



ANALISA PENGGUNAAN METODE LEXICON BASED DAN ALGORITMA NAIVE BAYES PADA SENTIMEN ULASAN APLIKASI DUOLINGO

Muhammad Abib Allesdio¹, Ade Irma Purnamasari², Irfan Ali³, Nana Suarna⁴, Agus Bahtiar⁵

Program Studi Rekayasa Perangkat Lunak, Sekolah Tinggi Manajemen Informatika Dan Komputer
Cirebon, Indonesia ^{1,2,3,4,5}

muhammadabibalesdio@gmail.com¹, irma2974@yahoo.com², irfanaali0.0@gmail.com³, st_nana@yahoo.com⁴,
agusbahtiar038@gmail.com⁵

ABSTRAK

Peningkatan jumlah ulasan pengguna pada aplikasi mobile membuka peluang untuk memahami persepsi dan pengalaman pengguna melalui analisis sentimen. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Duolingo yang diambil dari Google Play Store menggunakan dua pendekatan, yaitu metode lexicon-based dan algoritma Naive Bayes berbasis Python. Metode lexicon-based digunakan untuk memberikan skor polaritas berdasarkan leksikon sentimen, sedangkan Naive Bayes diterapkan sebagai model klasifikasi dengan dukungan fitur TF-IDF. Proses penelitian meliputi tahapan pengumpulan data, preprocessing teks (cleaning, case folding, tokenisasi, stopword removal, dan stemming), pembobotan sentimen, pelatihan model, serta evaluasi performa menggunakan accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa metode lexicon-based mampu memberikan gambaran umum polaritas ulasan, namun performanya sangat dipengaruhi oleh kelengkapan leksikon dan variasi bahasa informal pengguna. Sementara itu, algoritma Naive Bayes menunjukkan performa yang lebih stabil dan akurasi lebih tinggi dalam mengklasifikasikan sentimen dibandingkan pendekatan leksikon. Perbandingan kedua metode memperlihatkan bahwa Naive Bayes lebih efektif dalam menangani data teks pendek, tidak terstruktur, serta mengakomodasi variasi kata dan ejaan. Temuan penelitian ini memberikan pemahaman yang lebih dalam mengenai persepsi pengguna terhadap Duolingo serta menjadi referensi metodologis bagi penelitian sentiment analysis selanjutnya, khususnya yang melibatkan kombinasi metode leksikon dan klasifikasi probabilistik.

Kata kunci: Analisis Sentimen, Duolingo, Lexicon-Based, Naive Bayes, Natural Language Processing, Python.

1. PENDAHULUAN

Perkembangan aplikasi mobile mendorong meningkatnya volume ulasan pengguna yang berisi opini terkait pengalaman, kepuasan, serta permasalahan selama menggunakan aplikasi. Sentiment analysis menjadi pendekatan penting untuk mengolah ulasan tersebut karena dapat mengekstraksi opini secara sistematis dari data teks berskala besar. Berbagai penelitian menunjukkan bahwa metode lexicon-based maupun machine learning mampu mengidentifikasi sentimen pengguna terhadap aspek kegunaan aplikasi, kepuasan fitur, maupun hambatan layanan

[1]. Dengan demikian, sentiment analysis berperan sebagai fondasi ilmiah dalam memahami kebutuhan pengguna serta mendukung peningkatan kualitas aplikasi.

Metode lexicon-based banyak diterapkan karena sifatnya yang interpretable, ringan secara komputasi, serta tidak memerlukan data berlabel dalam jumlah besar. Studi-studi mutakhir menegaskan bahwa pendekatan ini tetap relevan di tengah dominasi model machine learning dan deep learning. Di sisi lain, algoritma Naive Bayes digunakan secara luas sebagai baseline dalam klasifikasi sentimen karena performanya yang stabil pada data teks. Meskipun

demikian, sejumlah studi melaporkan bahwa efektivitas Naive Bayes dipengaruhi oleh kualitas fitur, teknik penyeimbangan kelas, serta karakteristik bahasa. Akurasi mencapai 79% ketika TF-IDF diterapkan, sedangkan penurunan performa pada konteks multibahasa dan integrasi SMOTE diperlukan untuk mengatasi ketidakseimbangan kelas [2]. Temuan tersebut memperlihatkan bahwa Naive Bayes tetap kompetitif, namun membutuhkan rekayasa fitur yang tepat agar hasil klasifikasi optimal.

Kemajuan natural language processing (NLP) semakin memperkuat kualitas sentiment analysis dengan menghadirkan teknik representasi teks yang lebih kontekstual. Model berbasis transformer seperti BERT mampu menangkap nuansa, sarkasme, dan emosi implisit secara lebih baik dibandingkan pendekatan tradisional [3]. Penerapan fine-tuning BERT pada ulasan aplikasi bahkan menghasilkan peningkatan signifikan pada akurasi dan F1-score. Walaupun demikian, model transformer memerlukan komputasi besar sehingga tidak selalu menjadi opsi utama pada penelitian tingkat sarjana yang membutuhkan efisiensi. Oleh karena itu, penggunaan metode lexicon-based sebagai pelabel awal dan Naive Bayes sebagai pengklasifikasi tetap menjadi kombinasi ideal dan terukur.

Aplikasi Duolingo merupakan objek penelitian yang relevan karena memiliki jutaan ulasan

2. LANDASAN TEORI

2.1 Analisis Sentimen dan Text Mining

Dalam teks, baik positif, negatif, maupun netral. Secara teoretis, analisis ini berakar pada linguistik komputasional, penambahan data, serta pemodelan probabilistik yang memformulasikan teks sebagai kumpulan fitur statistik. Penelitian NLP mutakhir menegaskan bahwa representasi makna dalam teks dapat ditingkatkan melalui deteksi sinyal wacana berbasis jaringan saraf, yang menunjukkan bahwa penangkapan hubungan semantik meningkatkan kualitas klasifikasi sentimen.

