

ANALISIS SENTIMEN APLIKASI MYSILOAM MENGGUNAKAN METODE NAÏVE BAYES

Alvinalia^{1*}, Abdul Rahim², ³Taghfirul Azhima Yoga Siswa

^{1,2,3}Universitas Muhammadiyah Kalimantan Timur, Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124

Received: 1 Januari 2025

Accepted: 14 Januari 2025

Published: 20 Januari 2025

Keywords:

Naive Bayes, Aplikasi Mysiloam, Processing, Analisis Sentimen.

Correspondent Email:

2111102441140@umkt.ac.id

Abstrak. Aplikasi *MySiloam* yang dikembangkan oleh *Siloam Hospitals* merupakan platform yang menyediakan layanan kesehatan online, aplikasi ini dirancang untuk memudahkan pasien dalam mengakses berbagai layanan kesehatan secara efisien dan praktis, maka dari itu penting untuk memahami persepsi pengguna melalui analisis sentimen. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi *MySiloam* dengan menggunakan metode *Naive Bayes*. Metode *Naive Bayes* dipilih karena kesederhanaannya dan kemampuannya dalam menangani data teks dengan baik. Data yang digunakan dalam penelitian ini terdiri dari ulasan pengguna yang diambil dari laman *Google Play Store* pada aplikasi *MySiloam* sebanyak 1990 ulasan melalui tahapan *Scrapping*. Proses analisis dimulai dengan tahap *processing* data sebelum dilakukan analisis. Setelah data diproses, model dilatih menggunakan teknik *TF-IDF* dan *Confusion Matriks* untuk menguji ketepatan sentimen. Hasil penelitian menunjukkan bahwa model *Naive Bayes* berhasil mencapai akurasi sebesar 89%, yang menunjukkan efektivitas metode ini dalam mengsentimen sentimen positif dan negatif dari ulasan pengguna. Dari analisis yang dilakukan, ditemukan bahwa mayoritas pengguna memberikan ulasan positif, meskipun terdapat beberapa kritik terkait performa aplikasi. Temuan ini memberikan wawasan berharga bagi pengembang aplikasi *MySiloam* untuk meningkatkan pengalaman pengguna dan memperbaiki aspek-aspek yang kurang memuaskan. Secara keseluruhan, penelitian ini tidak hanya menyoroti potensi analisis sentimen dalam konteks aplikasi kesehatan tetapi juga menunjukkan bagaimana metode *Naive Bayes* dapat digunakan secara efektif dalam pengolahan data teks.

Abstract. The *MySiloam* application developed by *Siloam Hospitals* is a platform that provides online health services. This application is designed to make it easier for patients to access various health services efficiently and practically, therefore it is important to understand user perceptions through sentiment analysis. This research aims to analyze user sentiment towards the *MySiloam* application using the *Naive Bayes* method. The *Naive Bayes* method was chosen because of its simplicity and ability to handle text data well. The data used in this research consists of user reviews taken from the *Google Play Store* page on the *MySiloam* application, totaling 1990 reviews through the *Scrapping* stage. The analysis process begins with the data processing stage before analysis is carried out. After the data is processed, the model is trained using *TF-IDF* and *Confusion Matrix* techniques to test sentiment accuracy. The research results show that the *Naive Bayes* model managed to achieve an

accuracy of 89%, which shows the effectiveness of this method in quantifying positive and negative sentiment from user reviews. From the analysis carried out, it was found that the majority of users gave positive reviews, although there were some criticisms regarding the application's performance. These findings provide valuable insights for MySiloam application developers to improve user experience and improve unsatisfactory aspects. Overall, this research not only highlights the potential of sentiment analysis in the context of healthcare applications but also shows how Naive Bayes methods can be used effectively in text data processing.

1. PENDAHULUAN

Setiap tahun masyarakat Indonesia mengalami peningkatan mutu pelayanan kesehatan sehingga fasilitas layanan kesehatan harus memenuhi standar demi mengutamakan pasien [1]. Peningkatan tersebut melibatkan kualitas, aksesibilitas dan efisiensi sistem kesehatan yang harus di iringi dengan kemajuan teknologi, adanya pengembangan aplikasi kesehatan dapat meningkatkan kesadaran dan mendorong adopsi yang lebih luas di antara masyarakat, Pemrosesan Bahasa Alami atau *Natural Language Processing* analisis sentimen menyederhanakan operasi dan melambungkan kepuasan pelanggan ke tingkat yang lebih tinggi [2]. Oleh karena itu Ilmu Komputer mengembangkan kecerdasan buatan untuk menjadi bagian integral masyarakat dalam berbagai bidang kehidupan sehari-hari dengan teknologi yang membuat komputer bisa berpikir dan belajar seperti manusia hingga dapat menggantikan beberapa tugas pekerjaan layanan [3]. Salah satu keuntungan teknologi pada bidang industri adalah layanan kesehatan, karena dapat meningkatkan efisiensi dan kualitas dalam pelayanan lewat aplikasi, dan dapat mempelajari kesehatan melalui rekomendasi langsung dari dokter dimana rumah sakit atau klinik dapat mempromosikan aplikasi kesehatan yang relevan kepada pasien mereka melalui poster atau situs web [4].

