

# ANALISIS SENTIMEN PADA ULASAN PENGGUNA TERHADAP APLIKASI PEMBELAJARAN BAHASA DUOLINGO: MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR

Depro Winoto, Veven Desta Aditia, Cinda Sorisa, Ressa Priskila, Viktor Handrianus Pranatawijaya

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya

Jalan Yos Sudarso, Palangka Raya, Indonesia

deprowinoto3690@gmail.com

## ABSTRAK

Teknologi telah mengubah cara belajar bahasa, dengan populernya aplikasi seperti Duolingo di era digital. Studi ini diharapkan mampu untuk menganalisis reaksi pengguna terhadap aplikasi Duolingo, serta membandingkan keefektifan dan keefisienan algoritma yang berbeda antara *Naive Bayes* dengan *K-Nearest Neighbors* berdasarkan review pada *Google Play Store*. Data dikumpulkan menggunakan teknik *web scraping*, kemudian dilakukan serangkaian prosedur *pre-processing*. Setelah diproses, ulasan tersebut secara otomatis diklasifikasikan sebagai positif atau negatif menggunakan algoritma *Naive Bayes* dan *K-Nearest Neighbors*. Berdasarkan hasil analisis, *Naive Bayes* (NB) menunjukkan peningkatan akurasi yang lebih baik dari *K-Nearest Neighbors* (KNN), dengan akurasi 92%, presisi 93%, dan recall 92% untuk *Naive Bayes* (NB) dan akurasi 88%, presisi 91%, dan recall 90% untuk *K-Nearest Neighbors* (KNN) berdasarkan penelitian yang telah dilakukan.

**Kata kunci:** Sentimen Analisis, Naive Bayes, K-Nearest Neighbors, Aplikasi, Pembelajaran Bahasa

## 1. PENDAHULUAN

Pada era digital saat ini, terdapat banyak cara untuk belajar bahasa, salah satunya dapat diakses melalui aplikasi seperti Duolingo. Aplikasi ini telah di install oleh lebih dari 500 juta pengguna di belahan dunia dan menawarkan kursus bahasa gratis di semua tingkat level. Duolingo menawarkan pengalaman belajar yang menghibur dan menarik dengan metode permainan dan kuisnya. Teknologi telah berevolusi untuk memenuhi kebutuhan manusia dan memfasilitasi berbagai aktivitas. Peran teknologi menjadi dibutuhkan dalam menghadapi era digital masa sekarang [1].

Dalam konteks kebahasaan, Duolingo menjadi salah satu contoh sukses dalam mengintegrasikan AI untuk memfasilitasi belajar mandiri. Developer aplikasi ini mengatakan bahwa integrasi dengan AI berperan penting dalam memahami dan mengidentifikasi pengguna [2].

Aplikasi pastilah mempunyai kelebihan maupun kekurangannya masing-masing. Meskipun Duolingo dikenal secara luas, sangat penting untuk memahami bagaimana pengguna menggunakannya untuk memahami keinginan dan harapan mereka terhadap aplikasi tersebut. Masukan dari pengguna dapat memberikan informasi yang berharga bagi pengembang Duolingo untuk meningkatkan kualitas layanan yang mereka tawarkan. Media sosial telah menjadi cara penting bagi pengguna untuk mengekspresikan kepuasan, ketidakpuasan, dan pendapat mereka tentang aplikasi. Hal ini dapat digunakan sebagai subjek untuk menganalisis opini pengguna terhadap aplikasi Duolingo [3].

Tujuan utama pada penelitian ini adalah agar mengetahui review pengguna pada aplikasi Duolingo dengan menerapkan dua teknik klasifikasi, yaitu *Naive*

*Bayes* dan *K-Nearest Neighbors*. Tujuannya adalah untuk menentukan pendapat pengguna sebagai positif atau negatif, dan untuk dapat membandingkan kedua metode tersebut dengan data yang dikumpulkan dari *Google Play Store* selama satu tahun terakhir.