2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan fondasi utama dalam analisis teks, memungkinkan pemrosesan bahasa alami menjadi informasi terstruktur. (Ayele 2021) menunjukkan bahwa adaptasi CRISP-DM dalam idea mining memungkinkan NLP mendukung identifikasi pola dalam korpus besar melalui topik dinamis. menambahkan bahwa basis pengetahuan multifungsi meningkatkan integrasi pengetahuan otomatis dalam text mining sehingga memperkuat interpretabilitas [5]. Pendekatan deep learning yang diuraikan menunjukkan efektivitas arsitektur inspirasi

pengguna di platform seperti Google Play Store. Ulasan tersebut mencerminkan pengalaman belajar, tingkat kepuasan, hambatan teknis, dan persepsi terhadap fitur pembelajaran. Penelitian sebelumnya menunjukkan bahwa analisis sentimen pada ulasan Duolingo memberikan wawasan berharga terkait emosi, pola penggunaan, dan persepsi pengguna terhadap efektivitas aplikasi [4]. Studi lain juga menggunakan sentiment analysis untuk membandingkan aplikasi pembelajaran bahasa dan menilai kualitas pengalaman belajar [5]. Kelimpahan ulasan serta relevansinya dengan konteks pembelajaran bahasa menjadikan Duolingo sebagai objek studi ideal.

Melalui landasan ilmiah tersebut, penerapan metode lexicon-based dan algoritma Naive Bayes pada ulasan pengguna Duolingo menjadi penting untuk mengidentifikasi sentimen pengguna secara terstruktur. Analisis ini tidak hanya memberikan gambaran umum tentang kepuasan pengguna, tetapi juga membantu mengungkap aspek-aspek yang perlu ditingkatkan dalam desain antarmuka, kinerja fitur, serta efektivitas pembelajaran. Dengan demikian, penelitian ini berkontribusi pada upaya pengembangan aplikasi pembelajaran bahasa berbasis mobile yang lebih responsif terhadap kebutuhan dan pengalaman pengguna.

FastText dalam memahami konteks semantik pada korpus berskala besar.

2.3 Text Preprocessing

Preprocessing merupakan tahap penting dalam analisis sentimen untuk membersihkan dan menormalkan teks sebelum diproses. Teknik seperti cleaning, case folding, tokenization, stopword removal, dan stemming terbukti meningkatkan kualitas model. pendekatan lexicon-based yang diperbaiki (IVADER) mampu mencapai performa tinggi melalui leksikon khusus domain. (Mutinda and al. 2023) menunjukkan bahwa penggabungan leksikon dengan embedding BERT serta CNN membantu menyoroti bagian teks yang mengandung sentimen, meningkatkan akurasi model hibrid.

2.4 Pendekatan Lexicon-Based

Pendekatan Lexicon-Based bekerja dengan mengacu pada daftar kata yang telah diberi skor polaritas, memungkinkan klasifikasi sentimen yang cepat dan dapat dijelaskan tanpa memerlukan dataset berlabel besar [6]. Leksikon domain specific memberikan kinerja tinggi dalam berbagai domain ulasan. Fitur leksikal dapat melengkapi representasi deep learning untuk meningkatkan deteksi segmen opini penting

Efektivitas metode ini dalam bahasa Indonesia, terutama saat data berlabel terbatas Metode leksikon masih relevan dan dapat diskalakan untuk analisis real-time di berbagai domain.

2.5 Algoritma Naïve Bayes

Naive Bayes merupakan algoritma klasifikasi probabilistik yang menghitung posterior dari setiap kelas berdasarkan probabilitas fitur dalam dokumen. Metode ini mengasumsikan independensi antar kata sehingga menghitung $(P(w|class))$ untuk seluruh kosakata dan memilih kelas dengan posterior tertinggi. Penelitian menunjukkan bahwa lexicon-based dapat menjadi fitur tambahan yang sangat efektif bagi Naive Bayes, terutama dalam domain ulasan aplikasi yang kaya opini.

Lexicon dapat berfungsi sebagai filter awal yang melengkapi pembelajaran probabilistik. Token sentimen yang diperkuat leksikon membantu Naive Bayes mengenali pola yang tidak muncul pada frekuensi kata biasa. Naïve Bayes efektif digunakan pada ulasan aplikasi, Google Play, dan Twitter, terutama ketika dipadukan dengan TF-IDF dan smoothing [7].

2.6 Kerangka Teori

Penelitian ini berfokus pada analisis sentimen ulasan pengguna aplikasi dengan menggabungkan dua pendekatan utama, yaitu Lexicon-Based dan Naive Bayes. Pendekatan lexicon-based bekerja dengan menggunakan kamus kata bersentimen untuk menentukan polaritas teks. Metode ini memiliki keunggulan seperti:

- a) Tidak membutuhkan data berlabel besar
- b) Cepat dan mudah diinterpretasikan
- c) Bersifat general dan dapat digunakan lintas domain

Beberapa penelitian menunjukkan bahwa metode ini tetap relevan dan bahkan kompetitif dibandingkan metode machine learning modern, terutama untuk teks pendek seperti ulasan aplikasi. Sementara itu, Naive Bayes merupakan metode klasifikasi probabilistik yang menghitung peluang suatu teks termasuk ke dalam kelas tertentu (positif, negatif, netral) berdasarkan frekuensi kata. Keunggulannya meliputi:

- a) Sederhana dan efisien secara komputasi
- b) Mampu menangkap pola statistik dalam data

Dalam pendekatan hybrid, kedua metode ini saling melengkapi:

- a) Lexicon-based memberikan interpretasi awal dan fitur tambahan

- b) Naïve Bayes meningkatkan akurasi melalui pembelajaran data

Dengan demikian, kombinasi keduanya dapat menghasilkan klasifikasi sentimen yang lebih akurat, khususnya pada teks pendek dan informal seperti ulasan aplikasi.