Aplikasi *MySiloam* yang ada pada laman *Google Play Store* mampu memudahkan pasien dan memungkinkan adanya transformasi untuk mengakses informasi, berkomunikasi, dan berinteraksi dengan para medis (Rosyid, 2023). Dengan berbagai manfaat yang ada, sering juga

pengguna mengalami masalah teknis saat menggunakan aplikasi *MySiloam*, oleh karena itu, penelitian ini bertujuan untuk mengetahui seberapa banyak pendapat negatif yang ada serta mengeksplorasi sentimen positif yang dapat memberikan wawasan tentang bagaimana teknologi dapat berkontribusi dalam meningkatkan layanan kesehatan dan pengalaman pasien pada Aplikasi Layanan Kesehatan *MySiloam* yang ada di *Google Play Store* [5]. Untuk mengetahui hasil penelitian, peneliti menerapkan metode sentimen *Naïve Bayes* pada analisis sentimen dengan data dalam jumlah besar [6]

Penelitian mengenai metode *Naive Bayes* telah banyak dilakukan dalam berbagai bidang, terutama dalam analisis sentimen, metode ini merupakan algoritma yang digunakan untuk sentimen data berdasarkan *Probabilitas* kelas, dan *multinomial* mengsentimenkan teks dalam dokumen [7]. Model analisis sentimen dilatih menggunakan data latih yang telah diberi label sentimen (positif atau negatif) dan kemudian diklasifikasikan menggunakan algoritma *Naïve Bayes* [8].

Melalui Analisis Sentimen, peneliti berharap agar pengembang aplikasi *Mysiloam* dapat memahami ekspektasi pengguna yang tidak hanya memenuhi kebutuhan medis, tetapi juga memberikan pengalaman pengguna yang positif dan memuaskan [9]. Sehingga mampu meningkatkan aksesibilitas, memudahkan pengelolaan kesehatan, dan menawarkan potensi penghematan biaya, dan peneliti berhasil menarik lebih banyak pengguna, sehingga perkembangan aplikasi *Mysiloam* bisa

mendorong banyak orang untuk menggunakan aplikasi *Mysiloam*.

Dalam beberapa tahun terakhir, banyak penelitian yang mencoba menganalisis sentimen atau opini dalam teks menggunakan metode pembelajaran mesin atau *Machine Learning*. Penelitian [10] analisis sentimen ulasan game online, *Naïve Bayes* menunjukkan akurasi sebesar 88,5%, lebih tinggi dibandingkan dengan algoritma lain seperti *Regresi Logistik* dan *SVM*. Dan [11] analisis sentimen opini masyarakat terhadap kepopuleran produk skincare melalui *twitter* menggunakan algoritma *Naïve Bayes*, menghasilkan nilai *Accuracy*, *Precision*, dan *Recall* diatas 80%. Lalu [12] pada penelitian yang dilakukan sebelumnya tentang proses analisis sentimen mengenai respon pengguna terhadap Shopee dengan menerapkan algoritma *Naïve Bayes*. Hasil yang didapat dari analisis sentimen menggunakan algoritma *Naïve Bayes* mendapat 78% sentimen negatif dan 22% sentimen positif dari 200 dataset. Proses pengujian algoritma *Naïve Bayes* menggunakan *confussion matrix* memperoleh nilai akurasi yaitu 77,50%. Dalam beberapa penelitian, *Naïve Bayes* telah menunjukkan akurasi yang baik dalam analisis sentimen, dengan beberapa contoh mencapai di atas 70% dan bahkan di atas 85%. Algoritma *Naive Bayes* dipilih untuk penelitian ini karena kemampuannya dalam memproses sejumlah besar data teks dengan cepat dan efisien, serta telah terbukti efektif dalam berbagai kajian analisis sentimen, termasuk di bidang kesehatan.

[13]. Dalam konteks kesehatan, *Naive Bayes* dapat digunakan untuk menganalisis sentimen pasien terhadap layanan kesehatan, mengidentifikasi topik yang paling sering dibicarakan, serta memprediksi kebutuhan pasien. Dengan demikian, *Naive Bayes* dapat berkontribusi dalam meningkatkan kualitas layanan kesehatan dan aksesibilitas informasi kesehatan dan mengembangkan sistem untuk mengkaji aplikasi mana yang memiliki sentimen atau sinisme terbaik di kalangan

masyarakat yang ditentukan oleh interaksi pengguna[14].

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis adalah penelitian terhadap sistem yang telah ada dengan tujuan untuk merancang sistem baru atau memperbaiki sistem yang telah ada. Analisis merupakan penyerapan, pengkajian serta penggunaan informasi guna membuat kesimpulan. Hal ini dapat diartikan analisis sebagai kegiatan melakukann pembahasan atas suatu objek selanjutnya diadakan pengolahan data yang kemudian dapat dibuat kesimpulan [15].

