

Sentiment Analysis And Characteristics Of Youtube User Opinions Toward Samsung And Iphone Brands Using TF-IDF With Naive Bayes And KNN Comparison And Mcnemar

Naurah Atikah Nurpadhilah¹⁾; Surya Ade Saputera²⁾

¹⁾Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu

²⁾ Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu

Email: ¹⁾ naurahatikah.17@gmail.com ;²⁾ adesurya2012@gmail.com

How to Cite :

Nurpadhilah, N, A., Saputera, S, A. (2026). Sentiment Analysis and Characteristics of Youtube User Opinions Toward Samsung and Iphone Brands Using TF-IDF with Naive Bayes and KNN Comparison and Mcnemar Test. Jurnal Media Computer Science, 5(2)

ARTICLE HISTORY

Received [20 Maret 2026]

Revised [28 April 2026]

Accepted [30 April 2026]

KEYWORDS

IPhone, K-Nearest Neighbor, Naive Bayes, Samsung, Sentiment Analysis, TF-IDF, YouTube.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Perkembangan media sosial, khususnya YouTube, menghasilkan data opini publik dalam jumlah besar yang dapat dimanfaatkan untuk mengetahui persepsi pengguna terhadap suatu produk. Samsung dan iPhone merupakan dua brand smartphone yang memiliki tingkat persaingan tinggi dan sering menjadi bahan diskusi pada kolom komentar YouTube. Tujuan penelitian ini adalah mengetahui perbandingan kinerja algoritma Naive Bayes dan K-Nearest Neighbor (KNN) pada analisis sentimen komentar YouTube terkait kedua brand tersebut. Data diperoleh dengan melakukan scraping komentar YouTube memakai library youtube-comment-downloader. Proses penelitian diawali dengan pengumpulan data dan tahapan pre-processing teks berupa case folding, penghilangan tanda baca, penghapusan angka, stopword removal, serta stemming menggunakan pustaka Sastrawi. Berikutnya, data dikonversi menjadi representasi numerik melalui metode TF-IDF. Tahap klasifikasi dilakukan menggunakan algoritma Naive Bayes dan KNN, kemudian dievaluasi berdasarkan akurasi, classification report, confusion matrix, dan uji McNemar untuk mengetahui signifikansi perbedaan kinerja model. Selain itu, penelitian ini juga menganalisis distribusi kata berdasarkan sentimen dan brand menggunakan visualisasi WordCloud. Hasil penelitian menyatakan bahwa kedua algoritma mampu mengklasifikasikan sentimen komentar dengan baik serta memberikan gambaran mengenai karakteristik opini pengguna terhadap Samsung dan iPhone berdasarkan komentar YouTube.

ABSTRACT

The development of social media, particularly YouTube, has generated a large amount of public opinion data that can be utilized to understand user perceptions of products. Samsung and iPhone are two smartphone brands with intense market competition and are frequently discussed in YouTube comment sections. This study aims to compare the performance of the Naive Bayes and K-Nearest Neighbor (KNN) algorithms in sentiment analysis of YouTube comments related to these two brands. The research data were collected through a YouTube comment scraping process using the youtube-comment-downloader library. The research stages included data collection, text pre-processing consisting of case folding, punctuation removal, number removal, stopword removal, and stemming using the Sastrawi library. Furthermore, the text data were transformed into

numerical representations using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The classification process was carried out using the Naive Bayes and KNN algorithms and evaluated using accuracy, classification reports, confusion matrices, and the McNemar test to determine the significance of performance differences between the models. In addition, this study also analyzed word distribution based on sentiment and brand using WordCloud visualization. The results indicate that both algorithms are capable of classifying comment sentiments effectively and provide insights into user opinion characteristics toward Samsung and iPhone based on YouTube comments.

PENDAHULUAN

Perkembangan teknologi komunikasi yang pesat telah menciptakan ketergantungan masyarakat terhadap internet, menjadikan perangkat seluler sebagai kebutuhan utama di era digital [1]. Di tengah persaingan pasar global, Samsung dan Apple (iPhone) menjadi dua pemimpin pasar *smartphone* premium yang memiliki basis pengguna loyal dengan karakteristik yang berbeda [2]. iPhone mengandalkan eksklusivitas sistem operasi iOS, sementara Samsung menitikberatkan pada inovasi fitur yang beragam [3]. Dinamika persaingan ini terekam secara masif dalam kolom komentar platform YouTube, yang menjadi sumber data opini publik yang sangat besar namun tidak terstruktur [4].

Dalam pengolahan data opini YouTube, penggunaan bahasa gaul dan bahasa tidak baku Indonesia menjadi kendala utama yang dapat menurunkan performa klasifikasi [8]. Sehingga, diperlukan tahapan *pre-processing* yang komprehensif untuk membersihkan data dari karakter non-alfabet dan melakukan normalisasi teks [4]. Agar data teks dapat diproses oleh mesin, teks diubah menjadi representasi numerik menggunakan metode pembobotan TF-IDF yang mampu mengekstraksi kata-kata penting dan unik dari setiap komentar [6,7].