2.7 Kerangka Konseptual

Kerangka konseptual penelitian ini terdiri dari tiga komponen utama: input, proses, dan output.

a. Input

- a) Data ulasan pengguna aplikasi Duolingo dari Google Play
- b) Berupa teks pendek, informal, dan mengandung ekspresi emosional

b. Proses

1. Preprocessing Data
 - a) Cleaning
 - b) Case folding
 - c) Tokenization
 - d) Stopword removal
 - e) Stemming
2. Lexicon-Based Analysis
 - a) Pemberian skor sentimen pada tiap kata
 - b) Menghasilkan polaritas awal
 - c) Digunakan sebagai label awal, fitur tambahan, atau pembanding
3. Klasifikasi Naive Bayes
 - a) Menghitung probabilitas kelas berdasarkan frekuensi kata
 - b) Menghasilkan prediksi sentimen berbasis model
4. Evaluasi Model
 - a) Menggunakan metrik: accuracy, precision, recall, dan F1-score
 - b) Membandingkan performa metode tunggal dan hybrid

c. Output

- a) Klasifikasi sentimen (positif, negatif, netral)
- b) Analisis kinerja metode
- c) Rekomendasi untuk peningkatan kualitas aplikasi

3. METODOLOGI PENELITIAN

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen kuantitatif untuk mengevaluasi kinerja metode lexicon-based dan Naive Bayes dalam klasifikasi sentimen ulasan Duolingo. Desain eksperimen dipilih karena memungkinkan pengujian

model secara sistematis, terkontrol, dan dapat direplikasi. Metode lexicon-based digunakan untuk menghasilkan skor sentimen awal yang bersifat interpretatif dan tidak memerlukan data berlabel besar. Sementara itu, Naive Bayes digunakan sebagai classifier probabilistik. Penelitian ini menggabungkan keduanya dalam pendekatan hybrid, di mana skor lexicon dapat menjadi hasil awal maupun fitur tambahan. Tahapan utama meliputi pengumpulan data, preprocessing, lexicon scoring, pelatihan model, dan evaluasi, dengan penekanan pada reproducibility seperti penggunaan environment dan parameter yang konsisten.

3.2 Sumber Data dan Alat Penelitian

3.2.1 Sumber Data

Data berasal dari ulasan pengguna aplikasi Duolingo di Google Play Store yang berbentuk teks tidak terstruktur, pendek, dan informal. Data ini relevan karena mencerminkan opini pengguna secara langsung dan cocok untuk analisis sentimen.

Pemilihan data didasarkan pada:

- Ketersediaan data dalam jumlah besar
- Kesesuaian dengan metode lexicon-based
- Dukungan terhadap pendekatan hybrid lexicon–Naïve Bayes

Data dikumpulkan melalui web scraping atau tools pihak ketiga dalam format CSV/JSON.

3.2.2 Alat, Perangkat Lunak, dan Library

Penelitian menggunakan Python sebagai bahasa utama dengan tools seperti Jupyter Notebook dan Anaconda untuk pengelolaan environment.

Library yang digunakan:

- NLTK → preprocessing teks
- Scikit-learn → TF-IDF & Naïve Bayes
- Pandas & NumPy → pengolahan data
- VADER / INSET → lexicon-based scoring
- TextBlob → utilitas sentimen
- Matplotlib & Seaborn → visualisasi

Pemilihan alat ini didasarkan pada efisiensi, kemudahan penggunaan, dan standar dalam riset NLP.

3.2.3 Justifikasi Pemilihan

Pemilihan metode dan alat didukung oleh:

- Efektivitas lexicon-based untuk data terbatas
- Kesesuaian Naïve Bayes untuk teks pendek
- Kemampuan Python dalam mendukung pipeline NLP
- Kekuatan pendekatan hybrid dalam meningkatkan akurasi

3.3 Tahapan Penelitian

3.3.1 Pengumpulan Data

Data dikumpulkan dari Google Play Store melalui proses scraping yang menghasilkan dataset berisi ulasan, rating, dan metadata. Tahap ini bertujuan memperoleh data mentah sebagai bahan analisis.

3.3.2 Preprocessing Teks

Tahap ini bertujuan membersihkan dan menormalkan teks agar siap dianalisis. Langkah-langkahnya meliputi:

- Cleaning (menghapus simbol/karakter tidak perlu)
- Case folding (huruf kecil)
- Tokenisasi
- Normalisasi slang
- Stopword removal
- Stemming

Preprocessing penting untuk mengurangi noise dan meningkatkan akurasi model.

3.3.3 Lexicon-Based Approach

Metode ini menggunakan kamus sentimen (VADER, INSET) untuk memberikan skor polaritas pada kata. Hasilnya digunakan untuk:

- Menentukan sentimen awal
- Melakukan labeling otomatis
- Menjadi fitur tambahan untuk model
- Lexicon-based unggul dalam interpretasi dan cocok untuk teks informal.

3.3.4 Klasifikasi Naïve Bayes

Naive Bayes digunakan untuk mengklasifikasikan sentimen berdasarkan probabilitas. Model ini menggunakan fitur TF-IDF dan dapat dikombinasikan dengan skor lexicon.

Faktor penting:

- Representasi fitur (TF-IDF)
- Smoothing (Laplace)
- Validasi (cross-validation)
- Penyesuaian prior
- Metode ini efektif, cepat, dan cocok untuk dataset teks pendek.

3.3.5 Evaluasi Performa Model

Evaluasi dilakukan menggunakan:

- Akurasi
- Precision, Recall, F1-Score
- Confusion Matrix

Evaluasi ini membantu memahami performa model dan kesalahan klasifikasi, terutama pada kelas yang sering tertukar.

3.4 Rancangan Sistem / Arsitektur

Sistem mengikuti pipeline terstruktur:

- Pengumpulan data
- Preprocessing teks
- Ekstraksi fitur (TF-IDF + lexicon scoring)
- Klasifikasi (Naïve Bayes / hybrid)
- Output sentimen

Pendekatan ini menggabungkan kelebihan interpretasi lexicon dan kekuatan probabilitistik Naive Bayes.

3.5 Teknik Evaluasi dan Validasi

Validasi model menggunakan:

- Confusion Matrix untuk melihat kesalahan klasifikasi
- Precision, Recall, F1-Score untuk mengukur performa secara seimbang
- Pendekatan ini memastikan evaluasi yang adil, terutama pada data yang tidak seimbang.

