

Analisis Sentimen Terhadap Aplikasi KitaLulus Menggunakan Metode Naive Bayes dari Ulasan Google Play Store

Nadia Amalia Putri^{*1}, Agustina Srirahayu², Nugroho Arif Sudibyo³

^{1,2,3}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

Email: ^{*1}202020423@mhs.udb.ac.id, ²agustina@udb.ac.id, ³nugroho_arif@udb.ac.id

(Naskah masuk: 24 Juli 2024, diterima untuk diterbitkan: 5 Maret 2025)

Abstrak: Aplikasi pencarian pekerjaan seperti KitaLulus menjadi penting untuk membantu lulusan menemukan pekerjaan sesuai dengan keahlian dan minat dan bakat mereka. Analisis sentimen diperlukan untuk memahami pandangan pengguna terhadap aplikasi KitaLulus. Metode Naive Bayes digunakan dalam analisis ini karena efisiensinya dan akurasi tinggi dalam klasifikasi sentimen. Penelitian ini memanfaatkan 597 data ulasan yang diambil dari ulasan pengguna aplikasi didalam google play store dengan metode web scraping menggunakan library google-play-scraper. Pengolahan analisis sentimen pada dataset diberi label positif, netral, dan negatif. Data diolah menggunakan metode naive bayes berhasil mencapai tingkat akurasi sebesar 91%. Hasil evaluasi menunjukkan bahwa presisi, recall, dan f1-score untuk sentimen positif berturut-turut adalah 0.99, 0.94, dan 0.97. Di sisi lain, kinerja model lebih rendah pada sentimen negatif dan netral. Hasil analisis sentimen positif menunjukkan bahwa aplikasi KitaLulus telah diterima dan mendapatkan ulasan baik oleh para penggunanya. Penelitian ini bertujuan untuk meningkatkan pemahaman terhadap sentimen pengguna terhadap aplikasi KitaLulus dan memberikan kontribusi berharga bagi pengembang dalam usaha meningkatkan kualitas aplikasi tersebut.

Kata Kunci – Analisis Sentimen; Naive Bayes; KitaLulus

Sentiment Analysis of the KitaLulus Application Using the Naive Bayes Method from Google Play Store Reviews

Abstract: Job search applications like KitaLulus are essential to help graduates find jobs that match their skills, interests, and talents. Sentiment analysis is necessary to understand users' views on the KitaLulus application. The Naive Bayes method is used in this analysis due to its efficiency and high accuracy in sentiment classification. This study utilizes 597 review data collected from user reviews of the application on the Google Play Store through web scraping using the google-play-scraper library. Sentiment analysis processing on the dataset is labeled as positive, neutral, and negative. Data processed using the Naive Bayes method achieved an accuracy rate of 91%. Evaluation results show that the precision, recall, and F1-score for positive sentiment are 0.99, 0.94, and 0.97, respectively. On the other hand, the model's performance is lower for negative and neutral sentiments. The results of positive sentiment analysis indicate that the KitaLulus application has been well-received and has received good reviews from its users. This study aims to enhance understanding of user sentiment towards the KitaLulus application and provide valuable contributions to developers in their efforts to improve the quality of the application.

Keywords – Sentiment Analysis; Naive Bayes; KitaLulus

1. PENDAHULUAN

Pada era teknologi digital yang terus maju, aplikasi seluler sudah menjadi komponen penting dari keberadaan tiap-tiap hari. Di antara kategori aplikasi yang semakin disukai adalah yang didedikasikan untuk mencari pekerjaan dan pembuatan CV. Mengingat meningkatnya daya saing di pasar kerja, aplikasi semacam itu memberikan jalan baru dan efektif bagi individu yang mencari pekerjaan untuk mengidentifikasi lowongan pekerjaan yang selaras dengan kemampuan dan preferensi mereka.

Sayangnya, belakangan ini, banyak lulusan menghadapi tantangan dalam mengidentifikasi peluang kerja yang relevan, yang mengarah ke situasi di mana sebagian besar lulusan memilih

untuk beralih jalur karir dalam mengejar lowongan pekerjaan, memprioritaskan mengamankan pekerjaan daripada menyelaraskan dengan keahlian mereka. Akibatnya, tren ini menghasilkan skenario di mana keterampilan yang diperoleh selama pendidikan mereka tetap kurang dimanfaatkan dan gagal mencapai potensi penuh mereka. Masalah lowongan pekerjaan ini menimbulkan tantangan yang signifikan bagi individu dari berbagai latar belakang [1]. Pentingnya dilakukan cara yang efektif untuk mendukung lulusan baru dalam mengamankan peluang kerja yang selaras dengan keterampilan dan hasrat mereka. Oleh karena itu, ada kebutuhan mendesak untuk langkah-langkah proaktif untuk mengatasi ketidakcocokan antara pencari kerja dan posisi yang tersedia di dunia tenaga kerja.