## 2. TINJAUAN PUSTAKA

### 2.1. Penelitian Terdahulu

Penelitian sebelumnya yang dilakukan oleh Alifiana dkk.[4] melakukan sentimen analisis pada aplikasi duolingo menggunakan algoritma *naive bayes* dan *support machine learning*, penelitian tersebut mengumpulkan data dengan menggunakan teknik *web scrapping* di *google colab* pada bulan Juni 2023 yang menghasilkan ulasan sejumlah 1500 data. Tujuan dari penelitian tersebut adalah untuk memahami arah emosi yang dominan serta membandingkan tingkat akurasi kedua algoritma tersebut. Algoritma *Naive Bayes* memperoleh nilai *accuracy* sebesar 81%, *precision* 80%, *recall* 98%, dan *f1-score* 88%. Sedangkan metode Support Vector Machine menghasilkan nilai *accuracy* 85%, *precision* 87%, *recall* 94%, dan *f1-score* 90%, dari perbandingan performa kedua algoritma tersebut, bahwa algoritma SVM memiliki tingkat *accuracy* yang lebih baik.

Penelitian dilakukan oleh Fatma dkk.[5] dalam penelitian ini, data yang digunakan untuk berjumlah 1500 data ulasan aplikasi Duolingo yang tersedia di *Google Play Store*. Tujuan penelitian ini adalah untuk mengidentifikasi kecenderungan sentimen dan mengukur akurasi kedua algoritma tersebut. Hasilnya menunjukkan bahwa Algoritma *Naive Bayes* memiliki akurasi 81%, presisi 80%, recall 98%, dan skor f1 88%. Sebaliknya, metode *Support Vector Machine* memiliki akurasi 85%, presisi 87%, recall 94%, dan

skor f1 90% yang memiliki tingkat akurasi yang lebih baik.

Penelitian ini dilakukan oleh Muhammad Nanda Fahriza & Noviana Riza [6] tujuan dari penelitian ini adalah untuk mengumpulkan, menilai, dan mensintesis literatur yang relevan tentang penelitian yang telah dilakukan sebelumnya mengenai evaluasi opini pada aplikasi chat GPT dengan metode klasifikasi KNN. Berdasarkan penelitian sebelumnya, hasil penelitian menunjukkan bahwa metode-metode yang umum digunakan dalam analisis sentimen menunjukkan bahwa metode *K-nearest neighbor*, *support vector*, lebih baik pada metode klasifikasi KNN. Setelah evaluasi penelitian ini, ditemukan bahwa nilai K terbaik adalah 9. Analisis tinjauan *Chat Generative Pre-Trained Transformer GPT-a* dengan metode *klasifikasi K-Nearest Neighbor* (KNN) menghasilkan nilai K yang unggul dengan kurasi metode klasifikasi KNN berhasil dalam memprediksi opini dengan rata-rata akurasi 96,6% menunjukkan bahwa metode klasifikasi KNN berhasil dalam memprediksi opini berdasarkan tinjauan data aplikasi.

Penelitian dilakukan oleh Anam dkk.[7] penelitian dilakukan untuk menganalisis sentimen dalam interaksi antara pengguna dan pengelola Twitter menggunakan *Naive Bayes Classifier*, *K-Nearest Neighbor* (KNN), dan *Decision Tree*. Metode tersebut digunakan untuk mengklasifikasikan komentar pengguna terkait penggunaan teknologi setelah dilakukan analisis sentimen, sambil membandingkan akurasi dari masing-masing teknik. Hasil akurasi bervariasi, dengan *Naive Bayes* memperoleh tingkat akurasi sebesar 100%, *K-Nearest Neighbor* (KNN) sebesar 98,25%, dan juga *Decision Tree* sebesar 62,28%.

Penelitian oleh Annisa dkk. [8] telah dilakukan studi untuk membandingkan efektivitas metode *Naive Bayes Classifier* dengan *K-Nearest Neighbor* dalam menganalisis review emosi mengenai perangkat mobile JKN. Penelitiannya menerapkan kedua metode tersebut, NBC dan KNN, untuk mengevaluasi 2.847 data yang terbagi dalam tiga kelas: positif, netral, dan juga negatif. Data tersebut kemudian disubdivisikan menggunakan 10 uji *K-Fold Cross Validation* untuk mengevaluasi keakuratannya, dengan tingkat akurasi mencapai 61,15% untuk NBC dan 87,59% untuk KNN.