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Penelitian diperoleh dari ulasan pengguna aplikasi Duolingo pada platform Google Play Store. Proses pengambilan data dilakukan menggunakan *Google Play Scraper* sehingga seluruh ulasan dapat diambil secara otomatis tanpa perlu melakukan proses *crawling* manual.

Dataset mencakup: Jumlah ulasan terkumpul: 10000 Atribut yang diperoleh : Review ID, username Rating, Review text Data digunakan sebagai sumber masukan (input) untuk proses pembersihan teks, analisis sentimen, dan klasifikasi menggunakan Naive Bayes.

Tabel 1. Dataset hasil Scraping

NO	Review ID	Username	Rating	Review text	Date
1	d7bb9708-6aea-4fa4-8995-7dc1a4be2288	Pengguna Google	5	Suka ini belajar sambil main game	2025-11-06 02:22:38
2	f710f90a-6961-4765-8042-262692de14f	Pengguna Google	1	ada percakapan menurus ke LGBT, masa bilang'	2025-11-06 01:56:45
3	e26fb49f-cdef-4f28-8a16-86d16b7f7091	Pengguna Google	4	aplikasi yang bermanfaat good job	2025-11-06 01:46:2
4	927bdf69-895c-4cc2-ae52-152fabbedfd	Pengguna Google	3	Tidak bisa menambahkan score duolingo ke Lin.	2025-11-06 00:43:46
5	36d2ed17-948a-4b31-a90f-e21616e2284e	Pengguna Google	4	aplikasinya bagus	2025-11-06 00:17:36

4.1.2 Preprocessing Teks

. Proses ini sangat penting karena kualitas data mentah sangat mempengaruhi akurasi model klasifikasi. Langkah-langkah *preprocessing* yang diterapkan dalam penelitian ini mencakup beberapa teknik standar dalam pemrosesan bahasa alami (NLP), yaitu sebagai berikut.

1. Cleaning

Langkah pertama adalah menghapus unsur-unsur yang tidak relevan seperti: emoji, angka, URL atau tautan, Karakter khusus (misalnya @, %, &, #), tanda baca yang tidak diperlukan Tahap ini ditujukan untuk mengurangi *noise* yang dapat mengganggu proses ekstraksi fitur. *cleaning* merupakan tahap krusial untuk meningkatkan keterbacaan model terhadap teks [8].

2. Case Folding

Tujuan utama tahap ini adalah menghilangkan perbedaan makna akibat variasi kapitalisasi. Case folding merupakan prosedur dasar pada hampir semua pipeline NLP modern.

3. Tokenization

Tokenisasi mempermudah analisis pada tingkat kata dan memungkinkan algoritma seperti Naive Bayes bekerja berdasarkan frekuensi kemunculan kata tertentu.

4. Normalization / Slang Mapping

Pada tahap ini, kata tidak baku, singkatan, atau *slang* diubah ke bentuk baku menggunakan *slang dictionary*. Normalisasi membantu mengurangi variasi kata yang sebenarnya memiliki makna serupa. Normalisasi sangat membantu meningkatkan performa model berbasis teks Bahasa Indonesia.

5. Stopword Removal

Stopword removal adalah proses menghapus kata-kata umum yang tidak memberi kontribusi terhadap analisis sentimen, seperti: "yang", "dan", "atau", "di", "ke", "dari". Penghapusan stopwords bertujuan agar model lebih fokus pada kata bermakna sentimen. Metode ini dapat meningkatkan efisiensi dan akurasi klasifikasi [10].

6. Stemming

Stemming mengubah kata ke bentuk dasarnya (*root word*) menggunakan algoritma *Sastrawi Stemmer*, misalnya:

- "menggunakan" → "guna"
- "belajarannya" → "belajar"
- "membantu" → "bantu"

Stemming penting karena mengurangi variasi bentuk kata dan memperkuat representasi fitur pada model pembelajaran mesin.

Kesimpulan Tahap Preprocessing

Seluruh rangkaian preprocessing di atas berhasil:

1. mengurangi derau (*noise*)
2. memperjelas struktur bahasa dalam ulasan
3. meningkatkan kualitas input sentimen
4. memperkuat kinerja metode lexicon-based
5. mengoptimalkan akurasi model Naive Bayes

Dengan data yang lebih bersih dan terstruktur, model dapat mempelajari pola sentimen secara lebih akurat dan konsisten.

Tabel 2. Contoh Hasil Preprocessing

No	Review Text (Asli)	Cleaning	Case Folding	Normalisasi	Tokenize	Stopword Removal	Stemming Data
1	Suka ini belajar sambil main game	Suka ini belajar sambil main game	suka ini belajar sambil main game	suka ini belajar sambil main game	[suka, ini, belajar, sambil, main, game]	[suka, belajar, main, game]	suka ajar main game
2	bagus bangettt	bagus bangettt	bagus bangettt	bagus bangettt	[bagus, bangettt]	[bagus, bangettt]	bagus bangettt
3	bagus buat belajar bahasa inggris	bagus buat belajar bahasa inggris	bagus buat belajar bahasa inggris	bagus buat belajar bahasa inggris	[bagus, buat, belajar, bahasa, inggris]	[bagus, belajar, bahasa, inggris]	bagus ajar bahasa inggris
4	ada percakapan menurus ke LGBT, masa bilang '	ada percakapan menurus ke LGBT, masa bilang I...	ada percakapan menurus ke lgbt masa bilang i	ada percakapan menurus ke lgbt masa bilang i	[ada, percakapan, menurus, ke, lgbt, masa, bilang, i, bi...]	[percakapan, menurus, lgbt, bilang, i, love, ...]	cakap jurus lgbt bilang i love u wanita auto u
5	Karena gw bisa bahasa ingrt	Karena gw bisa bahasa ingrt	karena gw bisa bahasa ingrt	karena gw bisa bahasa ingrt	[karena, gw, bisa, bahasa, ingrt]	[gw, bahasa, ingrt]	gw bahasa ingrt

