

Sentiment Analysis of Patient Reviews on Public Health Center Services in Pekanbaru City Using the Naive Bayes Classifier Method

Analisis Sentimen Pasien Pada Ulasan Layanan Puskesmas Sekota Pekanbaru Menggunakan Metode Naive Bayes Classifier

Michael Jordan Sirait

Universitas Lancang Kuning, Indonesia

E-mail: Michael@gmail.com

*Correspondence: Michael@gmail.com¹

Abstract

The rapid development of information technology has encouraged the public to share opinions regarding healthcare services through digital platforms such as Google Maps. These patient reviews can be utilized to evaluate the quality of public health center (Puskesmas) services in a more objective and data-driven manner. This study aims to analyze patient sentiment toward Puskesmas services across Pekanbaru City using the Naive Bayes Classifier (NBC) method. The research data were collected through a web scraping technique from Google Maps, resulting in 856 patient reviews. The research stages included data preprocessing (case folding, cleansing, normalization, tokenizing, stopword removal, and stemming), sentiment labeling (positive, negative, and neutral), TF-IDF weighting, classification using NBC, and model evaluation. The initial sentiment distribution consisted of 480 neutral reviews, 259 positive reviews, and 117 negative reviews. To address data imbalance, the SMOTE method was applied. The evaluation results using a 70% training and 30% testing split showed an accuracy of 71.60%. After applying SMOTE, the accuracy increased to 78.21%, while the implementation of Chi-Square feature selection produced the highest accuracy of 80.56%. Meanwhile, PCA and LDA achieved accuracies of 71.06% and 57.41%, respectively. Word cloud visualization and TF-IDF analysis revealed dominant words such as "service," "friendly," "good," and "bad." The findings indicate that the Naive Bayes method is effective for classifying patient review sentiments, particularly when combined with Chi-Square feature selection. This study is expected to provide a basis for improving the quality of Puskesmas services in Pekanbaru City.

Keywords: Sentiment Analysis, Naive Bayes Classifier, Public Health Center, Web Scraping, TF-IDF

Abstrak

Perkembangan teknologi informasi mendorong masyarakat untuk menyampaikan opini terhadap layanan kesehatan melalui platform digital seperti Google Maps. Ulasan pasien tersebut dapat dimanfaatkan untuk mengevaluasi kualitas pelayanan Puskesmas secara lebih objektif dan berbasis data. Penelitian ini bertujuan untuk menganalisis sentimen pasien terhadap layanan Puskesmas sekota Pekanbaru menggunakan metode Naive Bayes Classifier (NBC). Data penelitian diperoleh melalui teknik web scraping pada Google Maps dan menghasilkan 856 ulasan pasien. Tahapan penelitian meliputi preprocessing data (case folding, cleansing, normalization, tokenizing, stopword removal, dan stemming), pelabelan sentimen (positif, negatif, dan netral), pembobotan TF-IDF, klasifikasi menggunakan NBC, serta evaluasi model. Distribusi awal sentimen menunjukkan 480 ulasan netral, 259 positif, dan 117 negatif. Untuk mengatasi ketidakseimbangan data, diterapkan metode SMOTE. Hasil pengujian dengan pembagian data 70% training dan 30% testing menunjukkan akurasi sebesar 71,60%. Setelah penerapan SMOTE, akurasi meningkat menjadi 78,21%, dan penggunaan Chi-Square Feature Selection menghasilkan akurasi tertinggi sebesar 80,56%. Sementara itu, penerapan PCA dan LDA masing-masing menghasilkan akurasi 71,06% dan 57,41%. Visualisasi wordcloud dan analisis TF-IDF menunjukkan kata dominan seperti "layan", "ramah", "bagus", dan "buruk". Hasil penelitian menunjukkan bahwa metode Naive Bayes efektif dalam

mengklasifikasikan sentimen ulasan pasien, khususnya setelah dilakukan seleksi fitur Chi-Square. Penelitian ini diharapkan dapat menjadi dasar evaluasi dan perbaikan kualitas layanan Puskesmas di Kota Pekanbaru secara berkelanjutan.

Kata Kunci: Analisis Sentimen, Naïve Bayes Classifier, Puskesmas, Web Scraping, TF-IDF

Received: 13-11-2025 | Revised: 05-02-2026 | Accepted: 11-02-2026

I-JAICL is licensed under a Creative Commons Attribution-Share Alike 4.0 International License



1. Pendahuluan

Perkembangan teknologi, terutama dalam bidang informasi dan komunikasi, telah membawa dampak signifikan terhadap sektor kesehatan. Layanan kesehatan merupakan salah satu pilar utama dalam pembangunan masyarakat yang sehat dan *produktif*. Dalam penyelenggaraan pelayanan kesehatan diperlukan adanya fasilitas kesehatan yaitu alat dan tempat yang digunakan untuk penyelenggaraan pelayanan kesehatan. Ada berbagai ragam penyedia fasilitas layanan kesehatan diantaranya adalah rumah sakit, balai kesehatan, puskesmas, hingga klinik kesehatan, semua fasilitas ini memiliki peran yang penting dalam menjaga dan meningkatkan derajat kesehatan masyarakat [1].

Pusat kesehatan masyarakat atau yang biasa disebut dengan Puskesmas berfungsi sebagai institusi kesehatan pemerintah yang menjangkau masyarakat secara langsung dan mudah di akses [2]. Seiring dengan bertambahnya jumlah penduduk dan kebutuhan akan layanan kesehatan yang berkualitas, penting bagi Puskesmas untuk terus meningkatkan kualitas layanan yang diberikan. Salah satu cara untuk memancarkan kualitas layanan adalah melalui umpan balik dari pasien. Ulasan atau *feedback* yang diberikan oleh pasien dapat memberikan wawasan yang berharga mengenai pengalaman mereka saat menggunakan layanan Puskesmas.

Ulasan ini mencakup berbagai aspek, seperti kualitas pelayanan, sikap tenaga medis, fasilitas yang tersedia, serta kepuasan secara keseluruhan. Dengan memahami sentimen pasien, pengelola Puskesmas dapat mengidentifikasi kekuatan dan kelemahan dalam layanan yang diberikan, serta menyusun strategi perbaikan yang tepat. Namun analisis terhadap ulasan pasien sering kali dilakukan secara manual, sehingga memakan waktu dan tenaga. Selain itu, manual analisis dapat menghasilkan interpretasi yang subjektif dan tidak konsisten. Oleh karena itu, diperlukan pendekatan yang lebih sistematis dan efisien untuk menganalisis sentimen dari ulasan pasien. Di era digital saat ini, di mana banyak pasien yang berbagi pengalaman mereka melalui *platform online*, sentimen berbasis data menjadi semakin relevan. Untuk mengetahui pendapat pasien terhadap layanan puskesmas di Google Maps, langkah yang perlu dilakukan adalah mengumpulkan data ulasan dan menganalisisnya menggunakan metode analisis sentimen. Analisis sentimen sendiri merupakan teknik untuk memahami pandangan atau opini seseorang terhadap suatu topik tertentu [3].