2.2 Naïve Bayes

Algoritma *Naïve Bayes* adalah metode yang menggunakan probabilitas dan statistik untuk memecahkan masalah klasifikasi, dengan menggunakan probabilitas dan statistik, Algoritma *Naive Bayes* memecahkan masalah klasifikasi dengan menghitung nilai probabilitas $P(A|B)$ berdasarkan informasi tentang probabilitas kelas X [16]. Penentuan untuk mengklasifikasikan, Kita memilih kelas untuk diklasifikasikan berdasarkan nilai maksimum dari probabilitas $P(A|B)$. dimana Probabilitas suatu dokumen B , berada di kelas A dapat dihitung menggunakan persamaan 1 [17].

$$(P(A|B) = \frac{P(B|A).P(A)}{P(B)}) \quad (1)$$

Keterangan:

$P(A|B)$ = Disebut juga posterior probability, yaitu peluang A terjadi setelah B terjadi.

$P(A \cap B)$ = Peluang B dan A terjadi bersamaan

$P(B|A)$ = Disebut juga likelihood, yaitu peluang B terjadi setelah A terjadi.

$P(A)$ = Disebut juga prior, yaitu peluang kejadian A

$P(B)$ = Peluang kejadian B

2.3 TF-IDF

Penelitian dengan data sentimen pasien terhadap ulasan pada aplikasi *Mysiloam* pada

laman *Google Play Store* dengan menggunakan algoritma *Naïve Bayes* setelah melakukan tahap *Preprocessing*, data ulasan diubah menjadi vektor fitur menggunakan teknik *TF-IDF*. Kata-kata seperti 'puas', 'sangat baik', dan 'cepat' yang sering muncul dalam ulasan positif diberikan bobot yang tinggi, sedangkan kata-kata seperti 'lama', 'sulit', dan 'kecewa' yang sering muncul dalam ulasan negatif juga diberikan bobot yang tinggi.

2.4 Evaluasi

Pada tahap terakhir dilakukan proses Analisis berdasarkan sentimen dalam dokumen. Semua tahapan ini akan menghasilkan model yang akan digunakan untuk menunjukkan akurasi dari hasil klasifikasi. Evaluasi dalam analisis sentimen digunakan untuk fokus menilai kualitas dan akurasi model klasifikasi dengan menggunakan metode *Confusion Matriks* pengukuran tingkat akurasi, *presisi*, *recall* dan *F1-Score* untuk menunjukkan hasil nilai dalam mengklasifikasi opini tersebut.

Tabel 1 *Confusion Matriks*

	Positif (1)	Negatif (0)
Positif (1)	TP	FP
Negatif (0)	FN	TN

Keterangan

TP (*True Positif*) = merupakan jumlah prediksi positif yang benar

TN (*True Negatif*) = merupakan jumlah prediksi negatif yang benar

FP (*False Positif*) = merupakan jumlah prediksi positif yang salah

FN (*False Negatif*) = merupakan jumlah prediksi negatif yang salah

a. Akurasi

Dalam analisis sentimen akurasi digunakan untuk menunjukkan seberapa baik model dapat memprediksi sentimen opini dengan benar hingga dapat dinyatakan sebagai presentase prediksi [18]. Dengan menggunakan persamaan matematika rumus akurasi yaitu:

$$(Accuracy = \frac{TP+TN}{TP+TN+FP+FN}) \quad (2)$$

b. Presisi

Presisi mengukur seberapa baik sistem dapat mengidentifikasi data yang sebenarnya positif dan mengklasifikasikannya dengan benar. Rumus presisi dalam analisis sentimen adalah:

$$(Presisi = \frac{TP}{TP+FP}) \quad (3)$$

c. Recall

Recall mengukur kemampuan sistem untuk menemukan semua dokumen yang relevan. Jika sistem memberikan 100 temuan dan di antaranya ada 50 dokumen yang relevan, maka nilai *recall*-nya adalah 50% (Romadloni et al., 2022).

$$(Recall = \frac{TP}{TP+FN}) \quad (4)$$

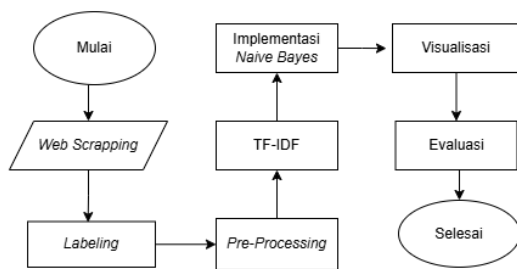
d. F1-Score

Digunakan untuk mengukur kinerja sistem secara lebih komprehensif dengan menggabungkan *Recall* dan *Precision*. Nilai *F1 Score* yang tinggi menunjukkan bahwa sistem memiliki kinerja yang baik dalam mengklasifikasikan dokumen. Rumus *F1-Score* dapat dilihat menggunakan rumus.

$$(F1\ Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}) \quad (5)$$

3. METODE PENELITIAN

Penelitian Analisis Sentimen Aplikasi *Mysiloam* menggunakan metode *Naïve Bayes* umumnya melibatkan beberapa tahapan yang sistematis, mulai dari penentuan objek penelitian, *Preprocessing*, hingga evaluasi model. Tahapan ini bertujuan untuk memastikan akurasi dan efektivitas sistem klasifikasi sentimen. Adapun metod penelitian yang digunakan dapat dilihat pada Gambar 1.