## 2.2. Analisis Sentimen

Analisis sentimen terdiri dari pemrosesan data dari kumpulan data yang tidak terstruktur [9]. Metode ini merupakan bagian dari bidang *Natural Language Processing* (NLP) dan bertujuan untuk mengekstrak dan memahami opini, emosi, hingga perasaan yang terdapat dalam sebuah teks. Kita sering menggunakan analisis sentimen, khususnya untuk melihat tren di jejaring sosial, memahami persepsi orang terhadap suatu produk atau layanan, atau meningkatkan kepuasan pelanggan. Analisis sentimen secara otomatis menggunakan data tekstual untuk

menentukan tren sikap orang terhadap suatu objek, apakah positif atau negatif [10]. Fokus utama pada sentimen analisis adalah agar dapat memahami pendapat seseorang mengenai suatu hal, sehingga dapat dimanfaatkan untuk mengambil sebuah keputusan [11].

## 2.3. Naive Bayes Classifier (NBC)

Metode *Naive Bayes Classifier* digunakan untuk mengklasifikasikan teks dalam analisis sentimen [12]. Algoritma *Naive Bayes Classifier* (NBC) melakukan klasifikasi label atau kelas berdasarkan probabilitas sederhana, dengan asumsi jika setiap elemen data adalah independen satu sama lain [13]. Algoritma *Naive Bayes Multinomial* adalah metode pembelajaran probabilitas yang didasarkan pada teorema *Bayes*, dan digunakan dalam bidang *Natural Language Processing* (NLP) [14]. Keuntungan penggunaan metode *Naive Bayes* baik jika jumlah data latihan yang kecil untuk menghitung parameter-parameter yang diperlukan prosedur klasifikasi [15]. Persamaan *Naive Bayes Classifier* (NBC) adalah sebagai berikut:

$$P(C|X) = \frac{P(X|C)P(c)}{P(x)} \quad (1)$$

## 2.4. K-Nearest Neighbor (KNN)

Algoritma KNN yaitu metode yang cukup efisien dalam mengklasifikasikan. Dalam banyak kasus, algoritma ini digunakan untuk mengklasifikasikan teks dan data [3]. KNN bekerja dengan mencari sejumlah objek terdekat dalam dataset pelatihan untuk objek yang akan diklasifikasikan [7]. Pendekatan umum dalam KNN meliputi:

1. Menentukan jumlah tetangganya yang akan dipertimbangkan.
2. Mengidentifikasi jarak antara objek yang akan diklasifikasikan menggunakan kumpulan data praktis, dan kemudian mengatur data sedemikian rupa untuk mengurutkannya dari yang paling dekat hingga paling jauh.

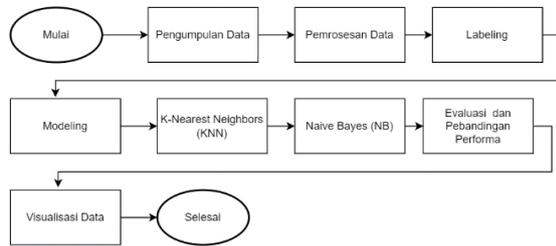
Menurut Irawaty dkk. [16] dimungkinkan untuk melakukan klasifikasi berdasarkan tetangganya yang terdekat, dengan menggunakan mayoritas kelas dari tetangganya tersebut untuk memprediksi kelas dari objek yang akan diklasifikasikan. Persamaan *K-Nearest Neighbor* (KNN) dapat digambarkan sebagai berikut [17]:

$$d_i = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

## 3. METODE PENELITIAN

Untuk melakukan penelitian analisis sentimen, perlu dilakukan beberapa tahapan. Hal ini meliputi pengumpulan data dari opini pengguna Duolingo di aplikasi *Google Play Store*, *pre-processing* (persiapan data sebelum di proses), pelabelan data secara manual, kemudian dilakukan pemodelan dari

data yang di dapatkan menggunakan algoritma *Naive Bayes* dan *K-Nearest Neigbort*. Kemudian dilakukan evaluasi dan perbandingan kinerja untuk menentukan algoritma yang paling efektif. Hal ini bisa diperhatikan pada alur penelitian.