Metode Naïve Bayes merupakan salah satu teknik yang populer dalam analisis sentimen, terutama dalam pengolahan bahasa alami (*Natural Language Processing*). Metode ini didasarkan pada teori probabilitas dan menggunakan asumsi independensi antar fitur, yang memungkinkan untuk mengklasifikasikan data teks dengan cepat dan akurat. *Naïve Bayes Classifier* telah terbukti efektif dalam berbagai aplikasi, termasuk analisis sentimen, karena kemampuannya menangani data yang besar dan kompleks. Dengan menggunakan metode analisis sentimen ini dapat dilakukan secara otomatis untuk mengidentifikasi apakah ulasan pasien bersifat *Positif*, *Negatif*, atau *Netral*. Hal ini tidak hanya menghemat waktu, tetapi juga meningkatkan akurasi dalam pengolahan data. [4]

Dalam konteks analisis penelitian ini, sentimen terhadap ulasan layanan Puskesmas di Kota Pekanbaru menggunakan metode *Naïve Bayes Classifier* bertujuan untuk memberikan gambaran yang lebih jelas mengenai persepsi pasien terhadap layanan yang diberikan. Dengan menganalisis sentimen dari ulasan pasien, diharapkan dapat diperoleh informasi yang mendalam mengenai tingkat kepuasan pasien, serta aspek-aspek layanan yang perlu diperbaiki. Hasil dari analisis ini diharapkan dapat memberikan rekomendasi yang konkret bagi pengelola Puskesmas dalam upaya meningkatkan kualitas layanan kesehatan. Selain itu, penelitian ini juga berkontribusi pada pengembangan ilmu pengetahuan di bidang analisis data dan kesehatan masyarakat. Dengan memanfaatkan teknologi dan metode analisis yang canggih, diharapkan dapat tercipta sistem yang lebih *responsif* terhadap kebutuhan masyarakat. Dalam jangka panjang, peningkatan kualitas layanan Puskesmas akan berdampak *Positif* pada kesehatan masyarakat secara keseluruhan, serta meningkatkan kepercayaan masyarakat terhadap layanan kesehatan yang disediakan oleh pemerintah. Dengan latar belakang tersebut, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam memahami sentimen pasien terhadap layanan Puskesmas di Kota Pekanbaru, serta memberikan dasar yang kuat untuk perbaikan dan pengembangan layanan kesehatan di masa yang akan datang.

2. Metode Penelitian

Adapun beberapa tahapan pada penelitian dari jelaskan pada gambar 1.

2.1 Studi Literatur

Studi literatur ini disusun untuk memberikan landasan teoritis yang kuat serta membandingkan penelitian sebelumnya yang relevan dengan topik analisis sentimen terhadap ulasan layanan Puskesmas menggunakan metode *Naïve Bayes*. Dalam beberapa tahun terakhir, analisis sentimen menjadi pendekatan yang banyak digunakan untuk menggali opini publik, terutama dalam bidang kesehatan dan layanan publik lainnya. Metode *Naïve Bayes Classifier* (NBC) dikenal sebagai salah satu algoritma klasifikasi teks berbasis probabilistik yang sederhana namun efektif dalam mengelola data dalam jumlah besar dan kompleks.

2.2 Pengumpulan Data/Scraping

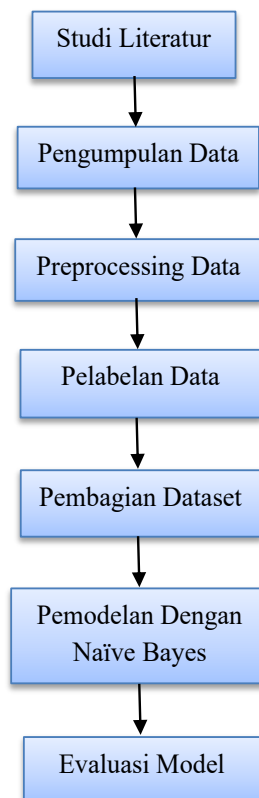
Dalam penelitian ini, teknik pengumpulan data dilakukan melalui pendekatan web *Scraping*, yaitu metode otomatis untuk mengekstraksi informasi dari situs web menggunakan bahasa pemrograman. Teknik ini dipilih karena data yang dianalisis berupa ulasan teks pasien terhadap layanan Puskesmas yang dipublikasikan secara online melalui *platform* Google Maps. Alasan pemilihan Google Maps sebagai sumber data adalah karena *platform* ini secara umum digunakan oleh masyarakat untuk memberikan ulasan dan penilaian langsung terhadap layanan fasilitas kesehatan, termasuk Puskesmas.

2.3 Preprocessing Data

Pada tahap praproses data, dilakukan text mining yang berfungsi untuk menganalisis teks guna memperoleh informasi yang berguna sesuai dengan tujuan tertentu. Proses ini mencakup dua tahap, yaitu penerapan struktur pada sumber data teks dan dilanjutkan dengan ekstraksi informasi serta pengetahuan yang relevan. Tahapan *Preprocessing* yang digunakan dalam penelitian ini meliputi *Case Folding*, *Cleansing*, *Normalization*, *Tokenizing*, *Stopword Removal* dan *Stemming*

2.4 Pelabelan Data

Pelabelan data merupakan salah satu tahap penting dalam proses analisis sentimen, terutama dalam konteks supervised learning, di mana algoritma seperti *Naïve Bayes Classifier* membutuhkan data latih yang telah diberi label kelas tertentu. Dalam penelitian ini, pelabelan dilakukan terhadap ulasan pasien yang telah dikumpulkan dari Google Maps, dengan tujuan untuk mengkategorikan setiap ulasan ke dalam kelas sentimen yang sesuai, yaitu *Positif* *Netral* atau *Negatif*.