Aplikasi KitaLulus dirancang untuk membantu individu di pasar kerja dalam mengidentifikasi peluang kerja yang selaras dengan keterampilan dan minat unik mereka. Selain itu, aplikasi ini tidak hanya memfasilitasi eksplorasi lowongan pekerjaan tetapi juga memungkinkan pengguna untuk membuat dan menyesuaikan daftar riwayat hidup mereka melalui platform online. Berfungsi sebagai perantara penting antara pencari kerja dan calon pemberi kerja, sangat penting untuk memahami umpan balik dan keterlibatan pengguna dengan aplikasi ini. Menyadari pentingnya respons dan interaksi pengguna dapat sangat berkontribusi untuk meningkatkan efektivitas dan pengalaman pengguna platform tersebut.

Dalam konteks khusus ini, analisis sentimen muncul sebagai alat penting untuk menjelaskan perspektif, emosi, dan pertemuan pengguna dalam kaitannya dengan aplikasi KitaLulus. Analisis sentimen diyakini sebagai teknik menguasai dan memproses dataset dengan langsung untuk memperoleh informasi. Analisis sentimen digunakan sebagai pendeteksi opini tentang topik dan suatu objek (seperti individu, organisasi, atau produk) dalam suatu kumpulan data [2]. Analisis sentimen yang akurat dapat memberikan wawasan mendalam mengenai kepuasan pengguna, memungkinkan pengembang mengambil tindakan pasti untuk membuat kualitas maupun daya saing aplikasi meningkat, serta merespons permintaan pasar yang dinamis [3].

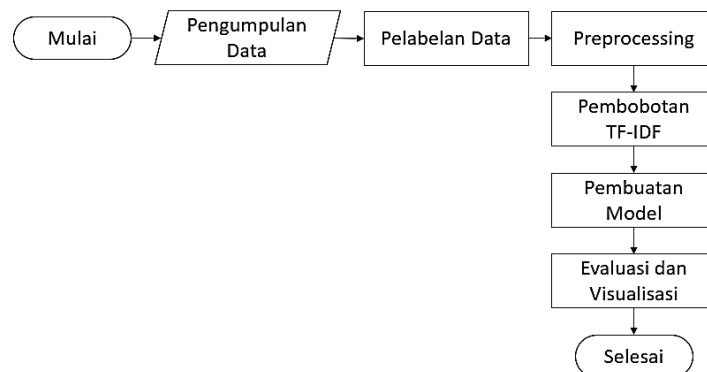
Metode *Naive Bayes* yang diterapkan pada analisis sentimen untuk pengklasifikasian sebuah ulasan pada aplikasi masuk pada golongan sentiment seperti positif, negatif, atau netral. Metode klasifikasi *Naive Bayes* memiliki arti sebagai metode yang memanfaatkan statistik sederhana berdasarkan *teorema Bayes*, yang mengasumsikan ada tidaknya kelas dengan karakteristik lain [4]. Saat melakukan analisis sentimen, banyak orang menggunakan metode *Naive Bayes* sebagai metodenya. Klasifikasi *Naive Bayes* menduga peluang keanggotaan suatu kelas beralaskan asumsi prediksi independen. Oleh karena itu, peluang waktu yang akan datang dapat ditentukan menurut pengalaman pada awalnya. Kelebihan metode *Naive Bayes* adalah efisien karena dapat mempersingkat proses analisis sentimen. Selain itu, metode *Naive Bayes* seringkali memiliki akurasi yang tinggi meskipun data pelatihannya sedikit [5].

Metode *Naive Bayes* terkenal dengan kemampuannya mencapai hasil akurasi yang tinggi. Pada kajian yang diteliti oleh Fikri dkk pada tahun 2020 yang bertajuk “Perbandingan Metode *Naive Bayes* dan *Support Vector Machine* pada Analisis Sentimen Twitter”, ditemukan bahwa kedua metode tersebut memiliki kinerja yang baik, namun metode *Naive Bayes* mengungguli metode SVM dalam hal akurasi, presisi, recall, dan f1-score [6]. Penelitian lain yang dikaji oleh Pattiha dan Hendry pada tahun 2022 membandingkan metode *K-Nearest Neighbor*, *Naive Bayes*, dan *Decision Tree* untuk menganalisis sentimen yang diungkapkan dalam tweet Twitter mengenai tanggapan dari PT PAL Indonesia. Hasil penelitian membuktikan bahwa metode *Naive Bayes* mempunyai nilai akurasi tinggi diantara ketiga metode tersebut, yaitu mencapai nilai akurasi sebesar 84,08% dengan standar deviasi +/- 1,28% [7].

