

Implementasi analisis sentimen pada ulasan aplikasi Duolingo di Google Playstore menggunakan algoritma Naïve Bayes

**Meli Apriliyani ¹⁾, Mirza Izzal Musyaffaq ¹⁾, Siti Nur'Aini ¹⁾,
Maya Rini Handayani ¹⁾, Khotibul Umam ²⁾**

^{1,2)}Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi,
UIN Walisongo Semarang

Jl. Prof. Dr. Hamka, Semarang 50185, Jawa Tengah, Indonesia.

Email: ²⁾khothibul_umam@walisongo.ac.id

Riwayat Artikel

Diterima:
25-04-2024

Direvisi:
27-05-2024

Disetujui:
28-05-2024

Abstract

This study investigates sentiment analysis to evaluate the Duolingo Application using the Naive Bayes method. The Duolingo program exemplifies the use of big data technology for processing vast and complex data. Google Play Store offers review and rating functions that can help with program development and fixing undesirable aspects. This project uses sentiment analysis techniques to automatically analyze Indonesian internet product reviews and obtain information about the feelings expressed in these reviews. The Naïve Bayes method is used to classify reviews as positive or negative. The research findings show that a dataset consisting of 1000 pieces of data originating from reviews of the Duolingo program on the Google Play Store was manually labeled before the preprocessing step. Of this number, 500 data have positive sentiments, while 500 have negative attitudes. Additionally, sentiment analysis shows an accuracy rate of 86 persen. The f1 score precision is 89 persen and recall 83 persen values, with f1 results for classification of 86 persen.

Keywords: Duolingo App, Naïve Bayes, sentiment analysis

Abstrak

Penelitian ini menyelidiki analisis sentimen evaluasi Aplikasi Duolingo menggunakan metode Naive Bayes. Program Duolingo mencontohkan penggunaan teknologi data besar untuk pemrosesan data yang luas dan rumit. Google Play Store menawarkan fungsi peninjauan dan pemeringkatan yang dapat membantu pengembangan program dan perbaikan aspek yang tidak diinginkan. Proyek ini menggunakan teknik analisis sentimen yang secara otomatis menganalisis ulasan produk internet Indonesia dan mendapatkan informasi mengenai perasaan yang diungkapkan dalam ulasan tersebut. Metode Naïve Bayes digunakan untuk menentukan klasifikasi ulasan menjadi positif atau negatif. Temuan penelitian menunjukkan bahwa kumpulan data yang terdiri dari 1000 data yang berasal dari ulasan program Duolingo di Google Play Store diberi label secara manual sebelum ke langkah prapemrosesan. Dari jumlah tersebut, 500 data memiliki sentimen positif, sedangkan 500 data memiliki sikap negatif. Selain

itu, analisis sentimen menunjukkan tingkat akurasi sebesar 86 persen. Skor f1 menunjukkan nilai presisi 89 persen dan recall 83 persen, dengan hasil f1 pada klasifikasi sebesar 86 persen.

Kata kunci: Aplikasi Duolingo, *Naïve Bayes*, Analisis Sentimen

Pendahuluan

Pesatnya kemajuan teknologi telah memberikan banyak manfaat bagi masyarakat di berbagai bidang kehidupan, memudahkan tugas sehari-hari, meningkatkan produktivitas, dan memperluas akses terhadap informasi dan layanan. Kemajuan teknologi yang pesat menghasilkan data dalam jumlah besar, yang dapat menghasilkan wawasan berharga jika ditangani dan dimanfaatkan secara efektif. Seiring dengan meningkatnya jumlah pengguna internet aktif, volume data yang dihasilkan juga meningkat. Teknologi *big data* memfasilitasi analisis sejumlah besar data yang rumit dan beragam, sehingga menghasilkan pengetahuan yang berharga [1].