Gambar 1. Tahap – tahap Penelitian

2.5 Pembagian Dataset

Dataset yang telah melalui proses pelabelan kemudian dibagi menjadi dua bagian, yaitu data latih dan data uji. Pada pembagian dataset ini melakukan 3 pembagian, yaitu pembagian dataset 70%: 30%, 80%: 20% dan 90%:10%. Pembagian ini dilakukan secara stratified, yaitu dengan mempertahankan distribusi proporsi setiap kelas sentimen (*Positif*, *Netral*, dan *Negatif*) agar tetap seimbang dalam masing-masing subset. Data latih digunakan dalam proses pembangunan model, sementara data uji dimanfaatkan untuk melakukan evaluasi kinerja model secara menyeluruh.

2.6 Pemodelan Dengan Algoritma Naive Bayes Classifier

Setelah proses pengumpulan data selesai, tahap selanjutnya adalah penyajian hasil dan pembahasan. Adapun langkah-langkah yang dilakukan pada tahap ini meliputi.

a. Pembobotan TF-IDF

Proses pembobotan dilakukan dengan memecah kalimat menjadi kata-kata individual, kemudian menetapkan bobot numerik untuk setiap kata menggunakan metode TF-IDF. Metode ini merupakan pendekatan statistik yang digunakan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen yang ada.

b. Klasifikasi

Pada tahap ini, dilakukan proses klasifikasi terhadap data teks yang telah diproses sebelumnya. Proses pemodelan dimulai dengan membagi data ke dalam dua kelompok, yaitu data pelatihan (*training*) dan data pengujian (*Testing*). Algoritma yang digunakan untuk melakukan klasifikasi ini adalah metode *Naïve Bayes*.

c. Hasil Analisis

Pada tahap ini, hasil klasifikasi diwujudkan dalam bentuk pengelompokan opini sentimen masyarakat berdasarkan kemunculan kata-kata tertentu, yang selanjutnya dikategorikan ke dalam kelas sentimen, yakni *Positif* dan *Negatif*.

d. Visualisasi Data

Pada tahap ini, dilakukan visualisasi kata-kata dalam dataset menggunakan *word cloud* berdasarkan frekuensi kemunculannya dalam masing-masing kategori sentimen, yaitu *Positif* dan *Negatif*. Visualisasi tersebut kemudian dianalisis dengan membandingkan antara frekuensi kata dan distribusi kelas yang dihasilkan dengan kondisi nyata yang terjadi di lapangan.

2.7 Evaluasi Model

Pengujian model merupakan tahapan penting dalam proses analisis sentimen untuk mengevaluasi kinerja algoritma klasifikasi yang digunakan, dalam hal ini adalah *Naïve Bayes Classifier*. Tujuan dari pengujian ini adalah untuk mengetahui seberapa baik model dalam mengklasifikasikan data ulasan ke dalam kelas sentimen yang sesuai, yakni *Positif*, *Negatif*, dan *Netral*. Untuk itu, dilakukan serangkaian langkah sistematis mulai dari pembagian dataset, pelatihan model, hingga evaluasi hasil klasifikasi.

3. Hasil dan Pembahasan

Pada tahap ini akan di tampilkan hasil yang di peroleh dari setiap proses yang telah di lakukan sebelumnya:

3.1 Pengambilan Data (Scraping)

Dalam penelitian ini, data ulasan pada Puskesmas sekota Pekanbaru dengan cara *Scraping* data yang diambil dari situs Google Maps menggunakan alat bantu *Web Scraper*, yaitu ekstensi yang tersedia pada peramban *Google Chrome*. *Web Scraper* memungkinkan pengguna untuk mengambil data secara otomatis dari halaman web dengan menyusun sitemap dan menentukan elemen-elemen HTML yang ingin dikumpulkan. Pengambilan data dilakukan dengan menggunakan ekstensi *Web Scraper* yang tersedia pada peramban *Google Chrome*. Langkah pertama yang dilakukan adalah menginstal ekstensi *Web Scraper* melalui *Chrome Web Store*. Setelah ekstensi berhasil diinstal, proses dilanjutkan dengan membuat sebuah sitemap, yaitu struktur yang menggambarkan alur navigasi halaman web yang akan diambil datanya. Pada tahap ini, URL halaman ulasan Puskesmas Pekanbaru di situs Google maps ditetapkan sebagai target *Scraping*.

Selanjutnya, dilakukan proses penandaan elemen-elemen HTML yang berisi data ulasan yang dibutuhkan. Elemen-elemen tersebut meliputi nama pengguna, tanggal ulasan, dan isi ulasan. Penandaan dilakukan dengan menggunakan fitur Selector yang disediakan oleh *Web Scraper*, di mana setiap elemen ditandai secara manual berdasarkan struktur HTML nya agar dapat dikenali dan dikumpulkan oleh sistem scraper. Setelah semua elemen berhasil ditandai dan sitemap dikonfigurasi dengan benar, proses *Scraping* dijalankan. *Web Scraper* akan secara otomatis mengakses setiap halaman ulasan sesuai struktur sitemap dan mengumpulkan data yang telah ditentukan. Data yang berhasil dikumpulkan kemudian diekspor ke dalam format *xlsx* agar lebih mudah digunakan dalam proses analisis data selanjutnya.

3.2 Pre – Processing Data

Tahapan preprocessing merupakan proses awal dalam pengolahan data teks sebelum dilakukan ekstraksi fitur dan klasifikasi. Tujuan utama preprocessing adalah untuk membersihkan dan menyiapkan data teks agar lebih terstruktur dan dapat dikenali dengan baik oleh algoritma klasifikasi. Pada penelitian ini, preprocessing dilakukan terhadap data ulasan yang telah diberi label sentimen. Beberapa tahapan yang dilakukan dalam preprocessing:

3.2.1 Case Folding

Case Folding merupakan proses normalisasi teks dengan cara mengubah semua huruf menjadi huruf kecil. Tujuan dari langkah ini adalah untuk memastikan bahwa kata-kata dengan makna yang sama namun memiliki perbedaan dalam penggunaan huruf kapital tidak dianggap sebagai entitas yang berbeda. Dengan demikian, proses ini membantu menyederhanakan data serta menurunkan jumlah variasi kata yang harus dianalisis dalam pengolahan teks.