Gambar 1. Alur Penelitian

### 3.1. Pengumpulan dataset

Dataset yang dimanfaatkan pada studi ini terdiri dari teks evaluasi pengguna pada aplikasi Duolingo. Data dikumpulkan dengan menerapkan metode *web scrapping tool* pada aplikasi *Google Play Store*. Kami menggunakan *Google Colab* untuk memproses data dan menghapus data yang tidak diperlukan dari tabel. *Web Scrapping* terdiri dari mengekstrak data dari halaman, sehingga dapat dianalisis, diproses dan digunakan untuk berbagai tujuan [18].

### 3.2. Pra-pemrosesan(Pre-processing)

*Pre-processing* yaitu tahap krusial dalam *cleaning* data hingga menjadi format yang lebih konsisten, umum, dan dasar [19]. Pra-pemrosesan adalah langkah pertama dalam mengelola data sebelum mengintegrasikannya ke dalam model. Prosedur ini melalui beberapa langkah, seperti:

1. *Replace*: Langkah ini terdiri dari mengganti karakter non-alfanumerik dengan string kosong.
2. *Transform case*: Digunakan untuk mengubah semua karakter menjadi huruf kecil.
3. *Tokenize*: terdiri dari memisahkan kata-kata dari setiap kalimat menjadi istilah yang berbeda, kemudian menghilangkan kata-kata yang tidak perlu.
4. *Filter tokens by length*: Digunakan untuk menghapus karakter yang kurang dari 4 atau lebih dari 25 kata.
5. *Replace tokens*: Diterapkan agar dapat mengganti frasa atau kata yang tidak sesuai dengan kbbi dengan kata-kata yang sesuai.
6. *Stopwords*: Digunakan untuk menghapus frasa atau kata yang tidak dibutuhkan.

### 3.3. Pelabelan data

Setelah serangkaian pra-pemrosesan telah dilakukan, maka dilakukan pelabelan data secara manual, yang akan dilakukan menggunakan excel, dengan label positi atau negatif.

### 3.4. Evaluasi dan Perbandingan

Pada tahap ini, dilakukan evaluasi dan perbandingan hasil pemodelan dengan memperhatikan akurasi, presisi, recall, dan *Area Under Curve* (AUC)

untuk menentukan model yang lebih unggul berdasarkan tingkat akurasi.

### 3.5. Visualisasi

Visualisasi digunakan untuk mempermudah identifikasi frasa atau kata yang sering muncul pada ulasan pengguna.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pengumpulan data

Studi menggunakan sampel evaluasi aplikasi Duolingo pada *Google Play Store*. Data diperoleh menggunakan metode *scrapping* dalam bahasa pemrograman *Python*. Data dikumpulkan sebanyak dua kali: pertama kali, dengan tingkat rating 1-2, dan kedua dengan tingkat rating 4-5, data yang di ambil menggunakan filter *Most\_Relevant*. Sehingga didapatkan adalah 600.

### 4.2. Data Awal

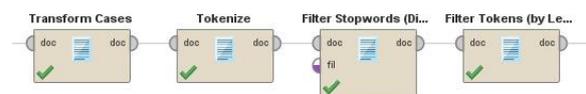
Data awal yang digunakan terdiri dari 600 evaluasi pemakai duolingo di *Google Play Store*, diambil dengan dua kali pengumpulan data sebelumnya. Tabel 1 menunjukkan data awal.

Table 1. Data awal

komentar
Aplikasinya bagus banget,berguna banget untuk belajar bahasa asing. Masukan dari aku sih tambahin fitur kayak google translate biar lebih bagus, biasanya aku pakai google translate kurang akurat,mung kin segitu aja masukan dari aku semoga fiturnya ditambahin yaaa Duolingo.

### 4.3. Pre-Processing Data Awal

*Pre-Processing* sebagai tahap awal dimana data akan di proses untuk menyiapkan data awal agar kemudian dapat digunakan dalam proses analisis. Data awal akan di proses menggunakan aplikasi rapidminer sebagaimana gambar 2.