Aplikasi Duolingo mencontohkan kemampuan teknologi dalam menangani kumpulan data yang luas dan rumit. Aplikasi Duolingo adalah program tanpa biaya yang dikembangkan oleh Luis Von Ahn dan Severin Hacker pada bulan November 2011. Tagline organisasi ini adalah “Menyediakan pendidikan bahasa gratis kepada orang-orang di seluruh dunia.” Situs web perusahaan mengklaim bahwa mereka memiliki lebih dari 30 juta orang yang telah mendaftar [2]. Aplikasi Duolingo mendapat rating terpuji menurut statistik yang diperoleh dari Google Playstore. Mengingat banyaknya ulasan pengguna untuk aplikasi Duolingo di Google Playstore, sulit untuk memastikan proporsi pasti dari tanggapan baik atau negatif. Memanfaatkan teknologi big data, aplikasi Duolingo memiliki kemampuan untuk menganalisis kumpulan data yang luas dan rumit, termasuk data penggunaan aplikasi, data penggunaan kata, dan data penggunaan fitur. Dengan menggunakan data ini, aplikasi Duolingo dapat menghasilkan wawasan berharga, seperti frekuensi penggunaan kata, pola penggunaan berbagai fitur, dan pola penggunaan aplikasi secara keseluruhan. Dengan memanfaatkan informasi ini, pengguna dapat meningkatkan pemanfaatan program, meningkatkan pengalaman belajar bahasa, dan mengoptimalkan pemanfaatan kemampuan aplikasi.

Play Store adalah layanan yang disediakan Google dan menyediakan berbagai macam materi digital termasuk permainan, aplikasi, film, musik, dan buku, yang disusun dalam beberapa kategori berbeda [3]. Play Store memiliki fungsi penilaian dan ulasan yang memungkinkan pelanggan memberikan komentar dan penilaian terhadap barang-barang yang telah mereka gunakan. Melalui penggunaan kemampuan ini pengguna dapat berkontribusi pada peningkatan aplikasi dan memperbaiki fungsionalitas yang tidak berfungsi. Hal ini memudahkan konsumen untuk memilih program yang sesuai dengan preferensi dan kebutuhan spesifik mereka. Ada lebih dari 500 juta unduhan aplikasi Duolingo yang tersedia di YouTube Play Store. Melalui

pemanfaatan pendekatan analisis *centime*, tujuan penelitian ini adalah untuk melakukan investigasi mendalam terhadap program Duolingo.

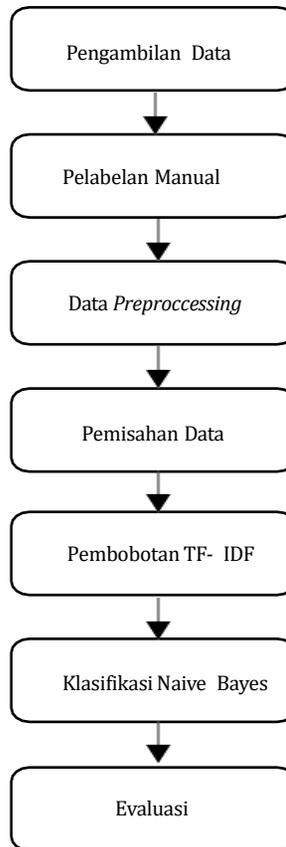
Analisis sentimen adalah salah satu pendekatan yang dapat diambil untuk mengatasi masalah pengkategorian evaluasi secara otomatis menjadi lebih positif atau lebih negatif [4]. Analisis sentimen sering digunakan dalam pemantauan media sosial untuk mendeteksi sentimen yang ada, termasuk opini positif dan negatif[5]. Tujuan dari penelitian ini adalah menggunakan metode Naïve Bayes untuk melakukan analisis sentimen pada evaluasi evaluasi aplikasi Duolingo. Namun demikian, penting untuk digarisbawahi bahwa hanya sejumlah kecil penelitian yang melakukan analisis sentimen pada program Duolingo. Tujuan dari penelitian ini adalah untuk memberikan penjelasan mengenai pendekatan analisis sentimen yang menggunakan algoritma Naïve Bayes, yaitu teknik *text mining* yang diimplementasikan dalam bahasa komputer Python. Untuk mengklasifikasikan ulasan sebagai baik atau negatif, metode Naïve Bayes akan digunakan. Pendekatan Naïve Bayes dipilih karena penerapannya yang luas dalam penambangan teks dan akurasinya yang luar biasa. Menerapkan pendekatan Naïve Bayes dapat mengurangi durasi yang diperlukan untuk kategorisasi analisis sentimen [6].

Penelitian sebelumnya yang membahas tentang analisis sentimen dan menggunakan data pada Google Play Store yaitu penelitian tentang *App Review in Google Play Store*. Penelitian ini menemukan bahwa Naïve Bayes dapat membantu memecahkan masalah pengklasifikasian sentimen pada ulasan aplikasi Shopee dengan akurasi yang tinggi sebesar 96,667 persen [7]. Penelitian lain yang membahas analisis sentimen menggunakan data Google Play adalah untuk menganalisis sentimen aplikasi BRIMO. Penelitian ini tidak menjelaskan kecenderungan sentimen pada pengguna terhadap topik yang dibahas dengan menghasilkan akurasi yang cukup tinggi yaitu 84,25 persen [8]. Kedua penelitian ini memiliki kesamaan dalam penggunaan algoritma dan hanya fokus pada pengkalsifikasian ulasan dan akurasi metodenya tidak menjelaskan kecenderungan sentimen pengguna terhadap topik yang dibahas.