Tabel 1. Hasil Data Case Folding

Data ulasan Awal	Data Hasil Case Folding
<u>Saya</u> lagi kontrol kehamilan saya, dan nanya ke petugas di meja farmasi tapi di jawab dengan TIDAK SOPAN, DAN MUKA JUDES.	<u>saya</u> lagi kontrol kehamilan saya, dan nanya ke petugas di meja farmasi tapi di jawab dengan <u>tidak sopan, dan muka judes. ...</u>
<u>Sumpah</u> ya pelayanan nya nol banget,dari cara ngomong smpe gestur tubuh nya ga mencerminkan org melayani dgn benar! <u>Petugas</u> nya dri mulai security smpe dokter poli gada yg bener,muka judes,org ...	<u>sumpah</u> ya pelayanan nya nol banget,dari cara ngomong smpe gestur tubuh nya ga mencerminkan org melayani dgn benar! <u>petugas</u> nya dri mulai security smpe dokter poli gada yg bener,muka judes,org ...

3.2.2 Cleansing

dibawah ini merupakan hasil yang di peroleh dari data cleansing, pada tahapan ini karakter selain huruf akan dihapus termasuk tanda baca, angka dan simbol emoticon.

Tabel 2. Hasil Data Cleansing

Hasil Case Folding	Hasil Data Cleansing
sumpah ya pelayanan nya nol <u>banget</u> ,dari cara ngomong smpe gestur tubuh nya ga mencerminkan org melayani dgn <u>benar!</u> petugas nya dri mulai security smpe dokter poli gada yg bener,muka <u>judes,org ...</u>	sumpah ya pelayanan nya nol <u>banget</u> dari cara ngomong smpe gestur tubuh nya ga mencerminkan org melayani dgn <u>benar</u> petugas nya dri mulai security smpe dokter poli gada yg <u>benermuka judesorg</u>
poli gigi dan mulut pelayanannya sangat jelek dan mulai kerja mereka telat dibandingkan <u>poli2</u> lain di <u>puskesmas</u> ,poli lain mulai <u>jam 08.15</u> mereka <u>jam 08.45</u> baru manggil <u>antrian</u> , <u>tolong</u> di tertibkan lagi poli gigi dan mulutnya	poli gigi dan mulut pelayanannya sangat jelek dan mulai kerja mereka telat dibandingkan <u>poli</u> lain di <u>puskesmas</u> poli lain mulai <u>jam</u> mereka <u>jam</u> baru manggil <u>antriantolong</u> ditertibkan lagi poli gigi dan mulutnya

3.2.3 Normalization

Setelah di lakukan Cleansing kemudian akan dilakukan Normalisation Teks, pada tabel dibawa ini merupakan hasil yang di peroleh setelah tahapan Normalisation teks, pada tahap ini ulasan awal yang memiliki kata singkat akan di rubah kedalam bentuk kata normal hal ini bertujuan untuk mempermudah saat melakukan proses pelabelan dan klasifikasi data.

Tabel 3. Hasil Data Normalization

Hasil Data Cleansing	Hasil Data Normalization
sangat tidak baik kepada masyarakat pribumi indonesia seperti saya hari ini <u>tgl</u> okt mendapat perlakuan tidak baik dari dr risa yang menangani suntik vaksin	sangat tidak baik kepada masyarakat pribumi indonesia seperti saya hari ini <u>tanggal</u> okt mendapat perlakuan tidak baik dari dari risa yang menangani suntik vaksin

Hasil Data Cleansing	Hasil Data Normalization
perawat yg tidak pakai hijab sangat judes tidak ada ramahnya poli umum jawaban nya ketus tdk ada keramahan tidak cocok jadi perawat	perawat yang tidak pakai hijab sangat judes tidak ada ramahnya poli umum jawaban nya ketus tidak ada keramahan tidak cocok jadi perawat

3.2.4 Tokenization

Setelah di lakukan Normalisation Teks, data yang di dapat sebelumnya akan di lakukan tahapan Tokenization. Pada tabel dibawa ini, Hal ini bertujuan untuk memotong kata sehingga menjadi bentuk tunggal, guna memudahkan dalam mengidentifikasi suatu kalimat.

Tabel 4. Hasil Data Tokenizing

Hasil Data Normalization	Hasil Data Tokenizing
kegiatan pagi ni nunggu antrian ni d puskesmas ttp semangat walau badan udah mintak ampun	kegiatan, pagi, ni, nunggu, antrian, ni, d, puskesmas ttp, semangat, walau, badan, udah, mintak, ampun
saya lagi kontrol kehamilan saya dan nanya ke petugas di meja farmasi tapi di jawab dengan tidak sopan dan muka judes	saya, lagi, kontrol, kehamilan, saya, dan, nanya, ke, petugas, di, meja, farmasi, tapi, di, jawab, dengan, tidak, sopan, dan, muka, judes

3.2.5 Stopword Removal

selanjutnya akan di lakukan tahapan Stopword Removal yaitu untuk menghapus kata sambung yang tidak memiliki makna, contoh kata yang akan di hilangkan adalah “dan”, “yang”, “lalu”, “gak”, “aja”, “di”, dll.

Tabel 5. Hasil Data Stopword Removal

Hasil Data Tokenizing	Hasil Data Stopword Removal
sumpah, ya, pelayanan, nya, nol, banget dari, cara, ngomong, smpe, gestur, tubuh, nya, tidak, mencerminkan, org, melayani, dgn, benar, petugas, nya, dri, mulai, security, smpe, dokter, poli, gada, yang, benermuka, judes org	sumpah, pelayanan, nya, nol, banget dari, cara, ngomong, smpe, gestur, tubuh, nya, mencerminkan, org, melayani, dgn, benar, petugas, nya, dri, mulai, security, smpe, dokter, poli, gada, benermuka, judes org
sangat, tidak, baik, kepada, masyarakat, pribumi, indonesia, seperti, saya, hari, ini, tanggal, okt, mendapat, perlakuan, tidak, baik, dari, dari, risa, yang, menangani, suntik, vaksin	sangat, baik, kepada, masyarakat, pribumi, indonesia, seperti, hari, tanggal, okt, mendapat, perlakuan, baik, risa, menangani, suntik, vaksin

3.2.6 Stemming

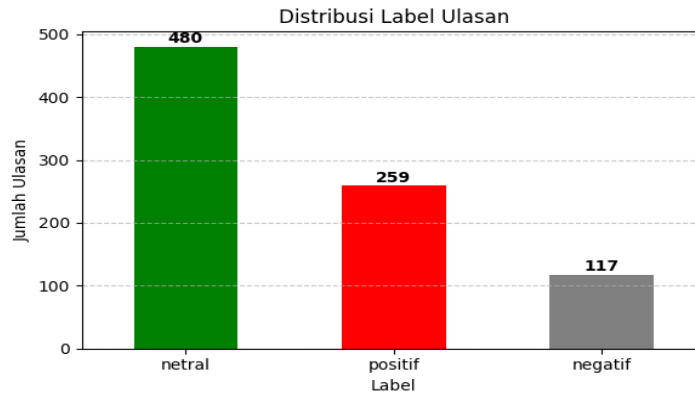
Setelah semua tahapan dilakukan kemudian akan di lakukan proses Stemming, hal ini bertujuan untuk menghilangkan kata imbuhan baik yang berada di awal maupun di akhir guna mendapatkan kata dasarnya.