Gambar 1 Metode Penelitian

3.1 Web Scrapping

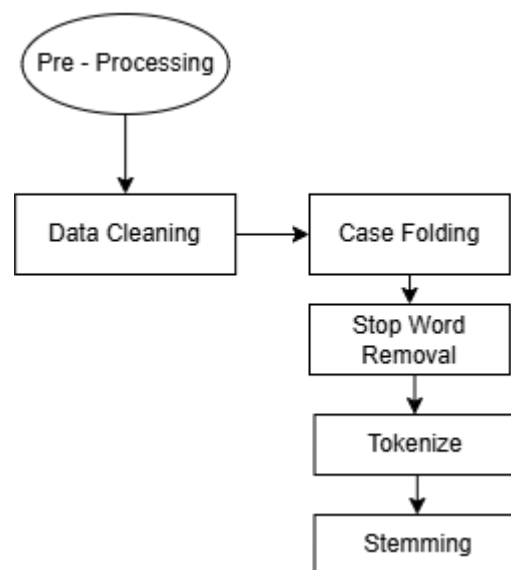
Data dikumpulkan melalui teknik *Web Scrapping* untuk mengambil ulasan pengguna dari *Google Play Store*, dengan fokus pada aplikasi kesehatan *Mysiloam* dan mengelola data 1990 ulasan dalam rentang waktu sejak awal Agustus 2024 sampai awal September 2024. Dengan fitur *Google Play Scraper* atau alat yang digunakan untuk *web scrapping* dalam bahasa pemrograman *Python* secara otomatis data diambil dari situs *web* menggunakan bahasa Indonesia.

3.2 Labeling

Dengan memberikan *Label* atau tanda pada setiap data dapat menentukan kelas pada ulasan konteks analisis sentimen, *Label* dapat berupa kategori, kelas, atau atribut tertentu yang ingin di identifikasikan dan proses pemberian tag atau label pada data teks bisa berupa positif, atau negatif. Tahapan *Labeling* juga penting karena memengaruhi kualitas model pada kualitas data latih yang digunakan.

3.3 Preprocessing Data

Pada tahapan ini *Preprocessing* dalam konteks *Data Science* dan *Data Mining* adalah proses penting yang dilakukan untuk mempersiapkan data agar lebih siap untuk analisis. Tujuan utama dari *Preprocessing* adalah untuk membersihkan dan menyederhanakan data teks sehingga model *machine learning* dapat memproses data tersebut secara efektif. Adapun Tahapan *Preprocessing* Data dapat dilihat Pada Gambar



Gambar 2 Preprocessing Data

3.4 TF-IDF

Penelitian dengan data sentimen pasien terhadap ulasan pada aplikasi *Mysiloam* pada laman *Google Play Store* dengan menggunakan algoritma *Naive Bayes* setelah melakukan tahap *Preprocessing*, data ulasan diubah menjadi vektor fitur menggunakan teknik *TF-IDF*. Kata-kata seperti 'puas', 'sangat baik', dan 'cepat' yang sering muncul dalam ulasan positif diberikan bobot yang tinggi, sedangkan kata-kata seperti 'lama', 'sulit', dan 'kecewa' yang sering muncul dalam ulasan negatif juga diberikan bobot yang tinggi.

3.5 Naive Bayes

Setelah data dikumpulkan dan dibersihkan, proses klasifikasi menggunakan algoritma *Naive Bayes* dilakukan. *Data Training* digunakan untuk melatih model *Naive Bayes* untuk mempelajari pola hubungan antara fitur dan kelas, dan data *Testing* digunakan untuk menguji model setelah dilatih. Model yang telah dilatih digunakan untuk memprediksi kelas berdasarkan pola yang telah dipelajari dari data *Training*. Akurasi model kemudian dievaluasi untuk mengukur kinerja klasifikasi.

3.6 Visualisasi

Visualisasi merupakan alat yang efektif dalam analisis sentimen menggunakan algoritma *Naive Bayes*. Dengan visualisasi dapat melihat distribusi sentimen dalam dataset serta mengidentifikasi kata-kata kunci yang mendominasi sentimen positif dan negatif. *WordCloud*, yang dapat dihasilkan menggunakan *library Python* seperti *WordCloud*, adalah salah satu teknik visualisasi yang populer untuk menampilkan frekuensi kata dalam teks [19].