Gambar 2. Pre-Processing

#### 4.3.1. Replace

Pada tahap ini, data yang di dapatkan akan dilakukan *cleaning*, pada tahapan ini data akan dibersihkan semua tanda baca lain selain huruf, baik itu tanda baca ((/),(;),(,),(")). Hasil Replace dapa dilihat pada Table 2.

Table 2. Hasil Replace

Replace
Aplikasinya bagus banget berguna banget untuk belajar bahasa asing Masukan dari aku sih tambahin fitur kayak google translate biar lebih bagus biasanya aku pakai google translate kurang akurat mung kin segitu aja masukan dari aku semoga fiturnya ditambahin yaaa Duolingo

### 4.3.2. Transform Case

Langkah ini bermaksud agar bisa mengganti huruf kapital menjadi huruf kecil. Data yang telah melalui proses *replace* akan beralih menjadi huruf kecil. Hasil dari transformasi kasus bisa diperhatikan pada Tabel 3.

Tabel 3. Hasil Transform Case

Transform Case
aplikasinya bagus banget berguna banget untuk belajar bahasa asing masukan dari aku sih tambahin fitur kayak google translate biar lebih bagus biasanya aku pakai google translate kurang akurat mung kin segitu aja masukan dari aku semoga fiturnya ditambahin yaaa duolingo.

### 4.3.3. Tokenize

Proses ini dilakukan guna memisahkan setiap kata atau frasa dalam setiap kalimat menjadi token tersendiri. Pembagian token bisa diperhatikan pada Tabel 4.

Table 4. Tokenize

Tokenize
"aplikasinya", "bagus", "banget", "berguna", "banget", "untuk", "belajar", "bahasa", "asing", "masukan", "dari", "aku", "sih", "tambahin", "fitur", "kayak", "google", "translate", "biar", "lebih", "bagus", "biasanya", "aku", "pakai", "google", "translate", "kurang", "akurat", "mung", "kin", "segitu", "aja", "masukan", "dari", "aku", "semoga", "fitunya", "ditambahin", "yaaa", "duolingo".

### 4.3.4. Filter Tokens By Length

Digunakan sebagai penghapus karakter yang kurang dari 4 atau lebih dari 25 kata. Hasil *filter tokens by length* dapat dilihat pada Table 5.

Table 5. Filter Tokens By Length

Filter Tokens By Length
aplikasinya bagus banget berguna banget belajar bahasa asing masukan tambahin fitur kayak google translate biar bagus pakai google translate akurat segitu mung masukan semoga fiturnya ditambahin yaaa duolingo

### 4.3.5. Replace Tokens

Langkah ini digunakan untuk mengganti kata-kata yang tidak sama seperti pada KBBI (Kamus Besar Bahasa Indonesia) dengan frasa atau kata yang benar. Sebagai contoh, kata "mung" dan "yaaa" akan diganti dengan kata yang sesuai.

Table 6. Replace Tokens

Replace Tokens
aplikasinya bagus banget berguna banget belajar bahasa asing masukan tambahin fitur kayak google translate biar bagus pakai google translate akurat segitu mungkin masukan semoga fiturnya ditambahin ya duolingo

### 4.3.6. Stopwords

Digunakan untuk menghapus kata yang tidak penting seperti : "aku", "adalah", "ada", "akulah". Kata-kata seperti ini akan di hapus karena tidak memiliki informasi yang penting.

### 4.4. Labeling

Pada tahap labeling dilakukan untuk menentukan apakah ulasan pengguna aplikasi bersifat positif atau negatif, yang dilakukan secara manual. Dan menghasilkan sisa data bersih yang siap di modeling berjumlah 416 data . Hasil dari labeling dapat dilihat pada Table 7.