Tabel 6. Hasil Data Stemming

Hasil Data Stopword Removal	Hasil Data Stemming
kontrol, kehamilan, nanya, petugas, meja, farmasi, jawab, sopan, muka, judes	kontrol, hamil, nanya, tugas, meja, farmasi, jawab, sopan, muka, judes
sumpah, pelayanan, nya, nol, banget dari, cara, ngomong, smpe, gestur, tubuh, nya, mencerminkan, org, melayani, dgn, benar, petugas, nya, dri, mulai, security, smpe, dokter, poli, gada, benermuka, judes org	sumpah, layan, nya, nol, banget dari, cara, ngomong, smpe, gestur, tubuh, nya, cermin, org, layan, dgn, benar, tugas, nya, dri, mulai, security, smpe, dokter, pol, gada, benermuka, judes org

3.3 Pelabelan Data

Setelah semua proses Pre – Processing di lakukan, maka tahap selanjutnya adalah pelabelan data, pelabelan ini di lakukan pada setiap data ulasan yang telah di peroleh sebelumnya. Hal ini di lakukan untuk mengetahui apakah ulasan tersebut dapat di katakan sentimen Positif dan Negatif.



Gambar 2. Hasil Pelabelan Data

Gambar 2 merupakan hasil persebaran sentimen yang di peroleh setelah proses Pelabelan data, setelah melalui beberapa proses maka didapatkan data yang bernilai *Netral* 480 *Positif* 259 dan *Negatif* 117.

Tabel 7. Hasil Ulasan Positif

No	Positif
1	layan cepat dokter pegawai ramah ramah senang obat sini
2	garcep buat surat badan sehat
3	layan sangat bagus sangat bagus
4	terimakasih sudah vaksin cepat ramah
5	tugas ramah layan hati
6	layan bagus ruang nya bagus
7	gedung tempat layan cukup bagus bersih sedikit kurang shg perlu perhati baik tugas maupun ujung klu perlu sediaan tunjuk maupun himbauan bersih layan cukup ramah cepat namun perlu tingkat agar
8	puskesmas harap raya mantap layan cepat tugas ramah sangat bantu dokter gigi ramah jadi anak senang cabut gigi sini
9	pertama datang puskesmas melur pegawai ramah utama bagi daftar langsung welcome layan pasien tahan tingkat kembali layan
10	obat puskesmas melur pekanbaru tempat sangat bersih ruang labornya pelayanannya sangat ramah hadap pasien

Tabel 7. menyajikan sejumlah data ulasan yang telah berhasil melalui proses pelabelan dan diklasifikasikan ke dalam kategori sentimen positif. Pelabelan dilakukan berdasarkan kandungan kata dan konteks kalimat yang menunjukkan kepuasan pelanggan terhadap layanan, fasilitas, maupun pengalaman berobat di puskesmas.

Tabel 8. Hasil Ulasan Negatif

No	Negatif
1	buruk layan enggak dokter gak satpam sama aja

No	Negatif
2	layan kurang baik apalagi tiap hari sabtu dokter datang lambat pasien sampai tunggu sampai jam lebih datang jam sampai belum tangan baru data saja
3	heran masak baru datang d depan pintu mau cabut gigi langsung d tolak kecewa ihh
4	tadi dateng puskesmas nya tutup asa mas jam operasional deh makan gaji buta kah
5	layan sangat sangat buruk kalian kerja ikhlas knpa sih gaji kalian rakyat bayar
6	tugas semua depan sama dokter judes jutek gaada ramah asa liat pasien tu kayak beban mending tutup aja nipuskesmas kasi pasien sakit malah jd tambah sakit
7	layan sangat buruk mau obat gigi lalau gak bisa alasannya ini itu liat kali dokter malas obat pasien layan tpt daftar suka mmbentak pasien
8	layan sangat buruk skringing jutek ramah kayak gak niat jadi dokter karna gaji kecil makanya jutek layan nya
9	lamban
10	layan nya jelek gak diarahin gimana prosedur nya nunggu lama oper belum ambil ini belum daftar gak nama sistem lah padahal dah antri minta ktp daftar alas jaring lelet lah tolong

Tabel 8 menampilkan data ulasan yang telah berhasil diberi label sentimen negatif. Penentuan label dilakukan berdasarkan identifikasi kata-kata yang mengindikasikan ketidakpuasan pelanggan, seperti keluhan terhadap kebersihan, kenyamanan, atau kualitas pelayanan yang dirasakan kurang memuaskan.

Selanjutnya akan divisualisasikan kedalam bentuk grafik wordcloud, tujuan penggunaan wordcloud dalam penelitian adalah untuk memvisualisasikan data teks secara efektif sehingga dapat memberikan wawasan yang lebih mendalam dan mudah di pahami.

Wordcloud sendiri memiliki fungsi dalam analisis sentimen antara lain:

- a) Identifikasi Pola Teks:
 - 1) Membantu mengidentifikasi kata-kata yang dominan atau sering muncul dalam teks positif dan negatif.
 - 2) Kata yang lebih besar menunjukkan frekuensi yang lebih tinggi, sehingga mempermudah pengenalan pola secara visual.
- b) Identifikasi Pola dan Tren:
 - 1) Membantu menggungkap kata – kata yang sering muncul dalam dataset teks, seperti tanggapan ulasan.
 - 2) Mempermudah peneliti dalam mengidentifikasi tema atau isu yang dominan dalam data.
- c) Segmentasi Berdasarkan Kategori Sentimen:
 - 1) Wordcloud dapat dibuat untuk masing-masing kategori sentimen (positif dan negatif), sehingga mempermudah perbandingan. Berikut adalah tampilan dari grafik wordcloud.

