen Positif	0,065972222	00.40	00.56	50
Rata-rata / Total	0,05	0,05	0,045833333	158
Akurasi Global			72.15%	(158 / 158)

Sumber: Peneliti 2026

Analisis Mendalam Performa Model Klasifikasi

Berdasarkan data eksperimen pada Tabel 1, model *Naive Bayes Classifier* yang dibangun mencatatkan nilai Akurasi Global sebesar 72,15**%**. Nilai akurasi ini dikategorikan cukup memuaskan (*fair classification*) mengingat data opini media sosial memiliki tingkat ambiguitas konteks yang sangat tinggi. Apabila dianalisis secara mendalam per kategori polaritas, terdapat performa unik yang dihasilkan oleh struktur probabilitas bersyarat khas Naive Bayes:

a. Analisis Kelas Positif: Nilai presisi (*precision*) mencapai 95%. Angka yang sangat tinggi ini mengindikasikan bahwa ketika model memprediksi sebuah komentar bermakna positif, probabilitas kebenaran prediksi tersebut hampir mendekati mutlak. Namun, nilai *recall* untuk kelas positif ini relatif rendah (40%), yang berarti ada cukup banyak dokumen positif asli di lapangan yang gagal dikenali dan justru salah diklasifikasikan ke dalam kelas netral oleh sistem.
b. Analisis Kelas Netral: Nilai *recall* menyentuh angka 99%. Hal ini menunjukkan model Naïve Bayes memiliki tingkat sensitivitas yang luar biasa tinggi dalam menyerap dokumen netral. Nyaris seluruh komentar netral dalam data uji berhasil dijaring dengan benar oleh sistem. Sisi negatifnya, nilai presisi jatuh di angka 69% karena dokumen dari kelas lain (terutama negatif

dan sebagian kecil positif) ikut terseret masuk dan salah dilabeli sebagai netral oleh model.

c. Analisis Kelas Negatif**:** Nilai performa mencatatkan angka 0,00. Hasil ekstrem ini terjadi akibat kendala klasik dalam pemelajaran mesin, yaitu Data Imbalance (Ketimpangan Data). Dari 158 data yang diuji, sampel untuk kelas negatif hanya berjumlah 13 ulasan (*support* = 13). Karena jumlah sebaran kelas negatif di data latih maupun data uji sangat minim, algoritma Naïve Bayes yang bekerja berdasarkan frekuensi kemunculan kata mengalami bias probabilitas prior. Model cenderung mengabaikan fitur-fitur kata bermakna negatif dan secara otomatis mengarahkan keputusan klasifikasi ke kelas mayoritas, yaitu kelas Netral.

KESIMPULAN

Penelitian ini berhasil menyusun kerangka kerja komputasional analisis sentimen untuk memetakan opini publik terkait polemik Chromebook Kemendikbudristek pada media sosial YouTube menggunakan algoritma *Naive Bayes Classifier*. Hasil ekstraksi data membuktikan bahwa opini masyarakat didominasi oleh respons Netral (60,2%), yang menunjukkan publik cenderung bersikap informatif dan menunggu implementasi riil dari kebijakan tersebut di sekolah-sekolah. Model *Naive Bayes* terbukti tangguh dengan menghasilkan Akurasi Global mencapai 72,15% serta nilai presisi kelas positif yang sangat superior sebesar 95%. Kendati demikian, kelemahan fatal algoritma NBC ditemukan saat dihadapkan pada data minoritas (kelas Negatif), di mana nilai metrik merosot hingga 0,00 akibat bias distribusi data (*data imbalance*).

DAFTAR PUSTAKA

Alita, D., & Shodiqin, R. A. (2023). Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 1–12.

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*, 10(1), 22–29.

Farah Zhafira, D., & Rahayudi, B. (2021). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(13), 5780–5787.

Fatah, Z., & Ismardani, M. T. (2026). Analisis Sentimen Komentar Pengguna YouTube Terhadap Kebijakan Penyitaan Tanah oleh Pemerintah Menggunakan Metode Naïve Bayes. *Jurnal Informatika dan Sistem Informasi*, 12(1), 59–66.

Febriyani, E., & Februariyanti, H. (2020). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter. *Jurnal Teknologi Informasi dan Komputer*, 17(1), 45–52.

Husna, A. N., & Rianto, P. (2021). Membaca Komentar di Media Sosial Sebagai Hiburan: Studi Etnografi Virtual Pengguna YouTube. *Jurnal Komunikasi*, 13(2), 145–158.

Khoirul, M., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi BRImo pada Ulasan Pengguna di Google Play Menggunakan Algoritma Naive Bayes. *Jurnal Mahasiswa*

Teknik Informatika (JATI), 7(1), 312–318.

Muslimin, M., & Lusiana, V. (2023). Analisis Sentimen Terhadap Kenaikan Harga Bahan Pokok Menggunakan Metode Naive Bayes Classifier. *Jurnal Media Informatika Budidarma*, 7(3), 1198–1206.

Nurian, A., & Nurina Sari, B. (2021). Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naive Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*, 11(3), 101–109.

Rizky Herdiansyah, M., & Yuliana, A. (2024). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naïve Bayes Berdasarkan Komentar pada YouTube. *Jurnal Mahasiswa Teknik Informatika (JATI)*, 8(6), 7120–7127.

Sebayang, S. K. H., & Sofyan, A. S. (2022). Analisis Kesalahan Berbahasa pada Sosial Media Instagram dalam Postingan, Komentar, dan Cerita Singkat. *Jurnal Serunai Bahasa Indonesia*, 19(2), 88–96.

Tutiasri, R. P., Laminto, N., & Nazri, K. (2020). Pemanfaatan Youtube Sebagai Media Pembelajaran Bagi Mahasiswa di Tengah Pandemi Covid-19. *Jurnal Komunikasi Pendidikan*, 4(2), 68–79.