Metode Penelitian

Penelitian ini fokus pada pemeriksaan penilaian pengguna terhadap program Duolingo di Google Playstore. Dengan menggunakan Algoritma Naïve Bayes untuk memprediksi suatu ulasan apakah positif atau negatif, dengan kemampuan yang dapat mengatasi data dengan dimensi yang tinggi [9]. Algoritma Naïve Bayes dipilih karena sangat akurat dan banyak digunakan dalam proses *text mining* [10]. Selain itu, algoritma Naïve Bayes adalah pengklasifikasi probabilistik yang paling sederhana dan mudah digunakan [11]. Metode penelitian ini terdiri dari tujuh tahapan penting: pengumpulan data, anotasi manual, pemrosesan awal data, pembobotan TF-IDF, partisi data ke dalam set pelatihan dan pengujian, klasifikasi menggunakan algoritma *Naïve Bayes*, dan

evaluasi. Dengan mengikuti proses ini secara sistematis, penelitian ini dapat menghasilkan hasil yang reliabel dan bermanfaat dalam memahami persepsi pengguna terhadap Aplikasi Duolingo. Gambar 1 menunjukkan alur penelitian.



Gambar 1 Alur penelitian

Pengumpulan Data. Proses pengumpulan data dilakukan melalui bahasa pemrograman Python dan *scraping tool* Google Play Store. Data yang diambil adalah 1000 data mulai *rating* bintang 1 sampai 5. Dengan menggunakan teknik *scraping*, data ulasan aplikasi Duolingo di Google Playstore diambil secara Otomatis. data yang dikumpulkan termasuk evaluasi pengguna, berupa nilai Bintang dan ulasan. Data yang telah dikumpulkan akan disimpan dalam *database* untuk penggunaan kasus yang akan datang.

Pelabelan. Setelah data tinjauan diperoleh, tugas pemberian label positif atau negatif menggunakan *library* NLTK. Setelah proses pelabelan selesai, data ulasan dibagi menjadi dua kategori berbeda: data ulasan positif yang mencakup ulasan dengan *rating* Bintang 4 atau 5 dan data ulasan negatif yang terdiri dari ulasan dengan *rating* bintang 1, 2, atau 3. Pembagian ini dilakukan untuk menyederhanakan prosedur analisis sentimen berikutnya dan memungkinkan fokus pada ulasan yang paling relevan dalam masing-masing kategori sentimen. Dengan demikian, analisis sentimen

dapat dilakukan dengan lebih efisien dan benar untuk lebih memahami opini pengguna Duolingo.

Data Preprocessing. Untuk melakukan analisis sentimen penting untuk melalui langkah *pretreatment* data terlebih dahulu. Pada langkah pra-pemrosesan dilakukan pemilihan data untuk menyempurnakan struktur data yang akan digunakan [12]. Prosedur penyiapan data terdiri dari empat tahapan yaitu *casefolding*, *tokenizing*, *stopwords*, dan *stemming*.

Case folding merupakan tahap pengolahan data atau mengkonversi semua huruf menjadi huruf kecil [13]. Proses ini membantu dalam menjaga konsistensi dalam teks, sehingga tidak ada perbedaan yang tidak perlu antara kata yang ditulis dalam huruf besar atau kecil. *Stopwords* adalah proses menghilangkan kata-kata yang tidak bermakna dari teks yang sedang diproses[14]. Kata *stopwords* adalah kata umum yang tidak memberikan nilai tambah pada analisis sentimen, contohnya “saya”, “dan”, “atau”, “di”, dan lainnya. Penghapusan kata *stopwords* membantu dalam focus pada kata-kata kunci yang lebih relevan dalam menentukan sentiment. *Tokenizing* adalah langkah di mana kalimat dipecah menjadi beberapa kata yang disebut sebagai token yang memisahkannya [15]. Teks dibagi menjadi unit-unit yang lebih kecil dengan prosedur ini, yang membuat analisis di masa depan lebih mudah dilakukan. *Stemming* terdiri dari proses menghilangkan kata-kata yang mempunyai imbuhan untuk diubah menjadi kata-kata sederhana [16]. Proses ini membantu dalam mengurangi variasi kata yang sebenarnya memiliki arti sama, sehingga memperkuat analisis sentimen dengan memperlakukan kata-kata tersebut secara konsisten. Dengan melalui tahapan-tahapan tersebut proses *preprocessing* dapat dipastikan data siap diolah lebih lanjut dalam analisis sentimen, sehingga dapat menghasilkan hasil yang lebih akurat.