Penelitian sebelumnya menyatakan bahwa algoritma Naïve Bayes Classifier memiliki efektivitas tinggi dalam pengklasifikasian sentimen otomatis pada data media sosial seperti YouTube dan Twitter [4,9,11]. Namun, untuk mendapatkan performa yang paling optimal, diperlukan perbandingan dengan metode K-Nearest Neighbor (KNN) yang dikenal memiliki performa baik dalam mengelola data dengan distribusi kompleks [5, 10]. Studi komparatif antara kedua algoritma ini telah banyak diterapkan untuk menguji tingkat akurasi dan efisiensi komputasi guna meminimalkan risiko kesalahan prediksi [12]. Meskipun berbagai penelitian telah membandingkan algoritma Naive Bayes dan KNN dalam analisis sentimen, sebagian besar penelitian tersebut belum mengkaji secara spesifik perbandingan persepsi pengguna terhadap brand *smartphone* yang berbeda berdasarkan data komentar YouTube. Selain itu, analisis yang menggabungkan pendekatan klasifikasi dengan eksplorasi distribusi kata berdasarkan sentimen dan brand masih terbatas. Untuk itu, penelitian ini dilakukan guna mengisi celah penelitian melalui analisis komparatif yang tidak hanya berorientasi pada performa algoritma, melainkan juga pada karakteristik opini pengguna dari setiap brand.

Penelitian ini bertujuan membandingkan efektivitas algoritma Naive Bayes dan KNN dalam mengelompokkan sentimen pengguna YouTube terhadap brand Samsung dan iPhone dengan mengintegrasikan metode TF-IDF. Kontribusi penelitian ini adalah memberikan analisis komprehensif terhadap performa algoritma klasifikasi serta karakteristik sentimen pengguna berdasarkan brand, sehingga mampu memberikan pemahaman yang lebih jelas mengenai pola persepsi pengguna serta efektivitas algoritma dalam mengklasifikasikan opini publik berbasis media sosial.

LANDASAN TEORI

Analisis Sentimen

Dalam pengolahan teks, teknik analisis sentimen digunakan untuk mengidentifikasi dan membagi opini atau emosi yang ada dalam teks ke dalam kategori positif, negatif, atau netral [4, 9]. Teknik ini dapat digunakan pada data media sosial secara luas karena dapat secara otomatis menggambarkan opini publik tentang suatu topik.

Dalam konteks platform YouTube, komentar pengguna menjadi sumber data yang kaya namun tidak terstruktur. Sejumlah penelitian mengungkapkan bahwa analisis sentimen terhadap komentar YouTube dapat merepresentasikan opini masyarakat dengan baik. Penelitian oleh Angdresey dkk. [13] menyatakan bahwa analisis sentimen pada komentar YouTube efektif dalam mengidentifikasi pandangan publik. Selain itu, Yaşa [14] dan Rout & Acharya [15] juga menegaskan bahwa komentar pengguna mencerminkan persepsi masyarakat terhadap berbagai isu yang dibahas dalam video.

Pre-processing Teks

Langkah awal dalam pengolahan teks adalah pre-processing, yang dilakukan untuk membersihkan dan menormalkan data sebelum analisis lanjutan. Penggabungan case, penghapusan tanda baca, penghapusan angka, penghapusan stopword, dan stemming adalah beberapa proses yang termasuk dalam kategori ini [8].

Dalam pengolahan teks berbahasa Indonesia, pre-processing memiliki peran penting karena banyaknya penggunaan bahasa tidak baku, singkatan, dan slang. Penelitian oleh Hadiprakoso dkk. [23] menyatakan bahwa penerapan pre-processing yang tepat dapat meningkatkan akurasi model secara signifikan. Hal serupa juga ditemukan oleh Lubis dkk. [24] dan Madyatmadja dkk. [25] yang menekankan pentingnya normalisasi teks dalam pengolahan dokumen. Selain itu, Khairunnisa dkk. [26] membuktikan bahwa tahapan pre-processing memiliki pengaruh langsung terhadap hasil analisis sentimen.

Pembobotan TF-IDF

Metode pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk mengubah data teks menjadi representasi numerik [6, 7]. Metode ini menilai seberapa penting suatu kata dalam dokumen relatif terhadap dokumen lain.

Secara matematis, TF-IDF dapat dirumuskan sebagai:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

dengan komponen *Inverse Document Frequency (IDF)*:

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

di mana:

1. $TF(t, d)$ = frekuensi kata t dalam dokumen d
2. N = jumlah total dokumen
3. $df(t)$ = jumlah dokumen yang mengandung kata t

Penggunaan TF-IDF dalam klasifikasi teks telah banyak diterapkan karena kemampuannya dalam mengekstraksi fitur penting. Yarsasi & Iskoko [16] menyatakan bahwa TF-IDF mampu meningkatkan performa klasifikasi dalam mendeteksi risiko *cyberbullying*. Selain itu, Zhang [17] serta Firizkiansah dkk. [18] juga menyatakan bahwa kombinasi TF-IDF dengan algoritma klasifikasi dapat menghasilkan representasi fitur yang lebih optimal.

Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi berbasis statistik yang menggunakan Teorema Bayes dengan asumsi fitur-fitur bersifat independen [4, 11]. Algoritma ini dikenal sederhana, cepat, dan efektif untuk mengolah data teks.

Secara matematis, Naive Bayes dirumuskan sebagai:

$$P(C|X) = \frac{(P(X|C) \times P(C))}{P(X)}$$

di mana:

1. $P(C|X)$ = probabilitas kelas C terhadap data X
2. $P(X|C)$ = probabilitas data terhadap kelas
3. $P(C)$ = probabilitas awal (prior)
4. $P(X)$ = probabilitas data

Naive Bayes kerap digunakan dalam analisis sentimen karena efektif menghasilkan klasifikasi yang akurat dengan kompleksitas komputasi rendah. Penelitian oleh Angdressey dkk. [13] menyatakan bahwa metode Naive Bayes tetap memiliki relevansi dalam klasifikasi sentimen modern.