Mengingat hasil penelitian sebelumnya yang mengkonfirmasi kinerja metode *Naive Bayes* yang baik dalam analisis sentimen. Harapan dari penelitian ini tidak hanya bertujuan untuk memberikan wawasan lebih dalam mengenai kepuasan pengguna terhadap aplikasi KitaLulus, namun juga memberikan masukan tambahan bagi pengembang dalam upaya meningkatkan kualitas aplikasi.

2. METODE PENELITIAN

Metodologi penelitian ini digunakan untuk memastikan keakuratan dan konsistensi dalam proses pengumpulan dan analisis data. Tahapan metodologi yang ditunjukkan pada Gambar 1 meliputi langkah-langkah penting yang dilakukan untuk memperoleh hasil penelitian yang valid.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Proses pengumpulan data dijadikan proses pembukaan, dimana data pengguna aplikasi KitaLulus diperoleh dari Google Play Store. Data dikumpulkan dengan cara *web scraping*, yang memungkinkan peneliti mengunduh ulasan pengguna secara otomatis dari halaman aplikasi Google Play Store. Dengan memanfaatkan cara *web scraping* yang memakai salah satu library pada bahasa pemrograman python yaitu *google-play-scrafer*, data review secara otomatis diekstraksi dan disusun menjadi bentuk CSV [8].

2.2. Pelabelan Dataset

Pelabelan data adalah proses pemeringkatan ulasan pengguna individual di aplikasi KitaLulus berdasarkan sentimen yang diungkapkan. Dalam studi ini, ulasan dinilai berdasarkan nilai emosionalnya. Komentar dengan peringkat 1 dan 2 ditandai sebagai "negatif", dengan peringkat 3 sebagai "netral" dan "positif" untuk nilai lainnya. Akurasi dan konsistensi dalam pelabelan menjadi penting pada cara ini memastikan bahwa data terorganisir kemudian dipakai sesuai harapan yang dimaksudkan [9].

2.3. Preprocessing

Preprocessing data merupakan tindakan penting dalam analisis sentimen, di mana teks disiapkan untuk ditinjau oleh manusia untuk digunakan dalam membangun model Naive Bayes. Proses tersebut dilakukan terlebih dahulu untuk melakukan proses data, oleh karena itu peneliti akan terhindar dari masalah data tidak tetap. Tujuan dari informasi teknis adalah klasifikasi yang mempunyai tingkat akurasi yang tinggi [10]. Tahap dari *preprocessing* untuk mengolah data pada penelitian ini yaitu *word normalization*, *case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming*. Dalam konteks normalisasi kata, proses pengolah kata dilakukan, meliputi konversi kata-kata yang menyimpang dari norma bahasa standar menjadi kata-kata yang mematuhi konvensi linguistik dan dianggap normal [11]. *Case folding* adalah proses mengubah bentuk kata agar mempunyai arti yang sama, baik dengan mengubah semuanya menjadi huruf kecil atau huruf besar. *Cleansing* adalah proses menghilangkan kata-kata yang tidak perlu untuk mengurangi kekacauan atau gangguan pada data. *Tokenizing* digunakan untuk memisahkan serangkaian kata menjadi segmen-segmen yang dipisahkan oleh spasi atau karakter khusus. Tugas *stopword removal* menghapus kata-kata yang sering nampak dan sifatnya umum, yang menunjukkan bahwa teks

yang sedang dianalisis kurang penting. *Stemming* adalah proses mengembalikan kata-kata yang mengandung sufiks ke bentuk dasar atau aslinya [12].

2.4. Pembobotan TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan teknik yang dipakai untuk mengganti teks jadi representasi numerik yang bisa dipahami oleh model *Naive Bayes*. Teknik TF-IDF dapat dijelaskan sebagai teknik yang digunakan untuk menetapkan kekerapan yang hanya terlihat pada setiap kata atau token, dimana setiap kata diberi bobot berdasarkan kepentingannya dalam dokumen. Bobot ini dihitung berdasarkan seringnya sebuah kata yang keluar di dokumen dan mengevaluasi kata tersebut relatif kepada keberadaan semua dokumen [13]. Berikut persamaan TF-IDF:

$$TF - IDF(d, t) = TD(d, t) * IDF(t) \quad (1)$$

di mana:

$$TF(d, t) = \frac{\text{Jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} \quad (2)$$

$$IDF(t) = \log \frac{\text{total dokumen}}{\text{jumlah dokumen yang mengandung kata } t} \quad (3)$$

dan:

t = kata

d = dokumen

2.5. Pembuatan Model

Setelah melakukan pembobotan TF-IDF, tahapan berikutnya ialah penerapan model memakai metode *Naive Bayes* untuk mengklasifikasikan sentimen ulasan pengguna. Pembuatan model adalah tahapan penting dalam analisis sentimen, dimana model *Naive Bayes* dilatih untuk mengklasifikasikan sentimen ulasan pengguna aplikasi KitaLulus. Implementasinya dilakukan dengan menggunakan bahasa pemrograman Python. Pengklasifikasi *Naive Bayes* adalah teknik klasifikasi yang menggunakan probabilitas sebagai dasar prediksi kelas data. Metode ini mengacu pada *teorema Bayes* dan dinilai sangat efektif dalam melakukan klasifikasi [14]. Persamaan umum algoritma klasifikasi *Naive Bayes* adalah seperti pada (4).

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (4)$$

Penjelasan:

H: Membuat asumsi data untuk kategori tertentu.

X : Data yang lapisannya belum diketahui.

P(H): Peluang H.

P(X): Peluang X.

P(H | X): Peluang suatu asumsi sesuai dengan keadaan.

P(X | H): Peluang didasarkan pada keadaan dan asumsi.

2.6. Evaluasi dan Visualisasi

Langkah terakhir adalah evaluasi yang sangat penting karena bertujuan untuk mengukur *accuracy*, *recall*, *precision*, dan *f1-score* dari model yang dibangun [14]. Kemudian pada tahap visualisasi menggunakan diagram lingkaran untuk mendapatkan gambaran yang lebih detail tentang performa model dalam mengklasifikasikan setiap label. Untuk memberikan representasi visual yang lebih jelas tentang distribusi prediksi dan performa model, kami akan menggunakan diagram lingkaran untuk menunjukkan persentase setiap label.

2.7. Pembobotan TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan teknik yang digunakan untuk mengubah teks menjadi representasi numerik yang dapat dipahami oleh model *Naive Bayes*. Teknik TF-IDF dapat dijelaskan sebagai teknik yang digunakan untuk menetapkan kekerapan yang hanya terlihat pada setiap kata atau token, dimana setiap kata diberi bobot berdasarkan kepentingannya dalam dokumen. Bobot ini dihitung berdasarkan seringnya sebuah kata muncul dalam dokumen dan mengevaluasi kata tersebut relatif kepada keberadaan semua dokumen [13]. Berikut persamaan TF-IDF:

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Penelitian ini mengumpulkan data yang memanfaatkan Google Collab untuk mengunduh data berupa ulasan pada aplikasi Kitalulus dari Google Playstore lalu menyimpannya kedalam bentuk CSV. Dalam proses akumulasi data ini, peneliti berhasil mendapat data berupa ulasan dari 597 data ulasan yang dianggap relevan berupa komentar dan skor ulasan. Data yang berhasil didapatkan kemudian diunduh dijadikan file .csv. Dibawah ini Gambar 2 yang menjadi proses penyisiran data.

```
[4] from google_play_scraper import Sort, reviews_all
import pandas as pd

app_reviews = reviews_all(
    'com.kitalulus', # ID aplikasi Kitalulus
    lang='id', # Bahasa ulasan
    country='id', # Negara
)

# Konversi hasil ke DataFrame
reviews_df = pd.DataFrame(app_reviews)

# Simpan ulasan ke dalam file CSV
reviews_df.to_csv('kitalulus_reviews.csv', index=False)

print("Ulasan telah disimpan ke dalam file kitalulus_reviews.csv")
```

Ulasan telah disimpan ke dalam file kitalulus_reviews.csv

Gambar 2. Scraping Data

Gambar 2 berisi program proses akumulasi data berupa ulasan aplikasi Kitalulus dari Google Play Store menggunakan *google_play_scraper*, mengonversinya menjadi DataFrame dengan *pandas*, dan menyimpannya ke file CSV bernama 'kitalulus_reviews.csv'.

3.2. Pelabelan Dataset

Fase pelabelan data melibatkan pengklasifikasian setiap ulasan pengguna di aplikasi Kitalulus ke dalam sentimen positif, negatif, atau netral, yang digunakan untuk melatih model Naive Bayes dalam mengukur sentimen.