Pemisahan data. Pemisahan data yang juga dikenal sebagai produksi data pelatihan dan data pengujian dilakukan untuk membagi data pra-pemrosesan menjadi subset tertentu untuk tujuan perbandingan. Meskipun tujuan utamanya adalah memperoleh kumpulan data yang cukup representatif untuk tujuan pelatihan model, kumpulan data pengujian digunakan untuk tujuan mengevaluasi efektivitas model yang dilatih. Dengan melakukan hal ini, dapat dinilai sejauh mana model menunjukkan kemampuan generalisasi terhadap data baru yang sebelumnya tidak teramati. Memastikan perbandingan yang tepat antara data pelatihan dan data pengujian sangat penting untuk menjamin bahwa model dapat menghasilkan hasil yang memuaskan dalam skenario dunia nyata.

Pembobotan TF-IDF. Informasi yang telah diproses sebelumnya, yang mungkin berupa kata-kata, akan diubah menjadi representasi numerik dengan menggunakan prosedur pembobotan kata. Teknik pembobotan kata berupaya untuk menentukan proporsi kepentingan setiap kata yang dijadikan ciri dalam penelitian. Frekuensi suatu

istilah dalam teks olahan berkorelasi langsung dengan bobotnya. Akibatnya, ketika jumlah dokumen yang diproses bertambah, semakin banyak pula karakteristik yang dihasilkan. Hal ini memungkinkan pemeriksaan yang lebih komprehensif dan berpotensi meningkatkan kaliber dan presisi temuan analitis yang dilakukan. Perhitungan bobot kata pada pendekatan TF-IDF merupakan hasil penggabungan dua gagasan yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) [17]. Salah satu definisi TF adalah berapa kali setiap kata muncul dalam dokumen tertentu. Nilai TF meningkat sebanding dengan jumlah kata yang ada di setiap dokumen dan sering muncul.

Klasifikasi Naïve Bayes. Proses klasifikasi merupakan suatu metode penambangan data yang dapat digunakan untuk membangun suatu model hingga semua data termasuk dalam kategori yang sama [18]. Teorema Bayes berfungsi sebagai premis fundamental yang digunakan dalam pendekatan klasifikasi yang digunakan dalam algoritma Naïve Bayes. Teorema ini adalah yang pertama. Seorang ilmuwan Inggris bernama Thomas Bayes lah yang menemukan Naïve Bayes [19]. Teknik Naïve Bayes sering digunakan oleh beberapa akademisi dalam analisis sentimen. Klasifikasi Naïve Bayes memprediksi kemungkinan kepemilikan suatu kelas dengan mengasumsikan bahwa karakteristik yang digunakan tidak tergantung satu sama lain. Metode Naïve Bayes memiliki keunggulan efisiensi, memungkinkan prosedur analisis sentimen lebih cepat dan efisien. Hal lebih lanjut yang perlu dipertimbangkan adalah bahwa algoritma Naïve Bayes memiliki kecenderungan untuk memperoleh akurasi tinggi dengan jumlah data pelatihan yang terbatas.

Evaluasi. Evaluasi merupakan fase yang mengukur efektivitas algoritma kategorisasi yang digunakan dalam penelitian. Kinerja sistem dinilai menggunakan ukuran seperti akurasi, presisi, dan perolehan. Untuk melakukan komputasi menggunakan benchmark ini, penting untuk menggunakan metodologi *Confusion Matrix* yang terdiri dari komponen *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Adapun rumus *Confusion Matrix* untuk menghitung akurasi, presisi, dan *recall* adalah seperti yang terlihat pada Persamaan (1) sampai (3).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

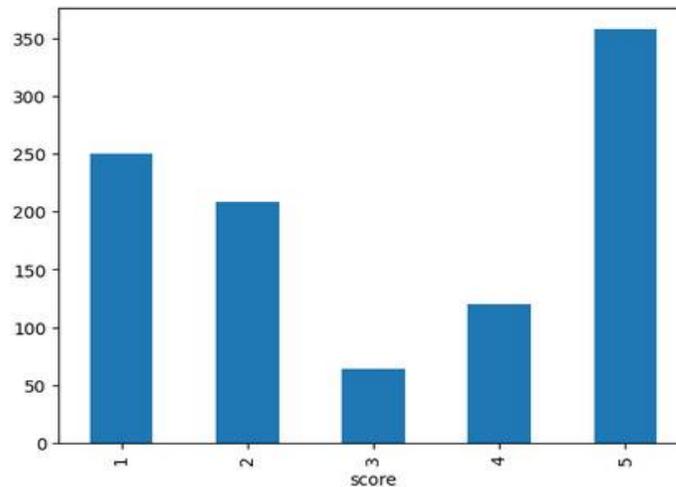
$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Hasil dan Pembahasan