Kategori sentimen di bagi menjadi positif dan netral, yaitu:

- a) Sentimen Positif: Merupakan kata – kata atau frasa yang memiliki konotasi positif, seperti “bersih”, “ramah”, “bagus”, atau “nyaman”. Ini menunjukkan bahwa orang – orang memiliki pandangan baik tentang suatu topik atau objek pembahasan.
- b) Sentimen Negatif: Merupakan kata – kata atau frasa yang memiliki konotasi negatif, seperti “kecewa”, “kotor”, “tidak nyaman”, atau “mengecewakan”. Ini menunjukkan bahwa orang – orang memiliki pandangan yang kuat tentang suatu topik atau objek pembahasan.

3.4 Pembobotan TF – IDF

Setelah dilakukan tahap pelabelan data, maka tahap selanjutnya melakukan pembobotan TF-IDF. Dapat dilihat kata lalu, puskesmas, layan, bulan, bagus, ramah, tahun, nya, baik, sarap, 10 kata yang sering muncul atau *top word* yang di tampilkan dalam bentuk diagram.

Tabel 9. Jumlah Kata Hasil Pembobotan

Rank	Word	Score	Count
1	lalu	0.090791	137
2	layan	0.085602	544
3	bulan	0.066074	78
4	bagus	0.052583	122
5	ramah	0.048688	198
6	tahun	0.045087	56
7	baik	0.042040	156
8	sangat	0.041346	177
9	puskesmas	0.037913	214
10	buruk	0.032034	83

3.5 Klasifikasi Algoritma Naive Bayes Classifier

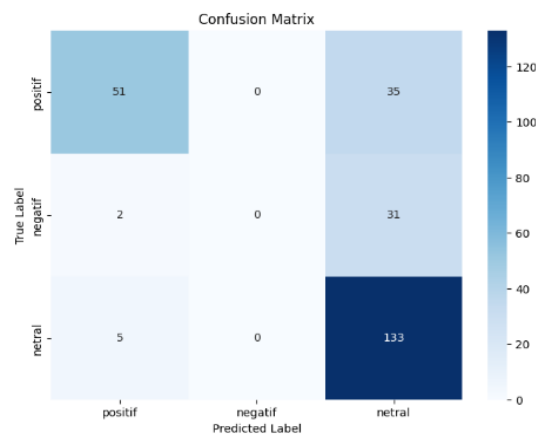
Pada tahap ini akan di lakukan klasifikasi untuk mendapatkan tingkat akurasi dari metode *Naive Bayes Classifier*. Kemudian akan dilakukan pembagian data, dari pembagian data yang dilakukan nantinya akan menghasilkan nilai akurasi. Data yang akan di klasifikasikan di bagi menjadi data *Training* dan data *Testing*, data akan di kelompokkan menjadi 70% data *Training* dan 30%

3.5.1 Pembagian 70% Data Training dan 30% Data Testing

Tabel 10. Hasil Analisis Sentimen Naive Bayes

	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>Support</i>
Positif	0.00	0.00	0.00	33
Negative	0.67	0.96	0.79	138
Netral	0.88	0.59	0.71	86
Accuracy			0.72	257
Macro avg	0.52	0.52	0.50	257
Weighted avg	0.65	0.72	0.66	257

Akurasi Model: 0.7160



Gambar 3. Hasil Confusion Matrix

<https://doi.org/xxxxxx>

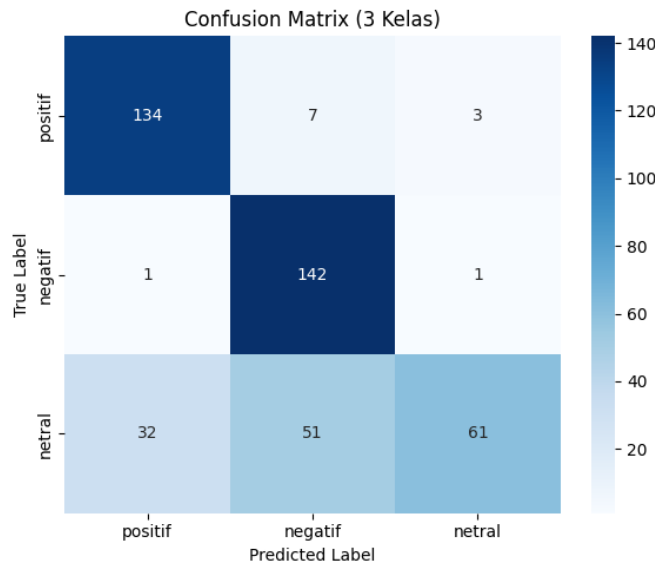
Berdasarkan Gambar 3, hasil klasifikasi menggunakan *Naïve Bayes* menunjukkan bahwa dari data berlabel *Positif*, 51 data terprediksi benar, sementara 35 data salah sebagai *Netral*. Pada data *Negatif*, tidak ada yang terklasifikasi benar, dengan 31 data salah menjadi *Netral* dan 2 data menjadi *Positif*. Untuk kelas *Netral*, model memprediksi dengan baik sebanyak 133 data, meskipun 5 data *Netral* salah terklasifikasi sebagai *Positif*. Hasil ini menunjukkan bahwa model cukup baik mendeteksi sentimen *Positif* dan *Netral*, namun masih lemah dalam mengklasifikasikan sentimen *Negatif* yang cenderung terbaca sebagai *Netral*. Dari hasil tersebut, diperoleh akurasi model sebesar 71,60%.

3.5.2 Pembagian 70% Data Training dan 30% Data Testing dengan Tambahan Smote.

Tabel 11. Hasil Analisis Sentimen Naive Bayes

	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>Support</i>
<i>Positif</i>	0.71	0.99	0.83	144
<i>Negatif</i>	0.94	0.42	0.58	144
<i>Netral</i>	0.80	0.93	0.86	144
<i>Accuracy</i>			0.78	432
<i>Macro avg</i>	0.82	0.78	0.76	432
<i>Weighted avg</i>	0.82	0.78	0.76	432