Algoritma K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan algoritma klasifikasi yang menentukan kelas data melalui kedekatan terhadap data latih di sekitarnya [5, 10]. Algoritma ini tidak memerlukan proses pelatihan yang kompleks dan mampu menangani data dengan distribusi yang tidak linear.

Pengukuran jarak pada KNN umumnya menggunakan Euclidean Distance:

$$d(x,y) = \sqrt{\sum (x_i - y_i)^2}$$

di mana:

1. x dan y = dua data yang dibandingkan
2. i = dimensi fitur

Selain itu, Kartarina & Sriwinarti [21] juga menyatakan bahwa KNN memiliki performa yang kompetitif dalam klasifikasi data. Shtino & Muça [19] serta Mardiani dkk. [20] menyatakan bahwa KNN dapat memberikan hasil yang baik dibandingkan metode lain. Selain itu, Hakim & Sugiyono [22] juga membuktikan bahwa KNN efektif digunakan dalam analisis 990entiment.

METODE PENELITIAN

Jenis Penelitian

Penelitian kuantitatif berbasis text mining ini bertujuan untuk menganalisis perasaan pengguna terhadap Samsung dan iPhone berdasarkan komentar yang mereka berikan di YouTube [13]. Metode yang digunakan adalah klasifikasi teks dengan menguji perbandingan performa antara algoritma Naive Bayes dan K-Nearest Neighbor (KNN).

Sumber dan Pengumpulan Data

Data sekunder yang diperoleh dari komentar pengguna YouTube adalah sumber data penelitian ini. Data dikumpulkan melalui proses *scraping* menggunakan library youtube-comment-downloader pada bahasa pemrograman Python.

Data dikumpulkan dari dua video YouTube yang merepresentasikan masing-masing brand, yaitu:

1. Samsung: <https://www.youtube.com/watch?v=ZGbp60qeIUM>
2. iPhone: <https://www.youtube.com/watch?v=9-poYwCZxDQ>

Pengambilan data dilakukan dengan batas maksimum 1500 komentar untuk setiap video, sehingga total dataset yang digunakan berjumlah sekitar 3000 komentar.

Setiap data terdiri dari:

1. teks komentar (*text_raw*)
2. label brand (Samsung atau iPhone)

Pengambilan data dilakukan pada waktu tertentu (23 April 2026) sehingga penelitian ini dapat direplikasi.

990 | Naurah Atikah Nurpadhilah, Surya Ade Saputera; Sentiment Analysis and Characteristics of Youtube User...

Variabel Penelitian

Variabel dalam penelitian ini dibagi menjadi tiga kategori utama, yaitu variabel input, variabel proses, dan variabel output.

1. Variabel Input

Variabel input berupa data mentah (raw data) yang diambil dari komentar pengguna YouTube. Data ini berbentuk teks tidak terstruktur (*unstructured text*) yang mengandung opini pengguna terhadap brand Samsung dan iPhone [1].

2. Variabel Proses

Variabel proses merupakan hasil transformasi data yang dilakukan melalui beberapa tahapan, yaitu:

- a) hasil pre-processing teks (clean_text)
- b) representasi fitur numerik menggunakan TF-IDF

Variabel ini berperan sebagai data yang akan diproses oleh algoritma klasifikasi.

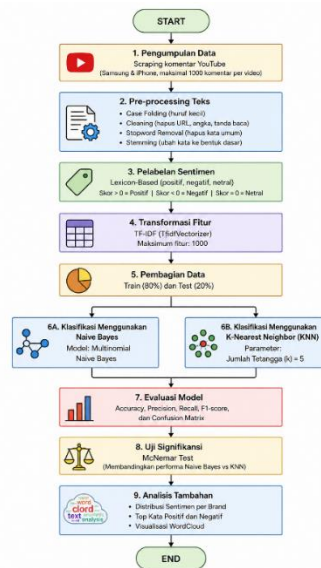
3. Variabel Output

Variabel output berupa label sentimen yang terdiri dari:

- a) positif
- b) negatif
- c) netral

Label ini dihasilkan melalui metode pelabelan berbasis leksikon (*lexicon-based labeling*).

Diagram Alur Penelitian



Gambar 1. Diagram Alur Penelitian

Proses penelitian, mulai dari pengumpulan data hingga analisis hasil, digambarkan secara sistematis dalam diagram alur penelitian di Gambar 1. Penelitian dimulai dengan mengumpulkan komentar pengguna YouTube melalui scraping. Data yang dikumpulkan masih terdiri dari teks tidak terstruktur dengan berbagai variasi Bahasa.

Tahap pre-processing kemudian digunakan untuk memproses data yang telah dikumpulkan. Pada langkah ini, pembersihan dan normalisasi teks dilakukan. Ini termasuk penyusunan case, penghapusan URL, angka, tanda baca, penghapusan stopwords, dan stemming. Tujuan dari tahap ini adalah untuk menghasilkan data teks yang lebih terstruktur dan siap untuk analisis. Selanjutnya, data dilabelkan ke dalam kategori sentimen dengan menggunakan metode berbasis leksikon untuk membagi komentar ke dalam kategori negatif, netral, atau positif. Label yang dibuat digunakan sebagai target untuk klasifikasi.