Pada tahap ini akan dijelaskan tahapan yang berkaitan dengan prosedur yang dilakukan dalam penelitian, sebagaimana dituangkan dalam metodologi penelitian

sebelumnya.

Tahap pertama adalah tahap pengambilan data. Pengambilan data dilakukan dengan memanfaatkan URL program Duolingo dari Playstore dimulai bulan Maret 2024. Gambar 2 dan Gambar 3 menyajikan hasil ekstraksi data yang dilakukan dengan menggunakan kata kunci “Duolingo” dan besarnya pencarian data dari 1000 kemunculan aplikasi Duolingo.



Gambar 2 Data mentah ulasan positif dan negatif

```
score
5    358
1    250
2    208
4    120
3     64
Name: count, dtype: int64
```

Gambar 3 Hasil output *scraping*

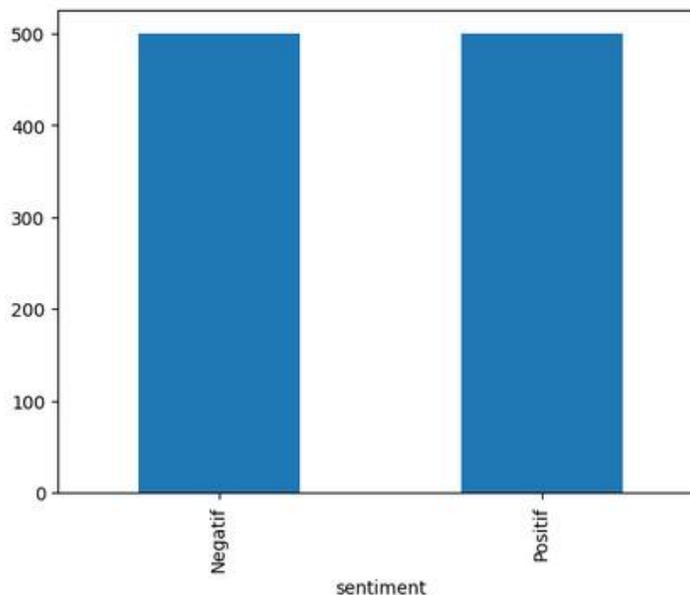
Setelah melakukan ekstraksi data, hasilnya menunjukkan bahwa aplikasi Duolingo memiliki beragam ulasan dari pengguna yang berbeda. Dari ulasan yang yang diberi bintang 1, 2, 3, 4 dan 5.

Setelah mengumpulkan data penelitian, dilanjutkan dengan melakukan pelabelan yang akan menggunakan *library* NLTK. NLTK (*Natural Language Toolkit*) adalah sebuah perpustakaan untuk pemrosesan bahasa alami Python yang dikembangkan oleh Stevan Bird dan Edward Loper [20]. Gambar 4 adalah hasil dari prosedur pelabelan yang telah dilakukan.

```
import nltk
from textblob import TextBlob
all_polarity = 0;
status=[]
total_positif = total_negatif = total = 0

for ulasan in data['content']:
    analysis = TextBlob(ulasan)
    all_polarity += analysis.polarity
    if (analysis.sentiment.polarity > 0.0):
        total_positif += 1
        status.append("Positif")
    elif (analysis.sentiment.polarity <= 0.0):
        total_negatif += 1
        status.append("Negatif")
    total += 1
print(f"Hasil Analisis Data: \nPositif = {total_positif} \nNegatif = {total_negatif}")
print(f"Total Data = {total}")
```

Gambar 4 Pelabelan NLTK



Gambar 5 Hasil Pelabelan NLTK

Hasil dari prosedur pelabelan NLTK ini memberikan wawasan penting tentang cara pengguna memandang platform pembelajaran bahasa online Duolingo. Dengan membedakan ulasan menjadi kategori positif dan negatif. Pada analisis pada data ini menghasilkan analisis data positif sebanyak 500 dan data negatif sebanyak 500.