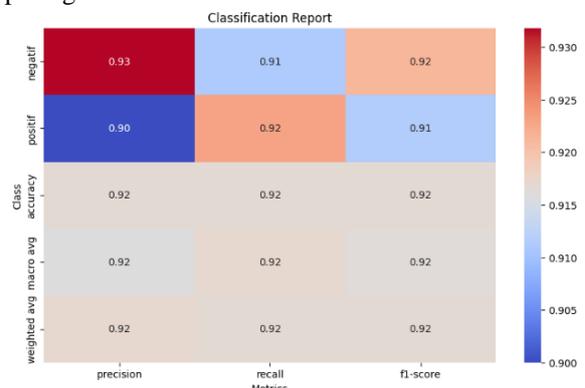
Table 7. Hasil Labeling

Label	Text
negative	kecewa lgbt contoh inggris anggota keluarga to...
positive	suka aplikasi membantu anak sekolah paham baha...
positif	karna bahasa inggris game terseru sedunia jaga...
negatif	mohon maaf pengembang harap propaganda lgbt ga...
negatif	aplikasi gini nyelipin unsur lgbtq bawa bawa j...
...	...
negative	kecewa lgbt contoh inggris anggota keluarga to...
positive	suka aplikasi membantu anak sekolah paham baha...
positif	karna bahasa inggris game terseru sedunia jaga...
negatif	mohon maaf pengembang harap propaganda lgbt ga...
negatif	aplikasi gini nyelipin unsur lgbtq bawa bawa j...

### 4.5. Modeling

#### 4.5.1. Modeling Naive Bayes

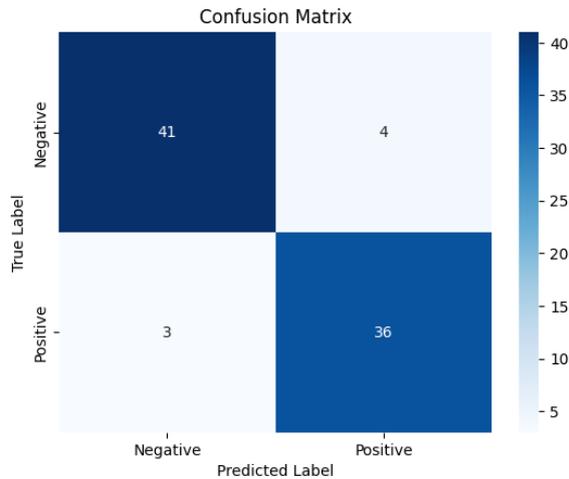
Modeling dilakukan menggunakan google colab, data di bagi sehingga 80% untuk train dan 20% untuk test, untuk kinerja modeling bisa diperhatikan pada gambar 3.



Gambar 3. Kinerja NB

Kinerja pemodelan algoritma NB menunjukkan akurasi sebesar 92%, presisi 93%, recall 92% dengan TP (*True positive*) 36 dan TN (*True Netative*) 41, dari

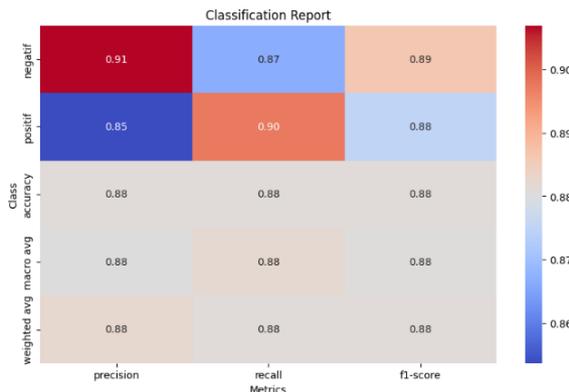
84 data yang di uji. Hasil *Confusion Matrix* terdapat pada gambar 4.



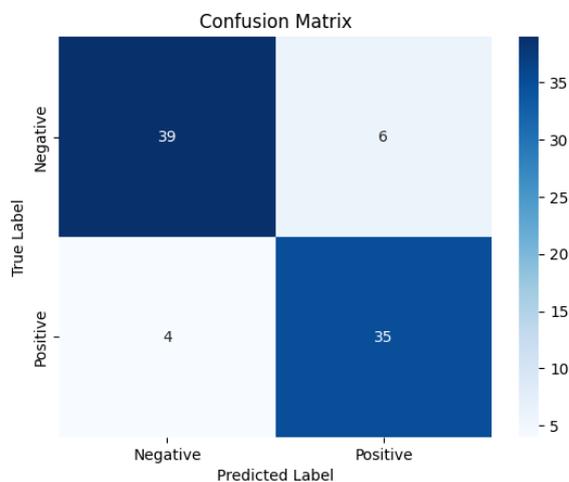
Gambar 4. *Confusion Matrix* NB

**4.5.2. Modeling K-Nearest Neighbor (KNN)**

Modeling dilakukan menggunakan google colab, data di bagi menjadi dua yaitu *training* dan *testing* dengan rasio train 80% dan test 20 % dari data, kemudian model knn dilatih dengan k=5, kemudian dilakukan pengujian method *predict algorithm* dan kinerja modeling bisa diperhatikan pada gambar 5.