	userName	content	score	label
0	Gunawan Arikusuma	Aplikasi yang mempermudah untuk mencari pekerjaan dengan fitur-fitur yang muda	5	positif
1	Aku Emmai	Aplikasi yang sangat membantu! Dengan fitur yang lengkap dan antarmuka yang intu	5	positif
2	Irwansyah Ramadan	Aplikasi yg mempermudah mencari pekerjaan sesuai pengalaman dalam bidang apa	5	positif
3	Iin Kamini	Aplikasinya sangat membantu dalam mencari pekerjaan. Diaplikasi ini selalu update	5	positif
4	Fitria Hasti	Aplikasi yg sangat membantu untuk mencari pekerjaan. Ada banyak pilihan jenis pek	5	positif
5	Nurul Aulia	Aplikasi paling terbaik di antara aplikasi2 pencaker yang lain di kelasnya semacam jo	5	positif
6	Mirna Aja	Aplikasi yang sangat membantu dalam mencari pekerjaan,kita juga bisa memilih pek	5	positif
7	Adinda Pujiwati	Apk sangat membantu dan rekomended sekali.. Terbantuu banget untuk buat CV cepa	5	positif
8	Rafni Rizqi Amelia	Aplikasi yang sangat membantu untuk mencari pekerjaan. Tampilannya menarik dan	5	positif
9	Nazwa Salsabila	Aplikasi yang sangat berguna dan membantu banyak orang. Bisa menyesuaikan peke	5	positif
10	Luwi Van Steve	Aplikasinya sangat simple, tidak lemot karena fiturnya sangat sederhana...hanya unt	5	positif
11	Lutvia Hapsiana	Gokil banget sihh sama aplikasi ini karena menyediakan loker sesuai dengan kriteria	5	positif
12	Fitri Noviantika	Aplikasinya sangat membantu dalam mencari pekerjaan. Tersedia berbagai macam k	5	positif
13	angga permana	Gimana sih ini aplikasinya,udah coba sampai 4 hp tapi tetap aja nggak mau di akses a	1	negatif
14	Anis Mu'awanah	Termasuk Aplikasi yang baik dan bagus serta mempermudah untuk mencari pekerja	4	positif

Gambar 3. Pelabelan Data

Berdasarkan nilai *score* tersebut diberi label sebagai sentimen yang dimodifikasi, yaitu “negatif” untuk nilai *score* 1 dan 2, “netral” untuk nilai *score* 3, dan “positif” untuk nilai *score* lainnya.

3.3. Preprocessing

Langkah pertama dalam *preprocessing* adalah normalisasi teks. Teks yang tidak terstruktur memerlukan proses transformasi menjadi bentuk tulisan yang teratur untuk mempersingkat proses eksplorasi data dalam analisis sentiment [15]. Pada langkah ini dilakukan normalisasi teks pada data frame untuk membakukan kata singkatan, misalnya mengubah "yg" dijadikan "yang", "dgn" dirubah "dengan", "udh" dijadikan "udah", "tp" dijadikan "tetapi", dll. Selain itu, kata-kata seperti "udah" dan "sudah", serta “tidak” dan "nggak", dianggap memiliki arti yang sama [16].

	userName	content	score	label
0	Gunawan Arikusuma	Aplikasi yang mempermudah untuk mencari pekerjaan dengan fitur-fitur yang muda	5	positif
1	Aku Emmai	Aplikasi yang sangat membantu! Dengan fitur yang lengkap dan antarmuka yang intu	5	positif
2	Irwansyah Ramadan	Aplikasi yang mempermudah mencari pekerjaan sesuai pengalaman dalam bidang a	5	positif
3	Iin Kamini	Aplikasinya sangat membantu dalam mencari pekerjaan. Diaplikasi ini selalu update	5	positif
4	Fitria Hasti	Aplikasi yang sangat membantu untuk mencari pekerjaan. Ada banyak pilihan jenis p	5	positif

Gambar 4. Normalisasi Teks

Setelah melakukan normalisasi teks, tindakan berikutnya yaitu *case folding*, dimana semua huruf pada teks review pengguna diubah menjadi huruf kecil, kemudian dilakukan tahap *cleansing* atau pembersihan teks.

	userName	content	score	label
0	Gunawan Arikusuma	aplikasi yang mempermudah untuk mencari pekerjaan dengan fitur-fitur yang mudah	5	positif
1	Aku Emmai	aplikasi yang sangat membantu dengan fitur yang lengkap dan antarmuka yang intu	5	positif
2	Irwansyah Ramadan	aplikasi yang mempermudah mencari pekerjaan sesuai pengalaman dalam bidang a	5	positif
3	Iin Kamini	aplikasinya sangat membantu dalam mencari pekerjaan. diaplikasi ini selalu update	5	positif
4	Fitria Hasti	aplikasi yang sangat membantu untuk mencari pekerjaan. ada banyak pilihan jenis p	5	positif

Gambar 5. Case Folding

Cleansing pembersihan teks dilakukan dengan menggunakan berbagai teknik termasuk penghapusan data duplikat, penghapusan teks kosong, penghapusan karakter dan tanda baca, penghapusan *Uniform Resource Locator (URL)* dan *HyperText Markup Language (HTML)*, serta penghapusan simbol dan referensi yang tidak terkait [17]. Hasil tahap pembersihan teks seperti pada Gambar 6.