Akurasi Model: 0.78,01



Gambar 4. Hasil Confusion Matrix

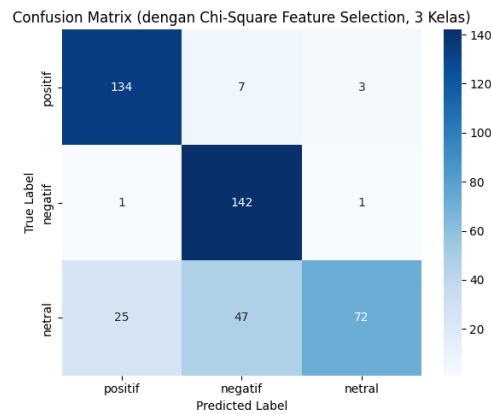
Berdasarkan Gambar 4, hasil klasifikasi menggunakan *Naïve Bayes* pada tiga kelas menunjukkan bahwa pada data positif, terdapat 134 data yang berhasil diprediksi dengan benar, sedangkan 7 data salah terklasifikasi sebagai negatif dan 3 data sebagai netral. Pada data negatif, model memprediksi dengan tepat sebanyak 142 data, sementara 1 data salah sebagai positif dan 1 data sebagai netral. Sementara itu, pada data netral, terdapat 61 data yang berhasil diprediksi dengan benar, namun masih terdapat 32 data yang salah terklasifikasi sebagai positif dan 51 data sebagai negatif. Hasil ini menunjukkan model memiliki kinerja yang baik dalam mendeteksi sentimen positif dan negatif, meskipun masih terdapat kesalahan klasifikasi pada sentimen netral yang sebagian terbaca sebagai positif dan negatif. Dari hasil tersebut, diperoleh akurasi model sebesar 78,01%.

3.5.3 Pembagian 70% Data Training dan 30% Data Testing dengan Chi Square

Tabel 12. Hasil Analisis Sentimen Naive Bayes

	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>Support</i>
Positif	0.72	0.99	0.84	144
Negatif	0.95	0.50	0.65	144
Netral	0.84	0.93	0.88	144
Accuracy			0.81	432
Macro avg	0.84	0.81	0.79	432
Weighted avg	0.84	0.81	0.79	432

Akurasi Model: 0.8056



Gambar 5. Hasil Confusion Matrix

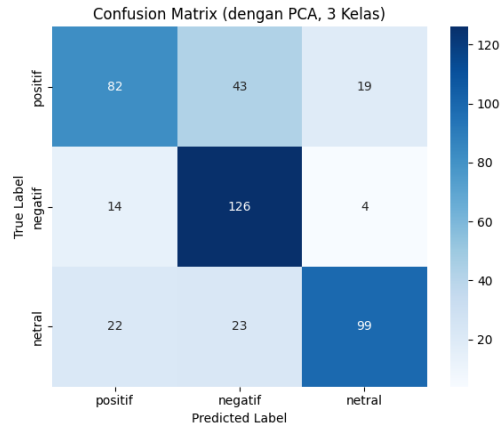
Berdasarkan confusion matrix pada gambar 5, hasil klasifikasi menggunakan Naïve Bayes dengan Chi-Square Feature Selection pada tiga kelas menunjukkan bahwa pada data positif, terdapat 134 data yang berhasil diprediksi dengan benar, sedangkan 7 data salah terklasifikasi sebagai negatif dan 3 data sebagai netral. Pada data negatif, model memprediksi dengan tepat sebanyak 142 data, dengan 1 data salah terklasifikasi sebagai positif dan 1 data sebagai netral. Sementara pada data netral, sebanyak 72 data berhasil diprediksi dengan benar, sedangkan 25 data salah terklasifikasi sebagai positif dan 47 data sebagai negatif. Hasil ini menunjukkan bahwa model memiliki kinerja baik dalam mendeteksi sentimen positif dan negatif, serta adanya peningkatan prediksi pada sentimen netral setelah menggunakan Chi-Square Feature Selection, meskipun masih terdapat sebagian data netral yang salah terklasifikasi sebagai positif atau negatif. Dari hasil tersebut, diperoleh akurasi model sebesar 80,56%

3.5.4 Pembagian 70% Data Training dan 30% Data Testing dengan Pca

Tabel 13. Hasil Analisis Sentimen memakai Pca

	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>Support</i>
Positif	0.66	0.88	0.75	144
Negatif	0.81	0.69	0.74	144
Netral	0.69	0.57	0.63	144
Accuracy			0.71	432
Macro avg	0.72	0.71	0.71	432
Weighted avg	0.72	0.71	0.71	432

Akurasi Model: 0.7106



Gambar 6. Hasil Confusion Matrix

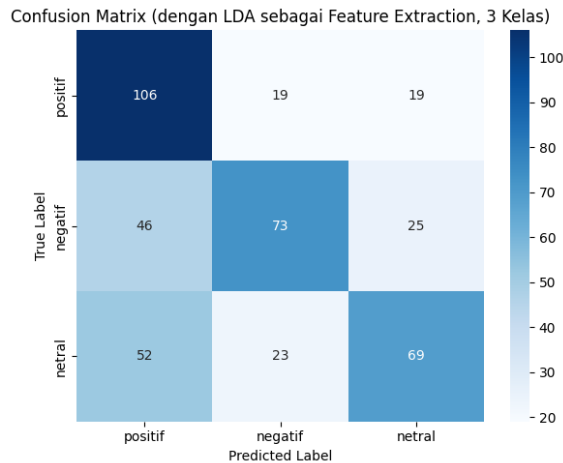
Berdasarkan confusion matrix pada gambar 6, hasil klasifikasi menggunakan Naïve Bayes dengan PCA pada tiga kelas menunjukkan bahwa pada data positif, sebanyak 82 data berhasil diprediksi dengan benar, sementara 43 data salah terklasifikasi sebagai negatif dan 19 data sebagai netral. Pada data negatif, model memprediksi dengan tepat sebanyak 126 data, sedangkan 14 data salah terklasifikasi sebagai positif dan 4 data sebagai netral. Pada data netral, terdapat 99 data yang berhasil diprediksi dengan benar, dengan 22 data salah terklasifikasi sebagai positif dan 23 data sebagai negatif. Hasil ini menunjukkan bahwa model memiliki kinerja yang cukup baik dalam mendeteksi ketiga kelas sentimen setelah menggunakan PCA, meskipun masih terdapat beberapa kesalahan klasifikasi terutama pada sentimen positif yang sebagian terdeteksi sebagai negatif. Dari hasil tersebut, diperoleh akurasi model sebesar 71,06%

3.5.5 Pembagian 70% Data Training dan 30% Data Testing dengan Lda

Tabel 14. Hasil Analisis Sentimen memakai Lda

	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>Support</i>
Positif	0.63	0.51	0.56	144
Negatif	0.61	0.48	0.54	144
Netral	0.52	0.74	0.61	144
Accuracy			0.57	432
Macro avg	0.59	0.57	0.57	432
Weighted avg	0.59	0.57	0.57	432

Akurasi Model: 0.5741



Gambar 7. Hasil Confusion Matrix

Berdasarkan Gambar Confusion Matrix, hasil klasifikasi menggunakan Naïve Bayes dengan LDA sebagai feature extraction pada tiga kelas menunjukkan bahwa pada data positif, terdapat 106 data yang berhasil diprediksi dengan benar, sedangkan 19 data salah terklasifikasi sebagai negatif dan 19 data sebagai netral. Pada data negatif, model memprediksi dengan tepat 73 data, sementara 46 data salah terklasifikasi sebagai positif dan 25 data sebagai netral. Pada data netral, terdapat 69 data yang diprediksi dengan benar, dengan 52 data salah terklasifikasi sebagai positif dan 23 data sebagai negatif.