Tahap selanjutnya adalah tahap *Preprocessing* yang merupakan langkah krusial yang harus diselesaikan sebelum menggunakan algoritma klasifikasi pada dokumen. Tahapan pra-pemrosesan adalah sebagai berikut:

Case Folding. Untuk menjamin keseluruhan teks yang dihasilkan konsisten, langkah ini dilakukan untuk mengubah semua huruf menjadi huruf kecil. Selama

prosedur ini, kata-kata yang tidak diperlukan dibersihkan dari tanda baca atau karakter tertentu. Tabel 1 merupakan hasil saat menggunakan *case folding*. Hasil dari *case folding*, semua huruf menjadi huruf kecil. Hal ini memberikan keseragaman dalam teks dan mempermudah analisis selanjutnya.

Tabel 1 Hasil *case folding*

<i>Score</i>	<i>Content</i>	<i>At</i>	<i>Content clean</i>
3	This is the one. Ive tried several apps and pr...	2024-05-13 03:46:51	this is the one ive tried several apps and pro..
4	It has a decent variety of exercises. it does ...	2024-05-13 10:02:02	it has a decent variety of exercises it does n...
3	Most important it's free, you can upgrade but ...	2024-05-09 19:21:19	most important it s free you can upgrade but i...
2	I really like the setup and the levels. I felt...	2024-05-04 11:01:20	i really like the setup and the levels i felt ...
4	Fantastic app designed to keep you engaged wit...	2024-05-10 03:38:26	fantastic app designed to keep you engaged wit...

Stopwords. Selanjutnya adalah tahapan *stopwords*, *Stopwords* merupakan tahap yang penting juga yaitu penghapusan kata- kata yang tidak penting dan tidak bermakna berdasarkan kata yang terdapat pada database *stopword*, seperti “dan”, “aku”, “dari”, dan lain-lain. Tabel 2 merupakan hasil dari *stopwords*. Penekanan pada kata-kata kunci yang lebih relevan dalam menentukan sentimen difasilitasi oleh hasil-hasil ini. Dengan demikian, proses ini membantu dalam meningkatkan kualitas analisis sentimen dengan membuang informasi yang tidak relevan, sehingga memastikan bahwa hasil analisis lebih akurat.

Tabel 2 Hasil *stopwords*

<i>Content clean</i>	<i>Text stopwords</i>
this is the one ive tried several apps and pro..	this is the one ive tried several apps and pro..
it has a decent variety of exercises it does n...	it has a decent variety of exercises it does n...
most important it s free you can upgrade but i...	most important it s free you can upgrade but i...
i really like the setup and the levels i felt ...	i really like the setup and the levels i felt ...
fantastic app designed to keep you engaged wit...	fantastic app designed to keep you engaged wit...

Tokenizing. Fase selanjutnya adalah *tokenizing*, yang melibatkan pembagian teks menjadi kata-kata individual untuk memfasilitasi proses *stemming* selanjutnya. Pada tahap ini teks ulasan telah dibagi menjadi kata-kata individual, yang memungkinkan untuk melakukan analisis lebih lanjut secara terpisah pada setiap kata. Tabel 3 merupakan hasil *tokenizing*. Dengan melalui tahap *tokenizing*, data siap untuk diproses lebih lanjut dalam analisis sentimen menggunakan algoritma klasifikasi.

Tabel 3 Hasil *tokenizing*

<i>Text_stopword</i>	<i>Text_tokens</i>
this is the one ive tried several apps and pro...	[this, is, the, one, ive, tried, several, apps...
it has a decent variety of exercises it does n...	[it, has, a, decent, variety, of, exercises, i...
most important it s free you can upgrade but i...	[most, important, it, s, free, you, can, upgra..
i really like the setup and the levels i felt ...	[i, really, like, the, setup, and, the, levels...
fantastic app designed to keep you engaged wit...	[fantastic, app, designed, to, keep, you, enga...