Selanjutnya, TF-IDF digunakan untuk melakukan proses transformasi fitur, yaitu mengubah data teks menjadi bentuk numerik yang dapat diproses oleh algoritma pembelajaran mesin. Selanjutnya, kumpulan data dibagi menjadi data latih dan data uji dengan rasio tertentu. Data uji digunakan untuk mengevaluasi kinerja model, sementara data latih digunakan untuk melatihnya.

Pada tahap klasifikasi, Data yang telah diproses kemudian diklasifikasikan menggunakan dua algoritma, Naive Bayes dan KNN, yang digunakan sebagai pembandingan untuk mengklasifikasikan sentimen teks. Untuk menilai performa model, hasil klasifikasi dinilai menggunakan metrik ketepatan, ketepatan, recall, skor f1, dan matriks kecacauan. Uji McNemar dilakukan untuk menguji signifikansi perbedaan performa kedua algoritma.

Tahap terakhir adalah analisis hasil, yang meliputi analisis distribusi sentimen berdasarkan brand, analisis kata-kata dominan untuk masing-masing kategori sentimen dan visualisasi dalam bentuk grafik dan wordcloud. Tujuan dari upaya ini adalah untuk mendapatkan pemahaman yang lebih mendalam tentang hasil klasifikasi.

HASIL DAN PEMBAHASAN

Tabel 1. Distribusi Label Sentimen pada Dataset

Kelas Sentimen	Jumlah Data	Persentase
Netral	2657	88,57%
Positif	232	7,73%
Negatif	111	3,70%
Total	3000	100%

Berdasarkan proses pelabelan, distribusi sentimen pada keseluruhan dataset menunjukkan ketimpangan kelas (class imbalance) yang signifikan, sebagaimana ditampilkan pada Tabel 4.1. Proses pengumpulan data menghasilkan sebanyak 3.000 komentar dari dua kanal YouTube resmi, masing-masing 1.500 komentar untuk Samsung dan iPhone. Preprocessing mencakup penyusunan case, penghapusan URL, angka, dan tanda baca, penghapusan stopword berbasis sastra, dan stemming untuk mengubah kata ke bentuk dasar. Pendekatan lexicon-based mendeteksi kata positif dan negatif untuk melakukan pelabelan sentimen otomatis.

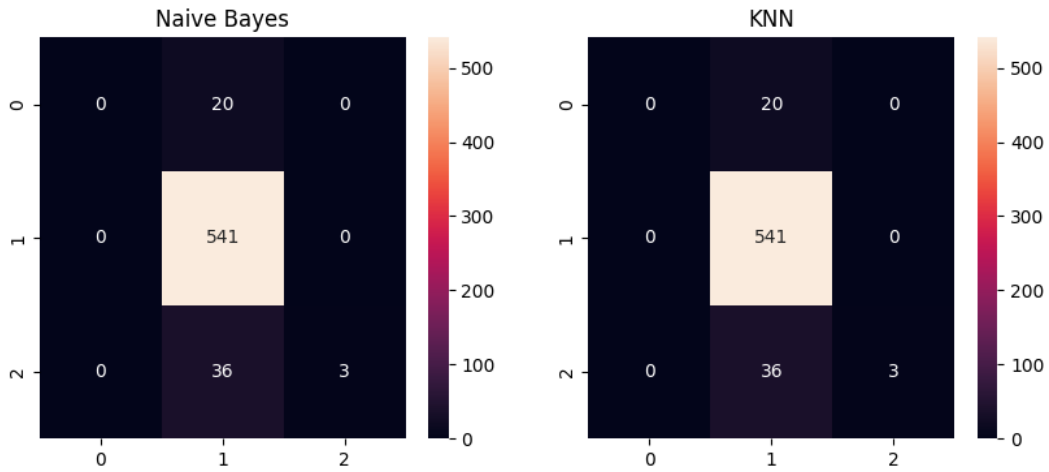
Dominasi kelas netral (88,57%) memperlihatkan bahwa sebagian besar komentar pengguna cenderung informatif, mengandung pertanyaan yang terkait dengan spesifikasi, atau tidak mengandung kata emosional yang ditemukan dalam leksikon. Selanjutnya, dataset dibagi secara acak dengan rasio 80:20 menjadi 2.400 data latih dan 600 data uji. Distribusi pada data uji adalah: Netral (541), Positif (39), dan Negatif (20).

Evaluasi Performa Klasifikasi Sentimen

Tabel 2. Metrik Evaluasi Performa Model pada Data Uji

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	F1-Score (Weighted)
Naive Bayes	90,67%	0,64	0,36	0,36	0,87
KNN (k=5)	90,67%	0,64	0,36	0,36	0,87

Data uji dievaluasi menggunakan dua algoritma, yaitu Naive Bayes (MultinomialNB) dan K-Nearest Neighbor (KNN, k=5), dengan fitur TF-IDF (max_features=1500). Hasil evaluasi kinerja model disajikan pada Tabel 2.



Gambar 2. Confusion matrix

Kedua model menunjukkan nilai accuracy yang identik (90,67%). Namun, tingginya nilai accuracy ini tidak serta-merta mencerminkan kemampuan model yang baik secara keseluruhan. Nilai Macro F1-Score yang rendah (0,36) mengonfirmasi bahwa performa model pada kelas minoritas (positif dan negatif) sangat lemah. Hal ini disebabkan oleh bias prediksi ke arah kelas mayoritas (netral) akibat distribusi data yang tidak seimbang dan keterbatasan pendekatan lexicon-based yang gagal menangkap konteks kalimat atau sarkasme. Kedua model cenderung memprediksi sebagian besar komentar sebagai netral, yang selaras dengan distribusi data uji.