Gambar 5. Kinerja KNN



Gambar 6. *Confusion Matrix* KNN

Setelah modeling di buat kemudian diuji maka didapatkan akurasi 88%, presisi 91%, dan recall 90% dengan TN(true negative) 39 dan TP (true positive) 35. Bisa diperhatikan pada gambar 5 dan 6.

**4.6. Visualisai**

Visualisasi bertujuan untuk menyampaikan frasa atau kata apakah yang selalu muncul dari ulasan positif dan negatif

**4.6.1. Ulasan Positif**



Gambar 7. Ulasan positif sering muncul

Pada ulasan positif prasa atau kata yang persentase kemunculan paling banyak yaitu bagus, belajar, duolingo, dan aplikasi.

**4.6.2. Ulasan Negatif**



Gambar 8. Ulasan negative sering muncul

Pada ulasan negatif kata lgbt dan iklan menjadi kata utama yang membuat aplikasi duolingo mendapatkan ulasan negatif.

**5. KESIMPULAN DAN SARAN**

Studi ini memanfaatkan data tinjauan pengguna aplikasi Duolingo di *Google Play Store* dengan total 600 data yang dikumpulkan menggunakan teknik *web scraping* dengan bahasa pemrograman Python. Data tersebut kemudian melalui beberapa tahap *pre-processing*. Ulasan kemudian diberi label secara manual untuk menentukan apakah bersifat positif atau negatif. Hasil menunjukkan bahwa NB mempunyai tingkat akurasi sedikit lebih tinggi daripada KNN. Secara khusus, *Naive Bayes* memiliki 36 *True Positives* (TP) dan 41 *True Negatives* (TN) dari total 84 data yang diuji, sedangkan *K-Nearest Neighbors*

memiliki 35 *True Positives* (TP) dan 39 *True Negatives* (TN) dari total 84 data yang diuji, dengan akurasi *Naïve Bayes* sebesar 92%, presisi 93%, *recall* 92% dan akurasi *K-Nearest Neighbors* sebesar 88%, presisi 91%, *recall* 90%. Terakhir, dilakukan visualisasi untuk melihat frasa-frasa yang paling sering keluar dalam ulasan positif dan negatif.

Didapati frasa atau kata yang sering muncul dalam ulasan positif termasuk "bagus", "belajar", "Duolingo", dan "aplikasi", sedangkan dalam ulasan negatif adalah "lgbt" dan "iklan". Masukan pada penelitian selanjutnya yaitu untuk mempelajari metode *pre-processing* data yang lebih baik dan efektif. Hal ini dapat mencakup penggunaan teknik pemrosesan bahasa alami baru atau model pembelajaran mesin yang lebih canggih. Penelitian lanjutan dapat memperluas cakupan sampel data dan melakukan analisis lintas platform untuk meningkatkan validitas dan aplikabilitas umum. Dengan demikian, penelitian selanjutnya di bidang ini diharapkan dapat memberikan pedoman yang lebih komprehensif bagi praktisi dan pengembang aplikasi serta meningkatkan pemahaman kita tentang analisis sentimen dalam konteks aplikasi pembelajaran bahasa.