Stemming. *Stemming* merupakan suatu metode yang digunakan untuk menghilangkan imbuhan, preposisi, kata ganti, dan kata keterangan sesuai dengan standar yang ditetapkan oleh KBBI [21]. Prosedur *stemming* sastra Indonesia ini dilakukan dengan bantuan Perpustakaan Sastrawi untuk kepentingan penelitian ini. Ilustrasi hasil dari metode *stemming* dapat dilihat pada Gambar 6. Proses ini membantu dalam mengurangi variasi kata yang sebenarnya memiliki arti yang sama, sehingga memperkuat analisis sentimen dengan memperlakukan kata-kata tersebut secara konsisten.

```
-----  
1 : this : this  
2 : is : is  
3 : the : the  
4 : one : one  
5 : ive : ive  
6 : tried : tried  
7 : several : several  
8 : apps : apps  
9 : and : and  
10 : programs : programs  
11 : to : to  
12 : help : help  
13 : me : me  
14 : learn : learn  
15 : japanese : japanese  
16 : but : but  
17 : that : that
```

Gambar 6 Hasil output *stemming*

Tahap kedua adalah tahap pemisahan data. Saat ini, data dibagi menjadi dua kategori berbeda: data pengujian dan data pelatihan. Data dipartisi menjadi 20 persen dataset pengujian dan 80 persen dataset pelatihan dalam penelitian ini. Dengan membagi dataset menjadi 20 persen untuk pengujian dan 80 persen untuk pelatihan, penelitian ini memastikan bahwa model yang dibangun dapat diuji dengan baik pada data yang tidak terlihat sebelumnya, sementara masih memiliki sejumlah besar data untuk melatih model dengan baik. Hal ini memungkinkan untuk mendapatkan evaluasi yang lebih obyektif terhadap kinerja model dalam memprediksi sentimen ulasan aplikasi Duolingo.

Tahap ketiga adalah tahap pembobotan TF-IDF. Saat ini, pendekatan pembobotan TF-IDF (*Term Frequency-Invers Document Frequency*) digunakan untuk mengubah data tekstual ke dalam format numerik. Kumpulan data memberikan bobot pada setiap kata berdasarkan frekuensinya dalam teks (TF) dan kelangkaannya di seluruh dokumen dalam kumpulan data (IDF). Dengan menggunakan pembobotan TF-IDF, diperoleh numerik yang lebih akurat dari teks ulasan, yang dapat digunakan oleh algoritma klasifikasi untuk analisis sentimen.

Tahap keempat adalah tahap klasifikasi dengan Naive Bayes. Setelah pembobotan TF-IDF diterapkan, dengan menggunakan algoritma klasifikasi Naive Bayes, komputer akan menggunakan data tersebut untuk tujuan pelatihan. Komputer akan menganalisis pola pada data pelatihan untuk membuat prediksi pada data baru berdasarkan pola yang diidentifikasi.

Tahap kelima adalah tahap evaluasi. Selanjutnya, tahap penilaian menjadi sangat penting dalam menilai kemandirian algoritma klasifikasi yang digunakan dalam penelitian. Evaluasi kinerja suatu algoritma dapat dilakukan dengan pemanfaatan metrik seperti akurasi, presisi, dan perolehan. Pendekatan yang dikenal sebagai matriks konfusi digunakan untuk menghitung ukuran ini. Pendekatan ini terdiri dari komponen-komponen berikut: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Temuan keluaran penilaian terlihat pada Gambar 9.

```
MultinomialNB Accuracy: 0.86
MultinomialNB Precision: 0.8854166666666666
MultinomialNB Recall: 0.8333333333333334
MultinomialNB f1_score: 0.8585858585858586
confusion_matrix:
[[85 17]
 [11 87]]
=====
              precision    recall  f1-score   support

 Negatif         0.89         0.83         0.86         102
 Positif         0.84         0.89         0.86          98

 accuracy              0.86              0.86              0.86         200
 macro avg              0.86              0.86              0.86         200
 weighted avg           0.86              0.86              0.86         200
```

Gambar 9 Hasil output evaluasi

Formulasi *Confusion Matrix* untuk menghitung akurasi, presisi, dan *recall* adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = \frac{87 + 85}{87 + 11 + 85 + 17} = \frac{172}{200} = 0,86$$

$$Precision = \frac{TP}{TP + FP} = \frac{87}{87 + 11} = \frac{87}{98} = 0,89$$

$$Recall = \frac{TP}{TP + FN} = \frac{87}{87 + 17} = \frac{87}{104} = 0,83$$

dimana TP memprediksi kata positif dan itu benar, TN memprediksi kata negatif dan itu benar, FP memprediksi kata positif dan itu salah, dan FN memprediksi kata itu negatif dan itu salah.