Uji Signifikansi Statistik (Uji McNemar)

Tabel 3. Hasil Uji McNemar

Statistic	p-value
2,0	1,000

Untuk menghindari klaim keunggulan model tanpa dasar statistik yang valid, dilakukan uji McNemar guna membandingkan signifikansi perbedaan prediksi antara Naive Bayes dan KNN. Hasil uji disajikan pada Tabel 3.

Nilai p-value sebesar 1,000 jauh melebihi tingkat signifikansi $\alpha = 0,05$, seperti yang ditunjukkan dalam Tabel 3. Hasil ini menunjukkan bahwa perbedaan prediksi Naive Bayes dan KNN tidak signifikan secara statistik. Secara metodologis, hal ini menegaskan bahwa kedua algoritma memberikan performa yang ekuivalen ketika dihadapkan pada distribusi data yang sangat condong ke kelas netral dengan fitur TF-IDF standar.

Analisis Distribusi Kata dan Sentimen per Brand

Tabel 4. Kata Dominan per Kategori Sentimen dan Brand

Brand	Sentimen	Top Kata Dominan (Berdasarkan Frekuensi)
Samsung	Positif	bagus, samsung, suka, hp, lebih
Samsung	Negatif	mahal, hp, harga, flagship, aja
iPhone	Positif	bagus, pro, iphone, suka, kamera, keren
iPhone	Negatif	mahal, iphone, hp, aja, flagship, beli

Selain evaluasi model, penelitian ini mengidentifikasi pola kata dominan yang membentuk sentimen positif dan negatif untuk masing-masing brand. Hasil ekstraksi frekuensi kata tertinggi ditampilkan pada Tabel 4.

Berdasarkan Tabel 4, terdapat kesamaan pola persepsi pengguna terhadap kedua brand. Kata mahal dan harga muncul secara dominan pada kelas negatif untuk Samsung maupun iPhone, mengindikasikan bahwa sensitivitas harga merupakan faktor utama keluhan pengguna di kedua segmen. Di sisi positif, iPhone lebih sering dikaitkan dengan kata pro, kamera, dan keren, yang mencerminkan fokus apresiasi pengguna pada aspek fotografi dan desain premium. Adapun, komentar positif Samsung lebih berpusat pada kata suka dan hp yang bersifat umum, tanpa penekanan spesifik pada fitur unggulan.



Gambar 3. Sentimen Samsung Negatif dan Positif

Gambar 3 menampilkan visualisasi *word cloud* komentar pengguna YouTube terhadap brand Samsung berdasarkan dua kategori sentimen, yaitu positif dan negatif. Dengan visualisasi ini, kata-kata yang paling sering muncul dalam setiap kategori sentimen dapat diidentifikasi untuk menggambarkan pola persepsi konsumen terhadap merek Samsung.

Pada bagian sentimen positif, terlihat bahwa kata seperti *bagus*, *samsung*, *suka*, *kamera*, *lebih*, dan *aja* muncul dengan ukuran yang lebih besar dibandingkan kata lainnya. Hal ini menyatakan bahwa sebagian besar pengguna memberikan apresiasi terhadap kualitas produk Samsung, khususnya pada aspek kamera dan pengalaman penggunaan secara umum. Dominasi kata *bagus* mengindikasikan bahwa pengguna memiliki persepsi yang cukup baik terhadap performa dan fitur yang ditawarkan Samsung. Selain itu, munculnya kata *kamera* memperlihatkan bahwa kualitas fotografi menjadi salah satu faktor yang paling sering dibicarakan secara positif oleh pengguna.

Di sisi lain, pada sentimen negatif terlihat dominasi kata seperti *mahal*, *samsung*, *harga*, *beli*, *android*, dan *flagship*. Kata *mahal* menjadi kata yang paling dominan dalam komentar negatif, yang menyatakan bahwa harga produk Samsung masih menjadi keluhan utama bagi sebagian pengguna. Kemunculan kata *flagship* dan *android* mengindikasikan bahwa komentar negatif banyak berkaitan dengan perbandingan harga perangkat premium Samsung terhadap fitur yang diperoleh pengguna. Selain itu, terdapat pula kata *udah* dan *aja* yang menunjukkan pola bahasa informal khas media sosial YouTube.

Berdasarkan visualisasi tersebut, dapat disimpulkan bahwa persepsi pengguna terhadap Samsung cenderung dipengaruhi oleh dua faktor utama, yaitu kualitas fitur dan harga produk. Sentimen positif lebih banyak berkaitan dengan kualitas kamera dan performa perangkat, sedangkan sentimen negatif lebih dominan membahas harga yang dianggap relatif mahal. Hasil ini menyatakan bahwa meskipun Samsung memperoleh apresiasi tinggi dari pengguna, faktor harga tetap menjadi aspek yang memengaruhi opini negatif pengguna terhadap brand tersebut.



Gambar 4. Sentimen Iphone Negatif dan Positif

Gambar 4 menampilkan visualisasi *word cloud* komentar pengguna YouTube terhadap brand iPhone berdasarkan kategori sentimen positif dan sentimen negatif. Visualisasi ini bertujuan untuk memperlihatkan distribusi kata yang paling sering digunakan pengguna dalam menyampaikan opini mereka terhadap produk iPhone.