3.7 Evaluasi

Pada tahap terakhir dilakukan proses klasifikasi berdasarkan sentimen dalam dokumen. Semua tahapan ini akan menghasilkan model yang akan digunakan untuk menunjukkan akurasi dari hasil klasifikasi. Evaluasi dalam analisis sentimen digunakan untuk fokus menilai kualitas dan akurasi model klasifikasi dengan menggunakan metode *Confusion Matriks* pengukuran tingkat akurasi, *Presisi*, *Recall* untuk menunjukkan hasil nilai dalam mengklasifikasi opini tersebut.

4. HASIL DAN PEMBAHASAN

4.1 Web Scrapping

Hasil pengumpulan data ulasan pada aplikasi *Mysiloam* menggunakan metode *Web Scrapping* sebagai tahap awal dalam penelitian ini berhasil dikumpulkan sebanyak 1990 ulasan. Data tersebut kemudian diolah untuk menghilangkan *noise* dan data yang tidak relevan. Adapun hasil pengumpulan data dapat dilihat pada Tabel 2.

Tabel 2 Data *Mysiloam*

UserName	Score	Content
Om Jony anwar effendi	1	Dibutuhkan
	2	sebahagian besar masyarakat tdk tahu bhs inggris, malah aplikasinya

Farida	4	pakai bhs Asing. ...buatlah dgn bhs indonesia
Sutias		cukup puas dg pelayanan
tini		registrasi opd
bayu		Baik
clausa	5	pelayanannya
edi		
kristianto	5	Aplikasi yang Baik dan Cepat

4.2 Labeling

Setelah proses pengumpulan data melalui teknik *scraping*, tahap selanjutnya adalah melakukan proses *labeling* pada data teks yang telah diperoleh dengan mengidentifikasikan sentimen menjadi positif, negatif atau netral, kemudian bagi data yang telah diberi label dibagi menjadi data latih dan data uji. Pembagian data merupakan langkah krusial dalam pengembangan model analisis sentimen. Data latih berfungsi membentuk model yang mampu mengidentifikasi sentimen pada data teks baru dan data uji digunakan untuk mengukur seberapa akurat model dalam memprediksi sentimen pada data yang baru. Adapun data yang telah dilabel dapat dilihat pada Tabel 3.

Tabel 3 Data Labeling

Content	Label
Dibutuhkan	Negatif
sebahagian besar masyarakat tdk tahu bhs inggris, malah aplikasinya pakai bhs Asing. ...buatlah dgn bhs indonesia	Negatif
cukup puas dg pelayanan registrasi opd	Positif
Baik pelayanannya	Positif

Aplikasi yang Baik dan Cepat

Positif

4.3 Preprocessing Data

Tahap *Processing* dalam analisis sentimen yang bertujuan untuk mempersiapkan data teks agar dapat diproses oleh algoritma *machine learning*, proses *Preprocessing* yang meliputi *tokenize*, *stopword*, dan *stemming* menunjukkan bahwa terdapat 1990 kata dan kalimat unik dalam dataset. Kata-kata ini diubah dengan tahapan *Preprocessing*.

a. Data Cleaning

Tahapan *cleaning* pada analisis sentimen sangat penting karena memastikan data yang digunakan dalam analisis adalah akurat, data *cleaning* mengidentifikasi data yang hilang, duplikat, data yang tidak konsisten dan tidak relevan. Hasil dari proses *Cleaning* ini menunjukkan data dapat diolah ke tahapan selanjutnya. Adapun data yang telah dicleaning dapat dilihat pada Tabel 4.

Tabel 4 Data Cleaning

Content	Score	Label	Text Clean
Baik pelayanannya	5	Positif	baik pelayanannya
Aplikasi yang Baik dan Cepat	5	Positif	aplikasi yang baik dan cepat
ter the best	4	Positif	ter the best
Dibutuhkan	1	Negatif	dibutuhkan
Mau berobat dokter umum gak bisa online	1	Negatif	mau berobat dokter umum gak bisa online

b. Case Folding

Proses dari *Case Folding* bertujuan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil untuk memastikan bahwa

perbedaan kapitalisasi tidak memengaruhi analisis, sehingga meningkatkan konsistensi data yang akan diproses. Adapun data yang telah di *Case Folding* dapat dilihat pada Tabel 5.

Tabel 5 Case Folding

Content	Score	Label	Text Clean
Baik pelayanannya	5	Positif	baik pelayanannya
Aplikasi yang Baik dan Cepat	5	Positif	aplikasi yang baik dan cepat
ter the best	4	Positif	ter the best
Dibutuhkan	1	Negatif	dibutuhkan
Mau berobat dokter umum gak bisa online	1	Negatif	mau berobat dokter umum gak bisa online

c. StopWord

Untuk menghapus kata-kata yang dianggap tidak memiliki makna signifikan dalam konteks analisis. Kata-kata ini biasanya adalah kata-kata umum seperti "dan", "yang", "di", yang sering muncul dalam teks tetapi tidak memberikan informasi penting tentang sentimen yang terkandung di dalamnya. Adapun data yang telah di *StopWord* dapat dilihat pada Tabel 6.