#### DAFTAR PUSTAKA

- [1] S. A. Saputra, D. Rosiyadi, W. Gata, dan S.M. Husain, "Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization," *Jurnal Resti: Rekayasa Sistem dan Teknologi Informasi*, vol. 3, no. 3, pp. 377-382, 2019, doi: 10.29207/resti.v3i3.1118.
- [2] A. B. Saputra, "Peran AI dalam Dunia Pendidikan," CV Brimedia Global, 2023.
- [3] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," *Jurnal Teknoinfo*, vol. 14, no. 2, pp. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [4] F. Alifiana, M. F. Asnawi, I. A. Ihsannudin, M. A. M. Baihaqy, dan D. Asmarajati, "Analisis Sentimen Aplikasi Duolingo Menggunakan Algoritma Naïve Bayes dan Support Machine Learning," *Device*, vol. 13, no. 2, pp. 223-230, 2023.
- [5] F. Alifiana, F. Asnawi, I. A. Ihsannudin, M. Alif, M. Baihaqy, and D. Asmarajati, "ANALISIS SENTIMEN APLIKASI DUOLINGO MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT MACHINE LEARNING," *JURNAL DEVICE*, vol. 13, pp. 223-230, 2023.
- [6] M. N. Fahriza and N. Riza, "Analisis Sentimen pada Ulasan Aplikasi Chat Generative Pre-Trained Transformer GPT menggunakan Metode Klasifikasi K-Nearest Neighbor (KNN)," *Systematic Literature Review*, *Jurnal Mahasiswa Teknik Informatika (JATI)*, vol. 7, no. 2, pp. 1351-1358, Apr. 2023.
- [7] M. K. Anam, B. N. Pikir, and M. B. Firdaus, "Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 139-150, Nov. 2021, doi: 10.30812/matrik.v21i1.1092.
- [8] C. Annisa, M. Afdal, and T. Khairil Ahsyar, "Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest Neighbor Pada Sentimen Review Aplikasi Mobile JKN," *Media Informatika Budidarma*, vol. 7, no. 3, pp. 1033-1040, 2023, doi: 10.30865/mib.v7i3.6242.
- [9] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, dan A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store: The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews On Google Play Store," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47-54, 2022, doi: 10.57152/malcom.v2i1.195.
- [10] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "TWITTER SOCIAL MEDIA SENTIMENT ANALYSIS AGAINST BITCOIN CRYPTOCURRENCY TRENDS USING RAPIDMINER," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, pp. 1183-1187, Oct. 2022, doi: 10.20884/1.jutif.2022.3.5.289.
- [11] A. Dwiki, A. Putra, and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," vol. 8, no. 2, 2021, [Online]. Available: <http://jurnal.mdp.ac.id>.
- [12] M. Al Khadafi, Kurnia Paranitha Kartika, and Filda Febrinita, "PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN LEXICON BASED UNTUK ANALISIS SENTIMEN CYBERBULLYING PADA BPJS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 725-733, Oct. 2022, doi: 10.36040/jati.v6i2.5633.
- [13] O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, pp. 191-196, 2019, doi: 10.26418/jp.v5i2.32661.
- [14] N.Hidayah dan S. Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820-826, 2021. doi: 10.29207/resti.v5i4.3146.
- [15] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes

- menggunakan Genetic Algorithm dan Bagging," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 504-510, 2021.
- [16] I. Irawaty, R. Andreswari and D. Pramesti, "Vectorizer Comparison for Sentiment Analysis on Social Media Youtube: A Case Study," *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*, Yogyakarta, Indonesia, 2020, pp. 69-74, doi: 10.1109/IC2IE50715.2020.9274650.
- [17] M. Naja Maskuri, H. Harliana, K. Sukerti, and R. M. Herdian Bhakti, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Prediksi Penyakit Stroke", *j.ilm.intech*, vol. 4, no. 01, pp. 130-140, May 2022, doi: 10.46772/intech.v4i01.751.
- [18] A. U. Nabiylah Ramadhanty and I. Najiyah, "IMPLEMENTASI WEB SCRAPING PADA SITUS JURNAL SINTA MENGGUNAKAN FRAMEWORKSELENIUM WEBDRIVER PYTHON," *JIKA (Jurnal Informatika)*, vol. 7, no.1, pp. 29 -36, 2023, doi:10.31000/jika.v7i1.7037.
- [19] J. Muliawan and E. Dazki, "Sentiment Analysis of Indonesia's Capital City Relocation Using Three Algorithms: Naïve Bayes, KNN, and Random Forest," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 5, pp. 1227-1236, 2023. doi: 10.52436/1.jutif.2023.4.5.347.