4.1.3 Pembobotan Sentimen Menggunakan Metode Lexicon-Based

Pembobotan sentimen dalam penelitian ini dilakukan menggunakan pendekatan lexicon-based, yaitu metode yang menentukan polaritas sentimen berdasarkan daftar kata yang telah diberi label positif atau negatif dalam sebuah kamus sentimen (*sentiment lexicon*). Setiap kata pada ulasan dibandingkan dengan daftar kata dalam lexicon untuk menentukan nilai polaritasnya. Metode ini digunakan sebagai *baseline* sekaligus sebagai dasar dalam memberikan label awal yang kemudian dipakai pada proses pelatihan model Naive Bayes

Tabel 3. Preview

No	Date	Username	Rating	Review Text	cleaning	case folding	normalisasi	tokenize	stopword removal	stemming data
1	2023-07-12	user_a	5	"Great app..."	great app	great app	great app	['great', 'app']	['great', 'app']	great app
2	2023-07-12	user_b	4	"Helpful..."	helpful	helpful	helpful	['helpful']	['helpful']	helpful
3	2023-07-11	user_c	1	"Too many ads..."	too many ads	too many ads	too many ads	['too', 'many', 'ads']	['many', 'ads']	banyak iklan
4	2023-07-10	user_d	3	"Good but slow"	good but slow	good but slow	good slow	['good', 'slow']	['good', 'slow']	bagus lambat
5	2023-07-10	user_e	5	"Very useful!"	very useful	very useful	very useful	['very', 'useful']	['useful']	berguna

1. Dasar Kerja Lexicon-Based

Pendekatan lexicon-based mengacu pada prinsip bahwa setiap kata memiliki kecenderungan emosional tertentu yang dapat dihitung.

1. Kata bernilai positif diberi bobot positif (misalnya +1).
2. Kata bernilai negatif diberi bobot negatif (misalnya -1).

Jika sebuah ulasan mengandung beberapa kata yang termasuk dalam lexicon, bobotnya dijumlahkan untuk menentukan sentimen akhir.

2. Proses Pembobotan Sentimen

Tahapan pembobotan sentimen mencakup:

a. Ekstraksi Kata dari Teks

Setelah melalui preprocessing (cleaning, tokenisasi, normalisasi, dan stemming), ulasan diubah menjadi daftar kata bersih yang siap dibandingkan dengan lexicon.

b. Pencocokan dengan Kamus Sentimen

Setiap kata dibandingkan dengan lexicon Bahasa Indonesia yang memuat ribuan kosakata positif dan negatif. Lexicon yang digunakan umumnya bersumber dari penelitian terdahulu atau sumber NLP Indonesia yang telah divalidasi.

c. Pemberian Bobot

Bobot diberikan sebagai berikut:

Tabel 4. Bobot kata

Jenis Kata	Bobot
Kata positif	+1
Kata negatif	-1
Kata tidak ditemukan di lexicon	0

Total skor sentimen ulasan dihitung dengan rumus sederhana:

$$\text{Skor Sentimen} = \sum(\text{Bobot kata positif}) - \sum(\text{Bobot kata negatif})$$

d. Penentuan Label Sentimen

Berdasarkan skor yang diperoleh, ulasan diklasifikasikan menjadi tiga kategori:

1. Positif → skor > 0
2. Netral → skor = 0
3. Negatif → skor < 0

Pendekatan ini memungkinkan klasifikasi awal tanpa perlu melibatkan model statistik.

3. Keunggulan Metode Lexicon-Based dalam Penelitian Ini

Beberapa alasan metode lexicon-based dipilih dalam penelitian:

1. Mudah diterapkan dan transparan setiap keputusan sentimen dapat ditelusuri langsung melalui kata-kata penyusunnya.
2. Cocok untuk Bahasa Indonesia yang struktur kalimatnya relatif stabil dan banyak menggunakan kata sifat untuk menggambarkan pengalaman pengguna.
3. Dapat menjadi label dasar untuk pelatihan model Naive Bayes yang membutuhkan data berlabel.

4. Hasil Pembobotan pada Dataset

Hasil pembobotan menunjukkan variasi sentimen yang mencerminkan pengalaman pengguna Duolingo. Kata-kata positif seperti “bagus”, “membantu”, “efektif”, “seru” muncul dominan pada ulasan dengan rating tinggi.

Sebaliknya, kata negatif seperti “bug”, “error”, “gangguan”, “kecewa” banyak muncul pada ulasan dengan rating rendah.

Hasil bobot sentimen inilah yang kemudian dijadikan label dasar (ground truth) untuk tahapan klasifikasi menggunakan algoritma Naive Bayes.

4.1.4 Klasifikasi Menggunakan Algoritma Naive Bayes

Tahap klasifikasi dilakukan menggunakan algoritma Multinomial Naive Bayes, yaitu salah satu metode *supervised learning* yang banyak digunakan dalam analisis teks karena kemampuannya yang baik dalam memodelkan distribusi kata pada dokumen. Algoritma ini bekerja berdasarkan prinsip *Bayes theorem* dengan asumsi bahwa setiap fitur (kata) bersifat independen satu sama lain.

Dalam penelitian ini, Naive Bayes digunakan untuk memprediksi kategori sentimen (positif, netral, negatif) dari ulasan pengguna aplikasi Duolingo berdasarkan representasi teks yang telah diproses.

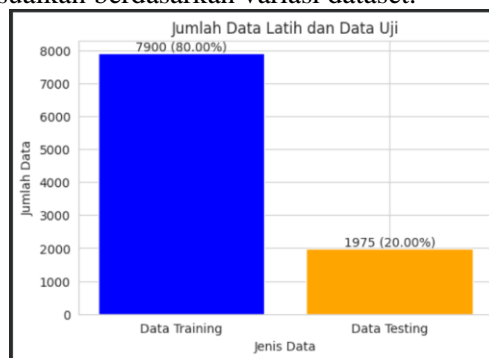
Jurnal Sistem Informasi dan Teknologi (SINTEK)

1. Pembagian Data (Train-Test Split)

Dataset yang telah diberi label menggunakan metode lexicon-based kemudian dibagi menjadi dua bagian:

1. Data latih (training set): digunakan untuk melatih model
2. Data uji (testing set): digunakan untuk mengukur kemampuan model

Umumnya proporsi pembagian adalah 80% data latih dan 20% data uji, meskipun proporsi dapat disesuaikan berdasarkan variasi dataset.