Tabel 6 StopWord

Content	Score	Label	Text Clean	Text StopWord
Baik pelayanannya	5	Positif	baik pelayanannya	pelayanannya
Aplikasi yang Baik	5	Positif	aplikasi yang baik dan cepat	aplikasi cepat

dan				
Cepat				
ter the			ter the	ter the best
best	4	Positif	best	
Dibutu		Negati	dibutuhka	dibutuhkan
hkan	1	f	n	
Mau				berobat
beroba				dokter gak
t			mau	onlinepas
dokter			berobat	kerumah
umum			dokter	sakitnya
gak		Negati	umum	dokter
bisa		f	gak bisa	bukanya
online	1		online	sore jam

d. Tokenize

Proses *Tokenize* dalam konteks pemrosesan bahasa alami (NLP) dan analisis sentimen adalah proses penting karena melibatkan pemecahan teks menjadi unit-unit yang lebih kecil yang disebut token, yang biasanya berupa kata atau frasa. Adapun data yang telah di*Tokenize* dapat dilihat pada Tabel 7.

Gambar 2 Data *Tokenize*

Content	Score	Label	Text CaseFolding	Text StopWord	Text Tokens
Baik pelayanannya	5	Positif	baik pelayanannya		['pelayanannya']
Aplikasi yang Baik dan Cepat	5	Positif	aplikasi yang baik dan cepat		['aplikasi', 'cepat']
ter the best	4	Positif	ter the best		['ter', 'the', 'best']
Dibutuhkan	1	Negatif	dibutuhkan		['dibutuhkan']
Mau berobat dokter umum gak bisa online	1	Negatif	mau berobat dokter umum gak bisa		['berobat', 'dokter', 'gak', 'onlinepas', 'kerumah', 'sakitnya', 'dokter', 'bukanya', 'sore', 'jam']

e. Stemming

Dalam analisis sentimen *Stemming* bertujuan untuk mengubah kata-kata berimbuhan menjadi bentuk dasar atau akar kata karena dapat mengklasifikasikan teks dengan teknik yang digunakan baik itu awalan (prefiks), akhiran (sufiks), maupun sisipan (infix). Adapun data yang telah di*Stemming* dapat dilihat pada Gambar 3.

Content	S c o r e	Label	Text Case Folding	Text StopWord	Text Tokens	Text Stemming
Baik pelayanannya	5	Positif	baik pelayanannya		['pelayanannya']	layan
Aplikasi yang Baik dan Cepat	5	Positif	aplikasi yang baik dan cepat		['aplikasi', 'cepat']	aplikasi cepat
ter the best	4	Positif	ter the best		['ter', 'the', 'best']	ter the best
Dibutuhkan	1	Negatif	dibutuhkan		['dibutuhkan']	butuh
Mau berobat dokter umum gak bisa online	1	Negatif	mau berobat dokter umum gak bisa		['berobat', 'dokter', 'gak', 'onlinepas', 'kerumah', 'sakitnya']...	obat dokter gak onlinepas rumah sakit...

Gambar 3 Data *Stemming*

4.4 TF-IDF

Setelah melakukan preprocessing data berhasil dikumpulkan sebanyak 1839 data untuk dilakukan proses klasifikasi Tahapan *TF-IDF* (*Term Frequency-Inverse Document Frequency*) dalam pembobotan kata, data dibagi menjadi dua set, yaitu data pelatihan (training) dan data pengujian (testing). Pembagian ini biasanya dilakukan dengan proporsi tertentu, peneliti melakukan pembagian data dengan uji coba *data training* 80% untuk *data testing* 20%, dengan menggunakan modul *sklearn.model_selection* sehingga jumlah data training didapatkan sebesar 1474 data dan data testing sebesar 369 data. Adapun rasio pembagian data dapat dilihat pada Tabel 8.

Tabel 7 Pembagian Data

Rasio	Data latih	Data Uji
80:20	1474	369

Setelah menghitung TF dan IDF untuk setiap kata dalam setiap dokumen, kita akan mendapatkan matriks *TF-IDF* yang berisi skor pentingnya setiap kata dalam setiap dokumen. *TF-IDF* adalah teknik yang sangat berguna dalam pemrosesan bahasa alami. Adapun *TF-IDF* dapat dilihat pada Gambar 4.

Kata-kata yang sering muncul:

layan: 2.62
aplikasi: 2.87
bantu: 2.91
bagus: 3.11
dokter: 3.11
mudah: 3.19
siloam: 3.2
nya: 3.49
pasien: 3.71
yg: 3.73

Kata-kata yang jarang muncul:

```
yet: 7.6
ygmuda: 7.6
yi: 7.6
yo: 7.6
you: 7.6
your: 7.6
ypho: 7.6
zoha: 7.6
zurich: 7.6
zv: 7.6
```

