

Analisis Sentimen Ulasan Duolingo dengan Metode Algoritma *Multinomial Naïve Bayes*

Sentiment Analysis of Duolingo Reviews Using the Multinomial Naïve Bayes Algorithm Method

Robbi Zulcharnain¹, Ginanjar Abdurrahman², Daryanto³

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Email: robizulcharnain@icloud.com¹, abdurrahmanginanjar@unmuhjember.ac.id², daryanto@unmuhjember.ac.id³

Received: 04.05.2024	Revised: 14.05.2024	Accepted: 21.05.2024	Available online: 07.06.2024
-------------------------	------------------------	-------------------------	---------------------------------

Abstract: Duolingo offers an effective solution by presenting challenges, games, and various engaging levels that maintain user interest. However, as its popularity grows, the platform has become a subject of debate regarding its approach to LGBT (Lesbian, Gay, Bisexual, Transgender) representation, seen as an important step toward greater inclusion in the learning space, while others feel that this issue sparks differing opinions. This study aims to analyze user sentiment toward the Duolingo application on the Google Play Store using the K-Fold Cross Validation and Multinomial Naïve Bayes methods. The K-Fold Cross Validation method achieved the highest scores, with an accuracy of 89.69%, precision of 0.89, recall of 0.92, and F1-Score of 0.90, indicating consistent model performance. Applying the Multinomial Naïve Bayes method to the data further demonstrated excellent results, with an accuracy of 89%, precision of 0.86, recall of 0.91, and F1-Score of 0.89. These findings indicate that the method is effective in sentiment analysis. The selection of Duolingo as the research object was based on the acceptance of its certification by prestigious universities abroad.

Keywords: Duolingo, Multinomial Naïve Bayes, Confusion Matrix, K-Fold Cross Validation

Abstrak: Duolingo menawarkan solusi efektif dengan menghadirkan tantangan, permainan, dan berbagai level menarik yang menjaga minat pengguna. Namun, seiring meningkatnya popularitas, platform ini menjadi subjek perdebatan terkait pendekatannya terhadap representasi LGBT (*Lesbian Gay Bisexual Transgender*) sebagai langkah penting menuju keterlibatan yang lebih besar dalam ruang pembelajaran, sementara yang lain merasa bahwa masalah ini memicu perbedaan pendapat. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Duolingo di *Google Play Store* dengan metode *K-Fold Cross Validation* dan *Multinomial Naïve Bayes*. Metode *K-Fold Cross Validation* menghasilkan nilai tertinggi dengan *accuracy* 89.69%, *precision* 0.89, *recall* 0.92 dan *f1-score* 0.90, yang menunjukkan performa model yang konsisten. Penerapan *Multinomial Naïve Bayes* pada data menunjukkan hasil yang sangat baik, dengan *accuracy* 89%, *precision* 0.86, *recall* 0.91, dan *f1-score* 0.89. Hasil ini menunjukkan bahwa metode ini efektif dalam menganalisis sentimen. Pemilihan Duolingo sebagai objek penelitian didasari oleh penerimaan sertifikat di Universitas ternama di luar negeri.

Kata Kunci: Duolingo, Multinomial Naïve Bayes, Confusion Matrix, K-Fold Cross Validation

1. PENDAHULUAN

Pembelajaran bahasa asing semakin diminati di era digital ini seiring dengan meningkatnya kebutuhan globalisasi dan keterampilan multibahasa. Aplikasi Duolingo telah menjadi salah satu platform terdepan dalam menyediakan pembelajaran bahasa secara mudah dan interaktif dengan menggunakan pendekatan gamifikasi. Duolingo memberikan pengalaman belajar yang menyenangkan melalui berbagai level tantangan dan permainan (Aufarakarsie, 2024). Namun, seiring dengan popularitasnya, platform ini mendapatkan berbagai tanggapan, baik positif maupun negatif, dari para pengguna yang tersebar di seluruh dunia.