Pada sentimen positif, kata yang paling dominan adalah *bagus*, *iphone*, *pro*, *suka*, *kamera*, dan *warna*. Dominasi kata *bagus* menyatakan bahwa mayoritas pengguna memberikan penilaian positif terhadap kualitas produk iPhone. Kemunculan kata *kamera* dan *pro* menyatakan bahwa seri iPhone Pro menjadi salah satu topik yang paling banyak diapresiasi, terutama terkait kualitas kamera dan performa perangkat. Selain itu, kata *warna* juga cukup sering muncul, yang mengindikasikan bahwa desain visual dan pilihan warna perangkat menjadi daya tarik tersendiri bagi pengguna iPhone.

Adapun, pada sentimen negatif terlihat bahwa kata *mahal* menjadi kata yang paling dominan dibandingkan kata lainnya. Selain itu, muncul pula kata seperti *iphone*, *harga*, *android*, *beli*, dan *flagship*. Dominasi kata *mahal* menyatakan bahwa harga perangkat iPhone menjadi faktor utama munculnya sentimen negatif pengguna. Banyak pengguna membandingkan harga iPhone dengan perangkat Android lain yang dianggap memiliki spesifikasi serupa dengan harga yang lebih rendah. Kemunculan kata *android* pada sentimen negatif menunjukkan adanya kecenderungan pengguna untuk melakukan perbandingan antar brand dalam menilai nilai produk yang diperoleh. Secara keseluruhan, hasil visualisasi pada Gambar 4 menyatakan bahwa persepsi pengguna terhadap iPhone didominasi oleh apresiasi terhadap kualitas kamera, desain, dan performa perangkat pada sentimen positif. Namun demikian, tingginya harga produk tetap menjadi sumber utama munculnya sentimen negatif. Pola ini memperlihatkan bahwa meskipun iPhone memiliki citra premium dan kualitas yang baik di mata pengguna, aspek harga masih menjadi pertimbangan penting dalam membentuk opini publik terhadap brand tersebut.

Pembahasan

Hasil penelitian ini memberikan beberapa temuan penting yang perlu didiskusikan secara kritis. Pertama, tingginya proporsi kelas netral (88,57%) dan rendahnya Macro F1-Score (0,36) mengonfirmasi bahwa pendekatan pelabelan lexicon-based tanpa validasi manual memiliki keterbatasan dalam menangkap nuansa bahasa informal YouTube. Banyak komentar yang bersifat netral karena hanya menanyakan spesifikasi, membandingkan seri, atau menggunakan bahasa gaul yang tidak terdaftar dalam leksikon. Temuan ini selaras dengan penelitian Khairunnisa dkk. (2021) yang menekankan bahwa preprocessing dan kamus leksikon yang tidak kontekstual akan menurunkan akurasi klasifikasi teks informal.

Kedua, hasil identik antara Naive Bayes dan KNN serta hasil uji McNemar ($p = 1,000$) menyatakan bahwa performa model pada studi ini lebih didominasi oleh karakteristik distribusi data daripada kompleksitas algoritma. Ketika dataset sangat tidak seimbang (*imbalanced*), algoritma sederhana cenderung bias ke kelas mayoritas. Hal ini sejalan dengan temuan Shtino & Muça (2023) dan Hakim & Sugiyono (2024) yang menyimpulkan bahwa tanpa teknik penanganan *imbalance* (seperti SMOTE atau *class weighting*), perbandingan akurasi antar-algoritma menjadi tidak

informatif. Ketiga, dari sisi implikasi praktis, analisis kata kunci menyatakan bahwa strategi pemasaran smartphone premium di Indonesia perlu menekankan value for money dan keunggulan fitur spesifik (seperti kamera untuk iPhone). Keluhan mahal yang konsisten pada kedua brand mengindikasikan bahwa persepsi nilai (perceived value) menjadi penentu utama sentimen positif, bukan sekadar merek.

Keterbatasan dan Rekomendasi Penelitian Selanjutnya Penelitian ini memiliki beberapa keterbatasan yang diakui secara transparan: (1) pelabelan leksikon otomatis tanpa inter-annotator agreement manusia, (2) ketidakseimbangan kelas yang ekstrem, dan (3) penggunaan hyperparameter default pada KNN dan Naive Bayes. Untuk penelitian selanjutnya, disarankan untuk: (1) melakukan pelabelan manual atau semi-supervised learning untuk meningkatkan kualitas ground truth, (2) menerapkan teknik resampling (SMOTE/ADASYN) atau class weights, dan (3) menguji model berbasis deep learning seperti IndoBERT yang mampu menangkap konteks kalimat serta slang secara lebih akurat.