Gambar 4 TF-IDF

4.5 Naive Bayes

Model atau metode *Naive Bayes* dilatih dengan menggunakan *dataset* yang berlabel positif dan negatif, setelah *preprocessing* (*cleaning, case folding, tokenisasi, stop word removal, stemming*), data diubah menjadi vektor TF-IDF. Model *Naive Bayes*, yang dilatih dengan *dataset* berlabel, berhasil mencapai akurasi 86%. Hasil ini menunjukkan kemampuan model dalam memprediksi kelas yang benar pada data baru, ini menandakan bahwa model tersebut mampu melakukan prediksi yang tepat dalam konteks analisis yang dilakukan, seperti analisis sentimen atau klasifikasi teks lainnya. Adapun hasil *Naive Bayes* dapat dilihat pada Gambar 4.

```
'Positif', 'Positif', 'Positif', 'Positif', 'Positif', 'Positif',
'Positif', 'Positif', 'Positif', 'Positif', 'Positif', 'Positif',
'Positif', 'Positif', 'Positif', 'Positif', 'Positif', 'Positif',
'Positif', 'Positif', 'Positif', 'Positif', 'Positif', 'Positif',
'Positif', 'Positif', 'Negatif'], dtype='<U7')
```

Gambar 4 Data Sentimen *Naive Bayes*

4.6 Visualisasi

Algoritma *Naive Bayes* dapat mengklasifikasikan data teks berdasarkan sentimen. Hasil klasifikasi ini dapat divisualisasikan menggunakan *WordCloud* untuk menampilkan kata-kata yang paling

sering muncul pada masing-masing sentimen. Kata-kata yang lebih sering muncul akan ditampilkan dalam ukuran font yang lebih besar, sehingga dapat dibedakan dengan mudah mengidentifikasi kata-kata kunci yang dominan. Adapun *WordCloud* dari Aplikasi *Mysiloam* dapat dilihat pada Gambar 5,



Gambar 5 *WordCloud* Aplikasi *Mysiloam*

WordCloud positif sering digunakan untuk menyoroti ulasan yang dihargai oleh pengguna. Adapun *WordCloud* dapat positif dapat dilihat pada Gambar 6



Gambar 6 *WordCloud* Kata Positif

WordCloud negatif berfungsi untuk mengidentifikasi masalah atau keluhan umum dari pengguna. Adapun *Wordcloud* kata negati dapat dilihat pada Gambar 7.



Gambar 7 *WordCloud* Kata Negatif

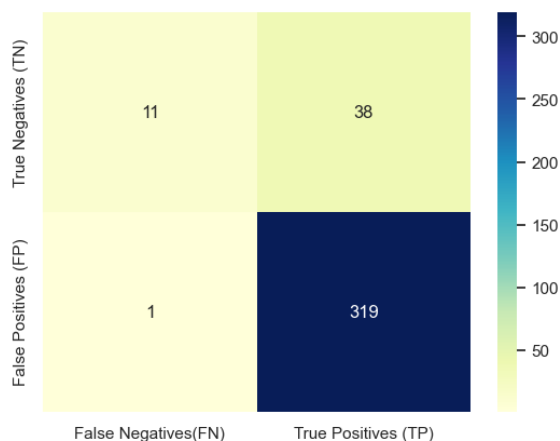
4.7 Evaluasi

Evaluasi model merupakan tahap akhir dalam proses klasifikasi yang dilakukan dengan tujuan untuk menguji model yang telah dilatih dan diuji. Adapun hasil evaluasi dapat dilihat pada Gambar 8.

	precision	recall	f1-score	support
Negatif	0.92	0.22	0.36	49
Positif	0.89	1.00	0.94	320
accuracy			0.89	369
macro avg	0.91	0.61	0.65	369
weighted avg	0.90	0.89	0.87	369

Gambar 8 Evaluasi

Adapun hasil *Confusion Matrix* dapat dilihat pada Gambar 9.



Gambar 9 *Confusion Matrix*

5. KESIMPULAN

Berdasarkan hasil pengujian dan pembahasan yang dilakukan, bahwa diperoleh kesimpulan algoritma untuk Analisis sentimen *Naive Bayes* dapat digunakan untuk melakukan analisis sentimen terhadap ulasan aplikasi *Mysiloam*,

1. pengguna secara umum memberikan penilaian positif terhadap fitur-fitur dasar aplikasi seperti pendaftaran dan pemesanan. Hal ini dapat ditunjukkan dari ulasan sentimen yang mana nilai positif sebesar 320 sentimen dan negatif 49 sentimen negatif
2. Nilai akurasi tinggi pada penelitian ini adalah dengan ulasan yang dikumpulkan melalui teknik *web scrapping* dengan jumlah 1990 data, diberi label sebelum dilakukan tahapan yang meliputi