Seiring meningkatnya penggunaan aplikasi Duolingo, pengguna sering kali meninggalkan ulasan di platform distribusi seperti *Google Play Store*. Ulasan tersebut merupakan sumber data yang sangat berharga untuk memahami pengalaman pengguna dan tingkat kepuasannya terhadap aplikasi. Untuk mengidentifikasi pola dalam ulasan tersebut, analisis sentimen berperan penting. Analisis ini tidak hanya membantu dalam memahami persepsi pengguna, tetapi juga memberikan pandangan yang dapat digunakan pengembang aplikasi untuk meningkatkan pelayanannya (Nashirul, 2024).

Salah satu metode yang efektif dalam klasifikasi ulasan adalah *Multinomial Naïve Bayes*, yang merupakan metode berbasis *probability*. Algoritma ini secara khusus cocok untuk data teks karena mampu menghitung kemungkinan suatu kata muncul dalam ulasan yang termasuk dalam kategori sentimen positif atau negatif. Pada penelitian ini, metode *Multinomial Naïve Bayes* diterapkan untuk mengklasifikasikan ulasan aplikasi Duolingo, dengan fokus pada sentimen positif dan negatif. Pendekatan ini menggabungkan *probability prior*, *likelihood*, serta *laplace smoothing* untuk meningkatkan *accuracy* prediksi. Berdasarkan hasil tersebut, penelitian bertujuan untuk mengeksplorasi

kinerja metode *Multinomial Naïve Bayes* dalam konteks ulasan Duolingo, serta mengukur *accuracy*, *precision*, *recall* dan *f1-score* untuk menilai kinerja model secara keseluruhan (Wibowo dkk., 2023).

Penelitian ini juga diharapkan memberikan kontribusi signifikan dalam analisis sentimen berbasis teks, khususnya untuk aplikasi edukatif seperti Duolingo. Dengan memanfaatkan data ulasan pengguna, pengembang dapat memperoleh wawasan yang lebih baik dalam mengelola konten aplikasi serta mengoptimalkan fitur-fitur yang tersedia, terutama dalam mengatasi keluhan dan meningkatkan pengalaman pengguna secara keseluruhan.

2. TINJAUAN PUSTAKA

Algoritma *Multinomial Naïve Bayes* telah banyak digunakan dalam berbagai penelitian yang berfokus pada analisis sentimen, seperti penilaian masyarakat terhadap Jakarta *International Stadium* (JIS) (Pratama dkk., 2024), ulasan aplikasi Gojek dan Grab ulasan Gojek dan Grab (Gunawan dkk., 2022), kebijakan publik COVID-19 di *Twitter* (Zulfikar dkk., 2023), ulasan aplikasi *Netflix* (Suhendra dkk., 2024), serta sentimen masyarakat terkait pengungsi rohingya (Ananda & Suryono, 2024). Penelitian ini berbeda dari yang sebelumnya dengan menggunakan *Maximum a Posteriori* dan *Vader* untuk meningkatkan *accuracy*.

Menurut (Pratama dkk., 2024), metode *Multinomial Naïve Bayes* untuk analisis sentimen ulasan aplikasi dengan hasil *accuracy* sebesar 85%. Proses *case folding*, *tokenizing*, dan *stemming* membantu meningkatkan kinerja model dalam mengklasifikasikan sentimen positif dan negatif. Meskipun ada keterbatasan pada asumsi independensi antar kata, metode ini tetap memberikan hasil yang memuaskan untuk analisis sentimen.

A. Kajian Teori

Penelitian ini melibatkan konsep-konsep utama seperti *Google Play Store* (*platform* distribusi digital), Duolingo (aplikasi pembelajaran bahasa asing), *Machine Learning* (teknik AI untuk belajar dari data), *Text Mining* (ekstrak informasi dari teks), *Text Preprocessing* (pengolahan teks untuk analisis), *TF-IDF*, *K-Fold Cross Validation*, *Confusion Matrix*, *Naïve Bayes*, *Multinomial Naïve Bayes*, dan Visualisasi *Word Cloud*.