KESIMPULAN DAN SARAN

Kesimpulan

1. Performa Algoritma Identik dengan Akurasi Tinggi namun Bias Kelas: Algoritma Naive Bayes dan K-Nearest Neighbor (KNN) yang dikombinasikan dengan metode TF-IDF menghasilkan tingkat akurasi yang identik, yaitu sebesar 90,67%. Namun, kinerja secara rinci menunjukkan ketimpangan yang signifikan: kelas netral mendominasi dengan recall mencapai 1,00 dan f1-score 0,95, sementara kelas positif dan negatif memiliki recall sangat rendah (0,08 dan 0,00). Hal ini mengindikasikan bahwa kedua model cenderung memprediksi mayoritas komentar sebagai netral, yang secara langsung selaras dengan distribusi data uji yang mengalami ketidakseimbangan kelas (class imbalance) ekstrem.
2. Tidak Ada Perbedaan Signifikan Secara Statistik: Hasil Uji McNemar menghasilkan p-value sebesar 1,000 dengan statistik uji sebesar 2,0. Berdasarkan batas signifikansi $\alpha = 0,05$, nilai p-value $> 0,05$ menyatakan bahwa tidak terdapat perbedaan signifikan secara statistik antara performa prediksi algoritma Naive Bayes dan KNN. Keduanya memberikan pola prediksi yang serupa akibat dominasi kelas netral pada data uji.
3. Keterbatasan Pelabelan Berbasis Leksikon: Rendahnya performa pada kelas minoritas (positif dan negatif) terutama disebabkan oleh keterbatasan pendekatan pelabelan berbasis leksikon (lexicon-based) yang digunakan. Pendekatan ini hanya mengandalkan kemunculan kata positif/negatif tanpa mampu menangkap konteks kalimat, struktur tata bahasa, maupun sarkasme yang sering muncul dalam komentar media sosial.
4. Pola Kata Dominan dan Persepsi Merek: Analisis frekuensi kata dominan mengungkapkan perbedaan karakteristik sentimen antar merek.
5. Samsung: Sentimen positif didominasi oleh kata bagus, samsung, suka, hp, lebih, sedangkan sentimen negatif didominasi oleh mahal, hp, harga, flagship, aja.
6. iPhone: Sentimen positif didominasi oleh bagus, pro, iphone, suka, kamera, keren, sedangkan sentimen negatif didominasi oleh mahal, iphone, hp, aja, flagship, beli. Kata mahal dan flagship muncul secara konsisten pada kedua sentimen negatif, mengindikasikan bahwa isu harga dan positioning produk merupakan titik kritik utama konsumen terhadap kedua merek smartphone premium tersebut.

Saran

Penanganan Ketidakseimbangan Kelas: Penelitian selanjutnya disarankan untuk menerapkan teknik penanganan class imbalance seperti SMOTE (Synthetic Minority Over-Sampling Technique), undersampling, atau penerapan class weights pada fungsi kerugian (loss function) algoritma. Hal ini

bertujuan untuk meningkatkan sensitivitas model terhadap kelas minoritas (positif dan negatif) agar tidak selalu terprediksi sebagai netral.

1. Peningkatan Metode Pelabelan: Mengganti atau mengombinasikan pelabelan berbasis leksikon dengan pendekatan kontekstual seperti fine-tuning model berbasis Transformer (misalnya IndoBERT atau IndoBERTtweet) atau melakukan pelabelan manual parsial pada subset data. Pendekatan ini mampu menangkap nuansa konteks, sarkasme, dan kalimat tidak baku yang sering muncul dalam komentar YouTube.
2. Pengembangan Model dengan Arsitektur Lanjutan atau Ensemble: Menerapkan model berbasis deep learning (seperti LSTM, Bi-LSTM, atau Transformer) atau metode ensemble (Voting/Stacking) yang terbukti lebih robust dalam menangani data tidak seimbang dan kompleksitas bahasa Indonesia informal di media sosial.
3. Ekspansi Dataset dan Multi-Platform: Memperluas cakupan dataset tidak hanya dari YouTube, tetapi juga mengintegrasikan data dari platform media sosial lain (Twitter/X, TikTok, Instagram, atau forum teknologi) untuk meningkatkan generalisasi dan validitas eksternal hasil klasifikasi.

Implikasi Praktis untuk Pengembang dan Pemasaran: Hasil analisis kata dominan dapat dimanfaatkan oleh tim riset pasar, pengembang produk, dan divisi pemasaran Samsung maupun Apple. Isu seperti mahal dan flagship yang muncul di sentimen negatif dapat menjadi bahan evaluasi strategi harga dan komunikasi nilai produk, sementara kata-kata seperti kamera, keren, dan pro pada iPhone atau bagus dan suka pada Samsung dapat dioptimalkan sebagai key message dalam kampanye pemasaran berikutnya.