Analisis sentimen evaluasi Aplikasi Duolingo diklasifikasikan menjadi sentimen positif dan negatif dengan menggunakan metode akurasi. Temuannya menunjukkan akurasi sebesar 0,86 atau 86 persen. Apabila dinyatakan seperti pada Gambar 9, sikap negatif menghasilkan skor presisi sebesar 89 persen dan hasil *recall* sebesar 83 persen. Skor *f1* digunakan untuk menilai kualitas akurasi dan nilai *recall*. Kategorisasi tersebut menghasilkan skor *f1* sebesar 86 persen. Mengenai hasil sentimen positif, tingkat akurasinya adalah 84 persen dan tingkat *recall*-nya adalah 89 persen. Skor *f1* digunakan untuk menilai kualitas akurasi dan nilai *recall*. Kategorisasi tersebut menghasilkan skor *f1* sebesar 86 persen.

Simpulan

Berdasarkan penelitian data yang dikumpulkan terdiri dari 1000 data yang diambil dari evaluasi program Duolingo di Google Play Store, seperti yang ditunjukkan oleh temuan penelitian. Sebelum tahap *pre-processing*, pelabelan manual mengungkapkan bahwa ulasan positif lebih banyak daripada ulasan negatif, dengan 500 data menunjukkan sentimen positif dan 500 data menunjukkan sentimen negatif. Selain itu, analisis sentimen mencapai tingkat akurasi sebesar 86 persen. Adapun saran untuk penelitian berikutnya adalah penggunaan algoritma lain untuk mendapatkan hasil perbandingan yang lebih baik, serta menerapkan analisis sentimen dengan menggunakan *tools* seperti *machine learning* atau yang lainnya.

Daftar Pustaka

- [1] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [2] S. Chohan, A. Nugroho, A. M. B. Aji, and W. Gata, "Analisis Sentimen Pengguna Aplikasi Duolingo Menggunakan Metode Naive Bayes dan Synthetic Minority Over Sampling Technique," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 139–144, 2020, doi: 10.31294/p.v22i2.8251.
- [3] A. Nurian, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naive Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: 10.23960/jitet.v11i3s1.3348.
- [4] E. Indrayuni, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks

- Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 29–36, 2019, doi: 10.31294/jki.v7i1.1.
- [5] A. K. Dewi, “Analisis Sentimen Ekspedisi Sicepat Dari Ulasan Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 796–805, 2022, doi: 10.35957/jatisi.v9i2.1802.
- [6] L. Ardiani, H. Sujaini, and T. Tursina, “Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak,” *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [7] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, “App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm,” *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012043.
- [8] M. K. Khoirul Insan, U. Hayati, and O. Nurdiawan, “Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373.
- [9] M. Birjali, M. Kasri, and A. Beni-Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends,” *Knowledge-Based Syst.*, vol. 226, p. 107134, 2021, doi: 10.1016/j.knosys.2021.107134.
- [10] I. Wickramasinghe and H. Kalutarage, “Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation,” *Soft Comput.*, vol. 25, no. 3, pp. 2277–2293, 2021, doi: 10.1007/s00500-020-05297-6.
- [11] P. Kaviani and S. Dhotre, “International Journal of Advance Engineering and Research Short Survey on Naive Bayes Algorithm,” *Int. J. Adv. Eng. Res. Dev.*, vol. 4, no. 11, pp. 607–611, 2017.
- [12] N. A. Rakhmawati, M. I. Aditama, R. I. Pratama, and K. H. U. Wiwaha, “Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19,” *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 2, pp. 90–92, 2020, doi: 10.26740/jieet.v4n2.p90-92.
- [13] S. Lestari and S. Saepudin, “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes,” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, pp. 163–170, 2021.
- [14] A. Imron, “Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier,” *Tek. Inform.*, pp. 10–13, 2019, [Online]. Available: <https://dspace.uii.ac.id/handle/123456789/14268>

-
- [15] G. Sanjaya and K. M. Lhaksana, "Lexicon Based),," vol. 7, no. 3, pp. 9698–9710, 2020.
- [16] F. Bei and S. Sudin, "Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm)," *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [17] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [18] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.
- [19] R. R. Syawal, L. Pratama, T. Wahyuni, and D. G. Passa, "Klasifikasi kualitas Ikan Nilem Berdasarkan Ukuran Menggunakan Algoritma Naive Bayes," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 7, no. 2, p. 72, 2023, doi: 10.51213/jimp.v7i2.514.
- [20] A. Nuzulia, "濟無No Title No Title No Title," *Angew. Chemie Int. Ed. 6(11)*, 951–952., no. 2015, pp. 5–24, 1967.
- [21] H. K. Wardana, I. Swanita, and B. W. Yohanes, "Sistem Pemeriksa Pola Kalimat Bahasa Indonesia berbasis Algoritme Left-Corner Parsing dengan Stemming," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 3, p. 211, 2019, doi: 10.22146/jnteti.v8i3.515.