1. Analisis Sentimen

Analisis sentimen adalah penelitian yang menganalisis opini, penilaian, evaluasi, dan emosi terhadap entitas seperti produk, layanan atau individu. Sejak awal 2000-an, bidang ini telah berkembang pesat dalam pemrosesan bahasa alami dan memberikan manfaat besar, terutama dalam memantau rekasi masyarakat terhadap produk (Sayogo dkk., 2024)

2. Naïve Bayes

Naïve Bayes adalah teknik pembelajaran mesin probabilistik yang mengklasifikasikan data berdasarkan *Teorema Bayes*. Metode ini efektif dengan sedikit data latih dan mampu menangani nilai yang hilang dalam data (Sholekhah dkk., 2024).

3. Multinomial Naïve Bayes

Multinomial Naïve Bayes adalah metode probabilistik yang mengasumsikan independensi antar atribut. Meskipun asumsi ini tidak selalu benar, metode ini berhasil dalam klasifikasi teks. Probabilitas kelas dokumen dihitung berdasarkan *prior* probabilitas bersyarat untuk setiap token dalam dokumen (Sabrani dkk., 2020).

Multinomial Naïve Bayes dipilih karena kesederhanaannya dalam menangani data teks dengan representasi berbasis frekuensi kata. Metode ini sangat efisien untuk dataset yang besar dengan banyak fitur, seperti ulasan teks, karena dapat menghitung probabilitas bersyarat secara langsung. Selain itu, penelitian sebelumnya menunjukkan bahwa *Multinomial Naïve Bayes* menghasilkan *accuracy* tinggi pada analisis sentimen teks berbahasa Indonesia (Suhendra dkk., 2024).

Rumus perhitungan *Multinomial Naïve Bayes* merupakan metode *supervised learning* sehingga setiap data perlu dilabeli sebelum dilakukan *training*, dapat digunakan persamaan 1.

$$P(c|d) \propto P(c) \prod_{k=1}^n P(t_k|c) \quad (1)$$

keterangan:

$P(c|d)$: Probabilitas dokumen d berada di kelas c

$P(c)$: *Prior* probabilitas suatu dokumen berada di kelas c

$\{t_1, \dots, t_n\}$: Token dalam dokumen d yang merupakan bagian dari *vocabulary* dengan jumlah n
 $P(t_k|c)$: Probabilitas bersyarat *term* t_k berada di dokumen kelas c

Klasifikasi dokumen bertujuan untuk menentukan kelas terbaik untuk suatu dokumen. Kelas terbaik dalam klasifikasi *Multinomial Naïve Bayes* ditentukan dengan mencari *Maximum a Posteriori* (MAP) kelas C_{map} , dapat digunakan persamaan 2.

$$C_{map} = \arg \max_{c \in C} \hat{P}(C) \prod_{k=1}^n P(t_k|c) \quad (2)$$

keterangan:

C_{map} : Kelas yang diprediksi dengan probabilitas *maksimum*

$\arg \max_{c \in C}$: Operator yang mencari nilai C dalam himpunan kelas C yang memaksimalkan ekspresi berikutnya

$\hat{P}(C)$: Probabilitas *a priori* dari kelas C

$\prod_{k=1}^n P(t_k|c)$: Produk dari probabilitas bersyarat $P(t_k|c)$, di mana t_k adalah fitur ke- k . Ini adalah probabilitas fitur t_k muncul dalam kelas C

Persamaan yang menggunakan logaritma dari probabilitas dapat dinyatakan dalam persamaan 3.

$$C_{map} = \arg \max_{c \in C} \left[\log \hat{P}(c) + \sum_{k=1}^n \log \hat{P}(t_k|c) \right] \quad (3)$$