DAFTAR PUSTAKA

- Putra, D. D., & Nurtantiono, A. (2021). Analisis pengaruh kualitas produk, harga dan layanan purna jual terhadap keputusan pembelian HP iPhone. *Jurnal Ekobis Dewantara*, 4(2), 69–77.
- Kulle, C. S., Amalina, N. A., Giska, G., Syam, M. I., & Husain, F. (2025). Analisis perbandingan persepsi konsumen terhadap Samsung dan iPhone di Indonesia. *Jurnal Interdisipliner*, 1(5), 88–100.
- Rizal, M., & Malihah, L. (2022). Analisis SWOT pada produk smartphone Apple dan Samsung. *Jurnal EKOBIS-DA: Jurnal Ekonomi dan Bisnis*, 3(1), 1–13.
- Hudha, M., Supriyati, E., & Listyorini, T. (2022). Analisis sentimen pengguna YouTube terhadap tayangan #MataNajwaMenantiTerawan dengan metode Naive Bayes Classifier. *JIKO (Jurnal Informatika dan Komputer)*, 5(1), 1–6.
- Rinanda, P. D., Delvika, B., Nurhidayarnis, S., Abror, N., & Hidayat, A. (2022). Perbandingan klasifikasi antara Naive Bayes dan K-Nearest Neighbor terhadap risiko diabetes pada ibu hamil. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 68–75.
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis sentimen kebijakan Kampus Merdeka menggunakan Naive Bayes dan pembobotan TF-IDF berdasarkan komentar pada YouTube. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, 2(1).
- Dyantono, A. M. D., & Putra, R. E. (2023). Perbandingan Sent2Vec TF-IDF Logistic Regression dan Word2Vec CNN pada hasil sentiment analysis YouTube comment. *Journal of Informatics and Computer Science (JINACS)*, 5(1), 63–72.
- Shevira, S., Suarjaya, I. M. A. D., & Buana, P. W. (2022). Pengaruh kombinasi dan urutan pre-processing pada tweets bahasa Indonesia. *JITTER: Jurnal Ilmiah Teknologi dan Komputer*, 3(2).
- Ismardani, T., & Fatah, Z. (2025). Sentiment analysis of YouTube user comments on government policies using the Naive Bayes method. *Journal of Data Insights*, 3(2), 113–123.
- Miftakhurahmat, M. A., Safitri, N., Kusnadi, P. A., & Rozikin, C. (2023). Klasifikasi pengguna hashtag pada aplikasi TikTok menggunakan perbandingan metode K-Nearest Neighbors dan Naive Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*, 11(3).

- Utami, E. (2022). Comparison Naive Bayes Classifier, K-Nearest Neighbor and Support Vector Machine in the classification of individual on Twitter account. *Jurnal Teknik Informatika (JUTIF)*, 3(3), 673–680.
- Utari, R. N. (2025). Identifikasi tingkat keaktifan media sosial terhadap pola belajar mahasiswa dengan K-Nearest Neighbor dan Naïve Bayes. *Jurnal Manajemen, Pendidikan dan Ilmu Komputer*, 2(2), 67–73.
- Angdresey, A., Sitanayah, L., & Tangka, I. L. H. (2025). Sentiment analysis for political debates on YouTube comments using BERT labeling, random oversampling, and Multinomial Naïve Bayes. *Journal of Computing Theories and Applications*, 2(3), 342–354.
- Yaşa, H. (2025). Sentiment analysis of news videos about artificial intelligence in Turkey: A YouTube case. *Revista de Comunicaci3n de la SEECI*, 58, 1–22.
- Rout, L., & Acharya, S. (2025). Does YouTube assist in mitigating global warming? A content analysis of YouTube videos and sentiment analysis through viewers' comments. *Transactions in Earth, Environment, and Sustainability*, 3(1), 51–71.
- Yarsasi, S., & Iskoko, A. (2025). A text classification approach for detecting cyberbullying risk on Twitter using Support Vector Machine with Naive Bayes and Random Forest comparison. *International Journal of Informatics and Information Systems*, 8(4), 175–187.
- Zhang, L. (2025). Improved web page categorization with semantic-aware focused crawling using GloVe and TF-IDF. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 127, 6569–6586.
- Firizkiansah, A., Muhammad, A., & Maulana, I. R. (2025). Optimasi klasifikasi data teks menggunakan algoritma Logistic Regression dengan TF-IDF dan SMOTE. *JIKOMTI: Jurnal Ilmiah Ilmu Komputer dan Teknologi Informasi*, 2(1), 29–36.
- Shtino, V. B., & Muça, M. (2023). Comparative study of K-NN, Naive Bayes and SVM for face expression classification techniques. *Balkan Journal of Interdisciplinary Research*, 9(3), 23–32.
- Mardiani, E., Rahmansyah, N., Ningsih, S., Lantana, D. A., Wirawan, A. S. P., Wijaya, S. A., & Putri, D. N. (2023). Komparasi metode KNN, Naive Bayes, Decision Tree, Ensemble, dan Linear Regression terhadap analisis performa pelajar SMA. *Innovative: Journal of Social Science Research*, 3(2).
- Kartarina, K., & Sriwinarti, N. K. (2021). Analisis metode K-Nearest Neighbors (K-NN) dan Naive Bayes dalam memprediksi kelulusan mahasiswa. *Jurnal Teknologi Informasi dan Multimedia*, 3(2), 106–112.
- Hakim, Z. R., & Sugiyono, S. (2024). Analisa sentimen terhadap kereta cepat Jakarta–Bandung menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor. *Jurnal Sains dan Teknologi*, 5(3), 939–945.
- Hadiprakoso, R. B., Setiawan, H., Yasa, R. N., & Girinoto. (2023). Text preprocessing for optimal accuracy in Indonesian sentiment analysis using a deep learning model with word embedding. In *AIP Conference Proceedings* (Vol. 2680, No. 1, p. 020050). AIP Publishing.
- Lubis, A. R., Safitri, H. R., Lubis, M., & Nugroho, O. (2023). Implementation of preprocessing in text summarization techniques for Indonesian language documents using the Flax T5 approach. In *2023 11th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1–6). IEEE.
- Madyatmadja, E. D., Yahya, B. N., & Wijaya, C. (2022). Contextual text analytics framework for citizen report classification: A case study using the Indonesian language. *IEEE Access*, 10, 31432–31444.
- Khairunnisa, S., Adiwijaya, A., & Al Faraby, S. (2021). Pengaruh text preprocessing terhadap analisis sentimen komentar masyarakat pada media sosial Twitter (studi kasus pandemi COVID-19). *Jurnal Media Informatika Budidarma*, 5(2), 406.