Hasil ini menunjukkan bahwa model cukup baik dalam mengenali ketiga kelas sentimen setelah menggunakan LDA sebagai feature extraction, meskipun masih terdapat kesalahan klasifikasi terutama pada sentimen netral dan negatif yang sebagian terbaca sebagai positif. Dari hasil tersebut, diperoleh akurasi model sebesar 57,41%

3.6 Hasil Perhitungan Naive Bayes Classifier

Setelah dilakukan tahapan preprocessing seperti case folding, cleansing, normalization, tokenizing, stopword removal, dan stemming, data ulasan kemudian diberi label sentimen positif dan negatif. Selanjutnya, dilakukan proses pembobotan kata menggunakan metode TF-IDF, yang berfungsi untuk menekankan kata-kata penting dalam setiap ulasan.

Klasifikasi dilakukan dengan algoritma Naive Bayes Classifier, menggunakan lima skenario pembagian data: 70% training dan 30% testing, 70%:30% setelah smote, serta 70%:30% setelah Chi square, 70%:30% setelah Pca, 70%:30% setelah Lda. Dari hasil pengujian, model menunjukkan performa terbaik pada skenario 70% training dan 30% testing setelah di Chi square dengan nilai akurasi sebesar 80,56%. Berikut tabel hasil pembagian yang telah dilakukan:

Tabel 15 Hasil Akurasi Dari 5 Pembagian

No	Split Data	Akurasi
1	70% : 30%	71,60%
2	70% : 30% setelah di smote	78,21%
3	70% : 30% setelah di Chi square	80,56%
4	70% : 30% setelah di Pca	71,06%
5	70% : 30% setelah di Lda	57,41%

4. Kesimpulan

Analisis terhadap 856 ulasan pasien di Google Maps menunjukkan bahwa mayoritas sentimen yang muncul adalah Netral, diikuti oleh sentimen Positif, dan kemudian Negatif. Hal ini mengindikasikan bahwa layanan Puskesmas di Kota Pekanbaru masih memiliki beberapa aspek yang perlu ditingkatkan guna meningkatkan kepuasan pasien. Hasil penelitian juga menunjukkan bahwa metode Naïve Bayes Classifier mampu mencapai tingkat akurasi sebesar 80,56%. Namun, setelah diterapkannya seleksi fitur Chi-Square, model masih menunjukkan kelemahan dalam mengenali sentimen Negatif, karena banyak ulasan Negatif yang salah diklasifikasikan sebagai Netral atau Positif. Proses pelabelan dilakukan dengan mengelompokkan ulasan ke dalam tiga kategori utama, yaitu sentimen Positif, Negatif, dan Netral, untuk mempermudah proses analisis sentimen serta visualisasi data. Selain itu, visualisasi data melalui wordcloud dan pembobotan TF-IDF turut membantu dalam memahami kata-kata yang paling sering muncul pada masing-masing kategori sentimen.

Daftar Pustaka

- [1] Sri Widagdo, K. N. Qodri, F. E. N. Saputro, and N. P. A. Rizky, "Analisis sentimen terhadap pelayanan kesehatan berdasarkan ulasan Google Maps menggunakan BERT," *Jurnal*

- FASILKOM (Teknologi Informasi dan Ilmu Komputer)*, vol. 13, no. 02, 2023. <https://doi.org/10.37859/jf.v13i02.5170>
- [2] D. Sinulingga, S. Batubara, A. Sembiring, and A. Fatikhus, "Hubungan pelayanan prima terhadap kepuasan pasien rawat jalan di Puskesmas Dolok Masihul tahun 2021," *Jurnal Penelitian Farmasi & Herbal*, vol. 4, no. 1, 2021, doi: <https://doi.org/10.36656/jpsy.v4i1.771>.
- [3] D. S. Utami, A. Erfina, and M. A. Id, "Analisis sentimen ulasan terkait UNESCO Global Geopark di Google Maps dengan algoritma Naive Bayes," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, no. 2, 2022.
- [4] G. Darmawan, S. Alam, and M. I. Sulisty, "Analisis sentimen berdasarkan ulasan pengguna aplikasi MyPertamina pada Google Playstore menggunakan metode Naive Bayes," *Storage*, vol. 2, no. 3, pp. 100–108, 2023. <https://doi.org/10.55123/storage.v2i3.2333>
- [5] M. Febri Yulianto, F. Maisa Hana, and A. Prihandono, "Penerapan algoritma Naive Bayes dalam analisis sentimen terhadap mobil listrik," *Sainteks*, vol. 22, no. 1, 2025. doi: <https://doi.org/10.30595/sainteks.v22i1.26036>.
- [6] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis sentimen penggunaan aplikasi Kinemaster menggunakan metode Naive Bayes," *Jurnal Ilmiah Computer Science*, vol. 2, no. 2, pp. 89–98, 2024. doi: <https://doi.org/10.58602/jics.v2i2.24>
- [7] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis sentimen menggunakan Naive Bayes untuk melihat review masyarakat terhadap tempat wisata pantai di Kabupaten Karawang pada ulasan Google Maps," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 28–38, 2022. doi: <https://doi.org/10.35969/interkom.v16i4.192>.
- [8] P. S. Kayaningtias and P. P. Adikara, "Analisis sentimen angket kepuasan pasien Puskesmas menggunakan metode Improved K-Nearest Neighbor," *J-PTIHK (Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer)*, vol. 6, no. 3, 2022. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10717>
- [9] W. Wahyuni, "Analisis sentimen terhadap opini feminisme menggunakan metode Naive Bayes," *Jurnal Informatika Ekonomi Bisnis*, pp. 148–153, 2022, doi: <https://doi.org/10.37034/infv4i4.162>.
- [10] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis sentimen berbasis aspek pada review Female Daily menggunakan TF-IDF dan Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: <https://doi.org/10.30865/mib.v5i2.2845>.
-