preprocessing, pembagaaian data dan klasifikasi kata menggunakan 1474 untuk melatih dan 369 untuk data model, lalu di implementasikan mampu mengklasifikasikan ulasan pengguna aplikasi *Mysiloam* dengan akurasi sebesar 89%. Dengan menunjukkan model yang dibangun memiliki akurasi yang tinggi dan sangat baik dalam mengidentifikasi baik ulasan positif maupun negatif. Hasil ini mengindikasikan bahwa aplikasi *Mysiloam* secara umum diterima dengan baik oleh pengguna.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Ardiansyah, Adika Sri Widagdo, Krisna Nuresa Qodri, F. E. N. Saputro, and Nisrina Akbar Rizky P, "Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT," *J. Fasilkom*, vol. 13, no. 02, pp. 326–333, 2023, doi: 10.37859/jf.v13i02.5170.
- [2] T. Tinaliah and T. Elizabeth, "Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.
- [3] A. Zein, "Kecerdasan Buatan Dalam Hal Otomatisasi Layanan," *J. Ilmu Komput. JIK*, vol. 4, no. 2, p. 18, 2021, [Online]. Available: <https://jurnal.pranataindonesia.ac.id/index.php/jik/article/download/96/49>
- [4] G. G. Sari and W. Wirman, "Telemedicine sebagai Media Konsultasi Kesehatan di Masa Pandemic COVID 19 di Indonesia," *J. Komun.*, vol. 15, no. 1, pp. 43–54, 2021, doi: 10.21107/ilkom.v15i1.10181.
- [5] R. Riskawati, F. Fatihanursari, I. Iin, and A. Rizki Rinaldi, "Penerapan Metode Naïve Bayes Classifier Pada Analisis Sentimen Aplikasi Gopay," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 346–353, 2024, doi: 10.36040/jati.v8i1.8699.
- [6] A. A. Darmawan, T. Informatika, F. I. Komputer, and U. L. Kuning, "Analisis Sentimen Terhadap Rumah Sakit Di Pekanbaru Menggunakan Metode Naïve Bayes," 2022.

- [7] F. Hadaina and U. Budiyo, "Implementasi Metode Multinomial Naïve Bayes Untuk Sentiment Analysis Terhadap Data Ulasan Produk Colearn Pada Google Play Store," *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 660–666, 2022, [Online]. Available: <https://senafiti.budiluhur.ac.id/index.php>
- [8] S. D. Mendoza *et al.*, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," *Nat. Microbiol.*, vol. 3, no. 1, p. 641, 2020, [Online]. Available: <http://dx.doi.org/10.1038/s41421-020-0164-0>
<https://doi.org/10.1016/j.solener.2019.02.027>
<https://www.golder.com/insights/block-caving-a-viable-alternative/>
<http://dx.doi.org/10.1038/s41467-020-15507-2>
<http://dx.doi.org/10.1038/s41587-020-05>
- [9] J. D. C. Aruan, B. Rahayudi, and A. Ridok, "Analisis Sentimen Opini Masyarakat terhadap Pelayanan Rumah Sakit Umum Daerah menggunakan Metode Support Vector Machine dan TermFrequency – Inverse Document Frequency," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 5, pp. 2072–2078, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] M. Safrudin, M. Martanto, and U. Hayati, "Perbandingan Kinerja Naïve Bayes Dan Support Vector Machine Untuk Klasifikasi Sentimen Ulasan Game Genshin Impact," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 3182–3188, 2024, doi: 10.36040/jati.v8i3.8415.
- [11] A. F. Setyaningsih, D. Septiyani, and S. R. Widiasari, "Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 224–235, 2023, doi: 10.37012/jtik.v9i1.1409.
- [12] H. Z. Muflih, H. D. Al Assyam, F. A. Pangestu, and M. Kamayani, "Analisis Sentimen Terhadap Ulasan Penggunaan Shopee Melalui Tweet Pada Twitter Menggunakan Algoritma Naïve Bayes," *J. Tek. Inform. dan Komput.*, vol. 2, no. 2, pp. 67–74, 2023, doi: 10.22236/jutikom.v2i2.12199.
- [14] N. R. Wardani and A. Erfina, "Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi Dokter Menggunakan Algoritma Naive Bayes," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform. Univ. Nusa Putra)*, pp. 11–18, 2021.
- [15] N. Pramudibyo, "Analisis Kepuasan Pengguna Terhadap Aplikasi Redbus Dengan Menggunakan Metode End User Computing Satisfaction (Eucs)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1099–1107, 2024, doi: 10.23960/jitet.v12i2.4133.
- [17] R. Indransyah, Y. H. Chrisnanto, P. N. Sabrina, and S. Kom, "Klasifikasi Sentimen Pergelaran MotoGP di Indonesia Menggunakan Algoritma Correlated Naive Bayes Classifier," *INFOTECH J.*, vol. 8, no. 2, pp. 60–66, 2022, [Online]. Available: <https://doi.org/10.31949/infotech.v8i2.3103>
- [18] G. Darmawan, S. Alam, and M. I. Sulisty, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes," *STORAGE – J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 100–108, 2023.
- [19] J. P. Setiadi and S. Sugiyamta, "Analisis dan Visualisasi Berbasis Web Sentimen Pengguna Jenius Menggunakan Naïve Bayes Classifier," *J. Teknol. Sist. Inf. dan Apl.*, vol. 7, no. 1, pp. 245–254, 2024, doi: 10.32493/jtsi.v7i1.37981.