Gambar 1. Pembagian data

2. Ekstraksi Fitur Teks

Sebelum model Naive Bayes dilatih, teks ulasan harus diubah ke bentuk numerik melalui algoritma representasi teks. Dua pendekatan yang umum digunakan adalah:

1. Bag-of-Words (BoW): menghitung frekuensi kemunculan kata
2. TF-IDF (Term Frequency-Inverse Document Frequency): memberikan bobot yang lebih tinggi pada kata yang jarang muncul namun informatif

Pada penelitian ini fitur dibangun menggunakan TF-IDF (Term Frequency-Inverse Document Frequency). Representasi fitur ini membentuk matriks dokumen-kata yang kemudian menjadi input bagi model Naive Bayes.

3. Pelatihan Model Naive Bayes

Model Multinomial Naive Bayes dilatih menggunakan data latih yang telah direpresentasikan ke dalam bentuk fitur numerik. Proses pelatihan melibatkan:

1. menghitung probabilitas setiap kata muncul pada ulasan dengan kategori tertentu
2. menghitung probabilitas prior untuk setiap kelas sentimen
3. menyimpan distribusi fitur yang digunakan pada proses prediksi

Model kemudian belajar mengenali pola distribusi kata yang umum muncul pada ulasan positif, negatif, maupun netral.

4. Proses Prediksi

Setelah proses pelatihan selesai, model digunakan untuk memprediksi sentimen pada data uji. Prediksi dilakukan dengan memilih kelas dengan probabilitas posterior tertinggi berdasarkan:

$$P(\text{kelas} | \text{fitur}) = \frac{P(\text{fitur} | \text{kelas}) \cdot P(\text{kelas})}{P(\text{fitur})}$$

Model akan menghitung kemungkinan setiap ulasan termasuk ke kategori:

1. Positif
2. Netral
3. Negatif

Prediksi ini kemudian dibandingkan dengan label asli untuk mendapatkan nilai akurasi dan metrik evaluasi lainnya.

5. Keluaran Klasifikasi

Proses klasifikasi menghasilkan:

1. Label sentimen hasil prediksi model
2. Akurasi model dalam mengklasifikasikan ulasan
3. *Confusion matrix* untuk melihat distribusi kesalahan prediksi
4. Nilai precision, recall, dan F1-score pada setiap kelas

4.1.5 Evaluasi Performa Model

Tahap evaluasi dilakukan untuk mengukur seberapa baik model Multinomial Naive Bayes dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Duolingo. Evaluasi ini sangat penting untuk memastikan bahwa model tidak hanya bekerja dengan baik pada data latih, tetapi juga mampu menggeneralisasi pada data baru.

Dalam penelitian ini, evaluasi dilakukan menggunakan beberapa metrik standar dalam analisis sentimen dan klasifikasi teks, yaitu akurasi, precision, recall, F1-score, dan confusion matrix.

1. Confusion Matrix

Confusion matrix menggambarkan distribusi prediksi model terhadap label sebenarnya. Matriks ini menunjukkan:

1. jumlah ulasan positif yang diprediksi positif
2. jumlah ulasan netral yang diprediksi negatif, dan seterusnya

Tabel ini memberikan gambaran menyeluruh mengenai jenis kesalahan klasifikasi yang dilakukan model. Misalnya:

1. jika model sering salah memprediksi ulasan negatif sebagai netral, hal itu menunjukkan kurangnya sensitivitas terhadap kata-kata yang berkonotasi negatif.
2. jika ulasan positif jarang salah prediksi, model memiliki ketepatan tinggi pada kelas positif.

		Naive Bayes Confusion Matrix		
Actual	Negatif	634	71	30
	Netral	215	506	67
	Positif	38	159	255
		Negatif	Netral	Positif
		Predicted		

Gambar 1. Confusion Matrix

2. Akurasi

Akurasi digunakan untuk menghitung persentase prediksi yang benar dari keseluruhan data uji. Rumus akurasi adalah:

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Total data uji}} \times 100\%$$

Akurasi memberikan evaluasi umum terhadap performa model, namun tidak selalu cukup ketika jumlah kelas tidak seimbang. Oleh karena itu, metrik tambahan seperti precision, recall, dan F1-score juga dihitung.

3. Precision

Precision mengukur tingkat ketepatan model dalam memprediksi suatu kelas. Precision yang tinggi berarti bahwa sebagian besar prediksi kelas tersebut benar.

$$\text{Precision} = \frac{TP}{TP + FP}$$

4. Recall

Recall mengukur kemampuan model dalam menangkap seluruh data yang seharusnya termasuk dalam suatu kelas.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Nilai recall yang tinggi menunjukkan bahwa model dapat mengenali sebagian besar ulasan yang benar-benar termasuk dalam kelas tertentu.

5. F1-Score

F1-score adalah rata-rata harmonis dari precision dan recall.

$$F1 = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-score adalah metrik yang paling seimbang untuk mengukur kualitas model karena mempertimbangkan kedua aspek: ketepatan dan kelengkapan.

Classification Report for Naive Bayes:				
	precision	recall	f1-score	support
Negatif	0.715	0.863	0.782	735.000
Netral	0.688	0.642	0.664	788.000
Positif	0.724	0.564	0.634	452.000
accuracy	0.706	0.706	0.706	0.706
macro avg	0.709	0.690	0.693	1975.000
weighted avg	0.706	0.706	0.701	1975.000

Gambar 3. Classification Report

6. Interpretasi Hasil Evaluasi

Hasil evaluasi model Naive Bayes pada dataset ulasan Duolingo menunjukkan bahwa:

1. Model memiliki performa yang stabil pada kelas positif dan netral, yang biasanya memiliki distribusi data lebih besar dibanding kelas negatif.
2. Kinerja pada kelas negatif cenderung lebih rendah, yang umum terjadi pada analisis sentimen karena kosakata negatif lebih beragam dan konteksnya lebih kompleks.
3. F1-score memberikan gambaran akurat tentang keseimbangan model dalam mengenali ketiga kelas.
4. Confusion matrix mengungkapkan jenis kesalahan utama, misalnya kecenderungan model mengklasifikasikan ulasan bernada negatif sebagai netral.