	userName	content	score	label
0	Gunawan Arikusuma	aplikasi yang mempermudah untuk mencari pekerjaan dengan fiturfitur yang mudah	5	positif
1	Aku Emmai	aplikasi yang sangat membantu dengan fitur yang lengkap dan antarmuka yang intuiti	5	positif
2	Irwansyah Ramadan	aplikasi yang mempermudah mencari pekerjaan sesuai pengalaman dalam bidang a	5	positif
3	Iin Kamini	aplikasinya sangat membantu dalam mencari pekerjaan diaplikasi ini selalu update c	5	positif
4	Fitria Hasti	aplikasi yang sangat membantu untuk mencari pekerjaan ada banyak pilihan jenis pe	5	positif

Gambar 6. Cleansing

Setelah tahap pembersihan teks, langkah selanjutnya adalah proses *tokenizing*, memecah teks sehingga jadi lebih kecil untuk memudahkan analisis.

	tokenized_content
0	aplikasi yang mempermudah untuk mencari pekerjaan dengan
1	aplikasi yang sangat membantu dengan fitur yang lengkap c
2	aplikasi yang mempermudah mencari pekerjaan sesuai pengalam
3	aplikasinya sangat membantu dalam mencari pekerjaan diaplikas
4	aplikasi yang sangat membantu untuk mencari pekerjaan ada

Gambar 7. Tokenizing Data

Berikutnya adalah fase *stopword removal*, dimana kata-kata umum yang biasanya tidak mempunyai arti khusus dalam analisis teks dihilangkan seperti terlihat pada Gambar 8.

	filtered_content
0	aplikasi mempermudah mencari pekerjaan fiturfitur mudah contc
1	aplikasi membantu fitur lengkap antarmuka intuitif kitalulus m
2	aplikasi mempermudah mencari pekerjaan sesuai pengalaman bi
3	aplikasinya membantu mencari pekerjaan diaplikasi update pilih
4	aplikasi membantu mencari pekerjaan pilihan jenis pekerjaan i

Gambar 8. Stopword Removal

Setelah proses penghapusan *stopwords* selesai, langkah terakhir dalam *preprocessing* adalah *stemming*. *Stemming* adalah tahap perubahan kata yang mempunyai sufiks ke bentuk dasarnya. Proses ini penting untuk memastikan bahwa semua variasi kata dianggap sebagai entitas yang sama oleh model [18].

	stemmed_content									
0	aplikasi	mudah	cari	kerja	fiturfitur	mudah	contoh	pilih	jenis	
1	aplikasi	bantu	fitur	lengkap	antarmuka	intuitif	kitalulus	bantu		
2	aplikasi	mudah	cari	kerja	sesuai	alam	bidang	apa	admin	
3	aplikasi	bantu	cari	kerja	aplikasi	update	pilih	jenis	kerja	c
4	aplikasi	bantu	cari	kerja	pilih	jenis	kerja	tipe	kerja	cv k

Gambar 9. Stemming

3.4. Pembobotan TF-IDF

Setelah data telah melalui seluruh proses *preprocessing* yang meliputi normalisasi teks, *case folding*, pembersihan teks, *tokenizing*, *stopword removal* dan *stemming*, langkah selanjutnya adalah menghitung bobot TF-IDF.

```
# Inisialisasi TF-IDF Vectorizer
tfidf_vectorizer = TfidfVectorizer(max_features=5000)
# Transformasi teks
X = tfidf_vectorizer.fit_transform(X)
```

Gambar 10. Proses Pembobotan TF-IDF

3.5. Pembuatan Model

Tahap selanjutnya adalah pembuatan model, dimana model *machine learning* dibangun untuk tindakan analisis sentimen dari ulasan pengguna. Dari tahap ini, model *Naive Bayes* digunakan untuk klasifikasi teks. Model ini dilatih menggunakan data yang diproses dan diberi bobot TF-IDF.

```
# Bagi data menjadi training dan testing set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Terapkan SMOTE pada data training jika diperlukan
smote = SMOTE(random_state=42)
X_train_resampled, y_train_resampled = smote.fit_resample(X_train, y_train)

# Inisialisasi dan Latih model Naive Bayes
nb_model = MultinomialNB()
nb_model.fit(X_train_resampled, y_train_resampled)

# Prediksi pada data test
y_pred = nb_model.predict(X_test)
```

Gambar 11. Proses Pembuatan Model

Pada gambar 11, proses pembuatan model dilakukan dengan membaginya menjadi data latih dan data uji menggunakan rasio 80:20. Metode SMOTE kemudian diterapkan pada data latih untuk menangani permasalahan asimetri data. SMOTE digunakan sebagai teknik yang efektif mengatasi adanya kendala ketidakseimbangan pada kumpulan data, memastikan distribusi yang lebih merata antar kelas [19].