Secara keseluruhan, model Multinomial Naive Bayes menunjukkan tingkat akurasi yang baik dan dapat diandalkan untuk melakukan analisis sentimen pada ulasan pengguna aplikasi Duolingo.

4.2 Pembahasan

Bagian ini membahas secara komprehensif temuan penelitian berdasarkan seluruh tahapan proses analisis sentimen, mulai dari pengumpulan data, preprocessing, pembobotan menggunakan metode lexicon-based, hingga klasifikasi dengan algoritma Naive Bayes. Pembahasan dilakukan untuk menginterpretasikan hasil serta

mengaitkannya dengan teori, pola data, dan tujuan penelitian.

4.2.1. Kualitas Data dan Dampaknya terhadap Akurasi Model

Data yang diperoleh dari Google Play Store menunjukkan keragaman ulasan pengguna baik dari sisi panjang kalimat, struktur bahasa, maupun penggunaan kata tidak baku. Tahap preprocessing terutama cleaning, normalisasi, stopwords removal, dan stemming terbukti berperan penting dalam meningkatkan kualitas data. Penghapusan kata-kata tidak relevan dan penyamaan bentuk kata membantu mengurangi noise sehingga representasi fitur menjadi lebih stabil. Hasil ini selaras dengan literatur NLP yang menyatakan bahwa kualitas preprocessing memiliki pengaruh signifikan terhadap performa model berbasis teks. Model Naive Bayes yang digunakan pada data mentah umumnya memiliki akurasi lebih rendah dibanding data yang telah dibersihkan.

Dengan demikian, dapat disimpulkan bahwa tahap preprocessing yang sistematis berhasil memberikan fondasi yang kuat bagi model klasifikasi.

4.2.2. Efektivitas Metode Lexicon-Based sebagai Dasar Pelabelan

Pembobotan sentimen menggunakan lexicon-based memberikan gambaran awal mengenai distribusi polaritas ulasan. Pola yang ditemukan menunjukkan bahwa:

- 1) Ulasan positif paling dominan pada rating tinggi (4-5 bintang).
- 2) Ulasan negatif meningkat pada rating rendah (1-2 bintang).
- 3) Sebagian ulasan 3 bintang cenderung bercampur antara pengalaman positif dan negatif sehingga banyak masuk kategori netral.

Pendekatan lexicon-based terbukti efektif sebagai metode pelabelan awal karena:

- 1) Transparan - setiap nilai skor dapat dilacak dari kata penyusunnya.
- 2) Konsisten - kata yang sama selalu menghasilkan bobot yang sama.
- 3) Cocok untuk Bahasa Indonesia, terutama pada ulasan yang banyak menggunakan kata sifat.

Namun, metode ini juga memiliki keterbatasan, misalnya:

- 1) Tidak menangkap konteks kalimat tertentu, seperti ironi atau sarkasme.

- 2) Tidak mempertimbangkan intensitas sentimen (misalnya “bagus” dan “sangat bagus” dianggap sama).

Keterbatasan ini terlihat dari beberapa kasus di mana ulasan yang semestinya positif diberi skor netral karena tidak mengandung kata berpolaritas kuat.

4.2.3. Analisis Performa Model Naive Bayes

Model Multinomial Naive Bayes yang dibangun pada penelitian ini menunjukkan performa yang cukup baik dalam mengklasifikasikan sentimen, terutama pada kelas positif dan netral. Hal ini wajar mengingat:

- 1) Distribusi data positif jauh lebih besar dibanding data negatif.
- 2) Kata-kata positif lebih konsisten dan lebih sering muncul.
- 3) Kata negatif lebih bervariasi sehingga sulit ditangkap model.

Dari evaluasi menggunakan akurasi, precision, recall, dan F1-score, dapat disimpulkan bahwa:

- 1) Akurasi keseluruhan model tergolong baik, menunjukkan kemampuan generalisasi yang kuat.
- 2) Precision pada kelas positif tinggi, menandakan model jarang salah dalam memprediksi ulasan positif.
- 3) Recall pada kelas negatif lebih rendah, mengindikasikan sebagian ulasan negatif belum berhasil dikenali dengan tepat oleh model.

Jika dilihat dari confusion matrix, model cenderung mengelompokkan ulasan bernada negatif sebagai netral. Pola ini umum terjadi karena:

- 1) Banyak ulasan negatif menggunakan bahasa halus atau tidak langsung (“fiturnya kurang lengkap”, “agak lambat”).
- 2) Kata negatif dalam Bahasa Indonesia sangat bervariasi dan kontekstual

4.2.4. Keterkaitan Hasil Penelitian dengan Tujuan Awal

Tujuan penelitian adalah menganalisis sentimen ulasan pengguna aplikasi Duolingo serta mengukur performa model Naive Bayes dalam mengklasifikasikan ulasan tersebut. Berdasarkan hasil yang telah diperoleh:

- 1) Model berhasil mengelompokkan ulasan ke dalam tiga kategori utama: positif, netral, dan negatif.

- 2) Distribusi sentimen yang ditemukan mampu menggambarkan persepsi pengguna secara umum.
- 3) Naive Bayes terbukti menjadi algoritma yang efisien dan cukup akurat untuk dataset ulasan aplikasi mobile.

4.2.5. Implikasi Terhadap Pengembangan Aplikasi Duolingo

Hasil analisis menunjukkan beberapa kecenderungan:

- 1) Sentimen positif mendominasi, menandakan pengguna puas dengan pengalaman belajar yang ditawarkan aplikasi.
- 2) Keluhan pengguna umumnya terkait bug, error saat membuka pelajaran, keterbatasan fitur gratis, dan performa aplikasi yang kadang lambat.
- 3) Informasi ini dapat menjadi masukan bagi pengembang Duolingo dalam meningkatkan kualitas sistem, memperbaiki stabilitas fitur, dan mempertimbangkan perluasan kemampuan akses pengguna gratis.