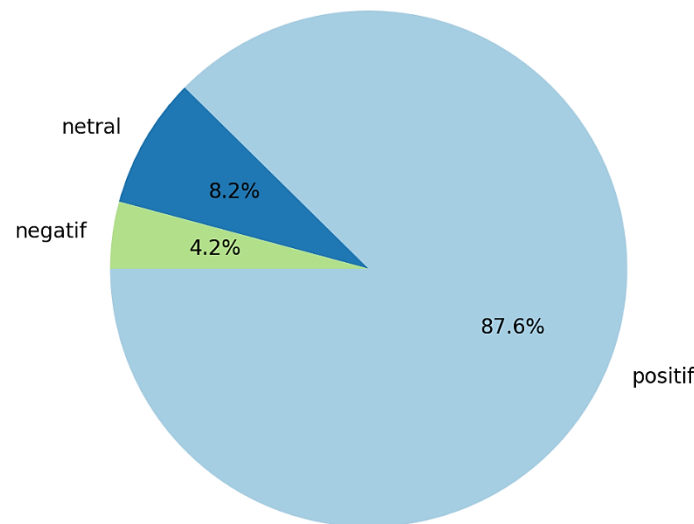
3.6. Evaluasi dan Visualisasi

Evaluasi dilakukan pada tahap ini dengan menggunakan *confusion matrix* dan *pie chart* pada visualisasi. Selama evaluasi, algoritma *Naive Bayes* mengukur *accuracy*, *recall*, *precision*, dan *f1-score*.

Tabel 1. Hasil Evaluasi

	Precision	Recall	F1-score	Support
Positif	0.99	107	0.97	0.94
Netral	0.33	4	0.46	0.75
Negatif	0.56	9	0.56	0.56
Accuracy	-	120	0.91	-
Macro avg	0.63	120	0.66	0.75
Weighted avg	0.94	120	0.92	0.91

Distribusi Kelas Ulasan



Gambar 12. Diagram Lingkaran

Berdasarkan hasil klasifikasi, terlihat bahwa model *Naive Bayes* mendapat nilai akurasi 91%. Menggali lebih dalam hasil untuk setiap kategori, kami menemukan bahwa model ini unggul dalam mengklasifikasikan kelas positif secara akurat, dengan *precision* 0,99, *recall* 0,94, dan *f1-score* sebesar 0,97. Namun, kinerjanya relatif lebih rendah ketika mengklasifikasikan kelas negatif dan netral. Kelas negatif menunjukkan *precision*, *recall*, dan *f1-score* masing-masing 0,56, sedangkan kelas netral menunjukkan *precision* 0,33, *recall* 0,75, dan *f1-score* 0,46. Dengan mempertimbangkan *macro avg*, menunjukkan *precision* (0,63), *recall* (0,75), dan *f1-score* (0,66). Di sisi lain, *weighted avg* menunjukkan bahwa model ini mempertahankan tingkat kinerjanya yang tinggi (*precision* 0.94, *recall* 0.91, *f1-score* 0.92), terutama karena jumlah sampel yang dominan di kelas positif. Confusion matrix menunjukkan bahwa model tersebut sangat baik dalam mengklasifikasikan ulasan positif dengan 101 prediksi yang benar, namun kurang akurat dalam mengklasifikasikan ulasan netral dan negatif.

4. KESIMPULAN

Temuan penelitian menunjukkan bahwa model klasifikasi *Naive Bayes* menunjukkan kinerja yang efektif dalam mengevaluasi sentimen pengguna terhadap aplikasi KitaLulus berdasarkan data tinjauan terbaru. Sebanyak 597 titik data tinjauan dikumpulkan selama fase pengumpulan data, terdiri dari komentar ulasan dan peringkat yang sesuai. Setiap ulasan dikategorikan sebagai "negatif" untuk peringkat 1 dan 2, "netral" untuk peringkat 3, dan "positif" untuk nilai peringkat lainnya. Selanjutnya, teknik pembobotan TF-IDF digunakan untuk mengubah teks ulasan menjadi format numerik. Kelemahan pada dataset yang digunakan memiliki sifat ketidakseimbangan pada setiap kelasnya. Untuk mengatasi ketidakseimbangan kelas, metode SMOTE digunakan pada data pelatihan. Hasil klasifikasi mengungkapkan akurasi model 91%. Kategori sentimen "positif" menunjukkan metrik kinerja yang patut dicontoh dengan dengan presisi 0,99, ingatan 0,94, dan skor F1 0,97.