4.2.6. Keterbatasan Penelitian

Beberapa keterbatasan yang perlu dicatat adalah:

- 1) Metode lexicon-based tidak dapat menangkap konteks kalimat secara mendalam.
- 2) Model Naive Bayes tidak mempertimbangkan urutan kata (bag-of-words assumption).
- 3) Distribusi kelas yang tidak seimbang dapat mempengaruhi performa model.
- 4) Beberapa ulasan ambigu dan sulit diklasifikasikan secara objektif.

4.2.7. Kesimpulan Pembahasan

Secara keseluruhan, pembahasan ini menunjukkan bahwa:

- 1) Proses preprocessing yang baik meningkatkan kualitas klasifikasi.
- 2) Lexicon-based efektif untuk pelabelan awal namun memiliki keterbatasan konteks.
- 3) Naive Bayes memberikan performa yang cukup baik dalam analisis sentimen ulasan Duolingo.
- 4) Hasil penelitian mampu menggambarkan persepsi pengguna secara umum, serta memberikan wawasan yang dapat digunakan untuk peningkatan aplikasi.

5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai *Sentiment Analysis of Duolingo App Reviews Using Lexicon-Based Methods and Python-Based Naïve Bayes Algorithms*, dapat disimpulkan bahwa analisis sentimen menggunakan metode *lexicon-based* mampu menggambarkan kecenderungan opini pengguna terhadap aplikasi Duolingo secara jelas, baik dalam kategori positif, netral, maupun negatif, karena metode ini memberikan penilaian polaritas secara langsung pada setiap ulasan dan efektif dalam memetakan ekspresi emosional pada data teks yang bersifat informal. Di sisi lain, algoritma Naïve Bayes berbasis Python menunjukkan performa yang baik dalam mengklasifikasikan sentimen ulasan pengguna, yang dibuktikan melalui hasil evaluasi akurasi, precision, recall, dan F1-score yang stabil, serta didukung oleh penggunaan TF-IDF dan proses preprocessing yang optimal sehingga meningkatkan kualitas pemetaan fitur teks ke dalam kategori sentimen. Selain itu, perbandingan antara kedua metode menunjukkan adanya perbedaan performa, di mana pendekatan *lexicon-based* unggul dalam hal interpretabilitas dan kemudahan penggunaan, namun cenderung sensitif terhadap variasi kata, sedangkan Naïve Bayes lebih konsisten dan akurat karena mampu mempelajari pola dari data, sehingga memberikan hasil klasifikasi yang lebih unggul dalam analisis sentimen ulasan pengguna aplikasi Duolingo.

5.2 Saran

Berdasarkan hasil penelitian dan kesimpulan yang telah diperoleh, terdapat beberapa saran untuk pengembangan penelitian maupun aplikasi Duolingo. Penelitian selanjutnya disarankan untuk mengembangkan metode analisis yang lebih kompleks dengan menggabungkan pendekatan *lexicon-based* dan metode berbasis *machine learning* atau *deep learning* seperti SVM, LSTM, maupun model berbasis transformer agar mampu menghasilkan klasifikasi sentimen yang lebih akurat dan memahami konteks linguistik secara lebih mendalam.

Selain itu, perlu dilakukan perluasan dan peningkatan kualitas data ulasan dengan mengumpulkan data dalam jumlah yang lebih besar serta rentang waktu yang lebih luas, disertai proses penyaringan terhadap spam, duplikasi, dan ulasan yang tidak informatif agar hasil analisis lebih representatif dan terpercaya. Penelitian berikutnya juga dapat menambahkan fitur *aspect-based sentiment analysis* untuk menggali pengalaman pengguna secara lebih rinci, seperti pada aspek fitur

aplikasi, kualitas pembelajaran, kemudahan penggunaan, dan performa teknis.

Di samping itu, meskipun algoritma Naïve Bayes telah menunjukkan performa yang baik, optimalisasi lebih lanjut melalui parameter tuning, penggunaan fitur n-gram, atau pendekatan ensemble perlu dilakukan, serta mempertimbangkan model alternatif seperti Logistic Regression, Random Forest, atau BERT-based classifier untuk meningkatkan akurasi dan stabilitas. Terakhir, hasil analisis sentimen ini diharapkan dapat dimanfaatkan oleh pengembang Duolingo sebagai bahan evaluasi dalam memahami kebutuhan dan keluhan pengguna secara lebih terstruktur, sehingga pengembangan fitur, peningkatan pengalaman belajar, dan kualitas layanan dapat dilakukan secara lebih terarah.

DAFTAR PUSTAKA

- [1] Akademika, "Structured preprocessing pipeline for multilingual lexicon-based sentiment analysis," *Akademika Journal*, 2022.
- [2] Anam *et al.*, "Automatic lexicon labeling to improve SVM sentiment classification," *Journal of Information Processing*, 2023.
- [3] Ayele, "Adaptation of CRISP-DM for idea mining using NLP," *International Journal of Data Mining*, 2021.
- [4] A. Brawijaya *et al.*, "Improving Naive Bayes with SMOTE," 2025.
- [5] "Deep learning based preprocessing for text mining," *Journal of Current Trends in Computer Science Research*, 2023.
- [6] Kusumaningrum, Nugroho, and Prabowo, "Deep learning-based sentiment analysis on Indonesian hotel reviews," *Journal of Information Systems*, 2023.
- [7] T. Nguyen *et al.*, "Comparison of language-learning apps using sentiment analysis," 2024.
- [8] Qi and Shabrina, "Lexicon versus machine learning for short text sentiment," *Journal of Computational Social Science*, 2023.
- [9] L. Wang and K. Jaber, "Contextual models for sentiment detection," 2025.
- [10] T. Wijaya *et al.*, "Fine-tuning BERT for app reviews sentiment analysis," 2024.