Sebaliknya, kategori “netral” menampilkan *f1-score* yang lebih rendah sebesar 0,46, meskipun tingkat ingatan relatif tinggi yaitu 0,75. Sebaliknya, kategori “negatif” menunjukkan nilai *precision*, *recall*, dan *f1-score* yang konsisten sebesar 0,56. Singkatnya, dapat disimpulkan bahwa aplikasi KitaLulus memiliki dampak menguntungkan bagi komunitas, secara efektif mengidentifikasi dan menyajikan sebagian besar ulasan positif.

DAFTAR PUSTAKA

- [1] Teguh Suprastiyo, W., & Airlangga, P. (2021). Penyebaran Informasi Lowongan Pekerjaan Berbasis Website (Studi Kasus Mahasiswa Dan Alumni Unwaha). *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 4(2), 230. <https://doi.org/10.37600/tekinkom.v4i2.400>
- [2] Hasri, C. F., & Alita, D. (2022). Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 3(2), 145-160. <https://doi.org/10.33365/jatika.v3i2.2026>
- [3] Septiani Gumilar, T., Astuti, R., & Arie Wijaya, Y. (2024). Analisis Sentimen Ulasan Aplikasi Lita Di Play Store Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 543-550. <https://doi.org/10.36040/jati.v8i1.8778>
- [4] Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 6, Issue 1). <http://dx.doi.org/10.30645/j-sakti.v6i1.443>
- [5] Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 785-795. <https://doi.org/10.35957/jatisi.v9i2.1835>
- [6] Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10(02), 71-76. <https://doi.org/10.32664/smatika.v10i02.455>
- [7] Pattiha, F. S., & Hendry, H. (2022). Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 506. <https://doi.org/10.30865/jurikom.v9i2.4016>
- [8] Rahma Firmansyah, D., & Lestariningsih, E. (2024). Analisis Sentimen Ulasan Aplikasi Smart Campus Unisbank di Google Playstore Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informasi Dan Komunikasi*, 8(2), 2024. <https://doi.org/10.35870/jti>
- [9] Wijaya, D., Saputra, R. A., & Irwiensyah, F. (2024). Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(4). <https://doi.org/10.30865/klik.v4i4.1738>
- [10] Anwar, K. (2022). Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 2(4), 148-155. <https://doi.org/10.30865/klik.v2i4.315>
- [11] Yutika, C. H., Adiwijaya, A., & Faraby, S. al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422. <https://doi.org/10.30865/mib.v5i2.2845>
- [12] Mastan, I. A., & Toni, Y. (2020). Analisis Sentimen Terhadap Tempat Kuliner Ayam Gedebuk Dari Komentar Pengunjung Dengan Menggunakan Metode Naïve Bayes Classifier. *JBASE - Journal of Business and Audit Information Systems*, 3(1). <https://doi.org/10.30813/jbase.v3i1.2062/>
- [13] Rozi, F., Sukmana, F., & Adani Nabil, M. (2021). Pengelompokan Judul Buku dengan Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Term Frequency - Inverse

- Document Frequency (TF-IDF). *JIMP: Jurnal Informatika Merdeka Pasuruan*, 3(1). <http://dx.doi.org/10.51213/jimp.v6i3.346>
- [14] Zhafran Muflih, H., Rizki Abdillah, A., & Noor Hasan, F. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(3), 1613–1621. <https://doi.org/10.30865/klik.v4i3.1303>
- [15] Tsani, M., Rupaka, A., Asmoro, L., & Pradana, B. (2020). Analisis Sentimen Review Transportasi Menggunakan Algoritma Support Vector Machine Berbasis Chi Square. *Smart Comp :Jurnalnya Orang Pintar Komputer*, 9(1), 35–39. <https://doi.org/10.30591/smartcomp.v9i1.1817>
- [16] Nehe, P. H., Berutu, S. S., & Budiati, H. (2024). Analisis Sentimen Opini Masyarakat Terhadap Presiden Jokowi Sebelum Dan Sesudah Pilpres 2024 Menggunakan Metode Naive Bayes Classification. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 13(1), 451. <https://doi.org/10.35889/jutisi.v13i1.1841>
- [17] Ramadhani, B., & Suryono, R. R. (2024). Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse. *Jurnal Media Informatika Budidarma*, 8(2). <https://doi.org/10.30865/mib.v8i2.7458>
- [18] Darmawan, G., Alam, S., & Imam Sulisty, M. (2023). Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes. *Storage - Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2(3), 100–108. <https://doi.org/10.55123>
- [19] Utami, H. (2022). Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network. *Indonesian Journal of Applied Statistics*, 5(1), 31. <https://doi.org/10.13057/ijas.v5i1.56825>