$\hat{P}(c)$ dan $\hat{P}(t_k|c)$ didapatkan dengan menghitung *maximum likelihood*, yang merupakan frekuensi relatif dari parameter (Starmer, 2020). Untuk *prior*, dapat digunakan persamaan 4.

$$\hat{P}(c) = \frac{N_c}{N} \quad (4)$$

keterangan:

$\hat{P}(c)$: *Prior* probabilitas suatu dokumen berada di kelas c

N_c : Jumlah dokumen dengan kelas c

N : Jumlah seluruh dokumen

$\hat{P}(t|c)$ merupakan probabilitas frekuensi relatif *term* t dalam dokumen berada di kelas c (Sabrani dkk., 2020), yang dapat dihitung menggunakan persamaan 5.

$$\hat{P}(t|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}} \quad (5)$$

keterangan:

$\hat{P}(t|c)$: Probabilitas bersyarat *term* t berada di dokumen pada kelas c .

T_{ct} : Jumlah kemunculan *term* t pada dokumen dengan kategori c

$\sum_{t' \in V} T_{ct'}$: Jumlah frekuensi seluruh *term* pada kelas c

Perhitungan *maximum likelihood* memiliki kelemahan, yaitu suatu kata dalam kelas yang tidak terlihat pada data *training* akan memiliki nilai 0. Hal ini menyebabkan perhitungan $P(c|d)$ menghasilkan nilai 0, karena setiap bilangan yang dikelikan dengan 0 akan menghasilkan 0. Untuk mengatasi masalah ini, diterapkan teknik *add-one* atau *Laplace smoothing*.

Smoothing pada algoritma *Multinomial Naïve Bayes*, digunakan teknik *laplace smoothing* untuk mengatasi masalah probabilitas 0. Masalah ini terjadi ketika suatu kata dalam kelas tertentu tidak muncul dalam data latih, sehingga probabilitas bersyarat kata tersebut menjadi 0. Dengan *smoothing*, nilai probabilitas setiap kata disesuaikan agar tetap positif, memastikan bahwa hasil prediksi lebih stabil dan representatif. Misalnya, konstanta kecil seperti $a = 1$ ditambahkan kedalam perhitungan probabilitas untuk memastikan tidak ada nilai 0, yang meningkatkan *accuracy* model dalam menangani dataset dengan frekuensi rendah atau data yang tidak seimbang (Ajay, 2023), sehingga persamaan 5 berubah menjadi persamaan 6.

$$\hat{P}(t|c) = \frac{T_{ct} + 1}{\sum_{t' \in V} (T_{ct'} + 1)} = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'}) + B'} \quad (6)$$

keterangan:

B : Jumlah seluruh *term* pada *vocabulary*

Sedangkan rumus *Multinomial Naïve bayes* dengan menggunakan pembobotan TF-IDF, dapat dilihat pada persamaan 7.

$$\hat{P}(t|c) = \frac{W_{ct}}{(\sum_{w' \in V} W_{ct'}) + B'} \quad (7)$$

keterangan:

W_{ct} : Bobot TF-IDF *term* t pada dokumen dengan kategori c

4. TF-IDF (*Term Frequency-Inverse Document Frequency*)

TF-IDF adalah transformasi data yang menghitung frekuensi setiap *term* yang muncul dalam dokumen. *Term Frequency* (TF) mengukur seberapa sering *term* muncul, sementara *Inverse Document Frequency* (IDF) menilai seberapa umum *term* tersebut dalam korpus. Nilai TF-IDF memberikan indikasi kekuatan relevansi *term* dalam konteks dokumen.

Rumus TF dapat didefinisikan pada persamaan 8.

$$TF(t_k, d_j) = f(t_k, d_j) \quad (8)$$

keterangan:

$f(t_k, d_j)$: Jumlah kemunculan *term* t_k dalam dokumen d_j

IDF dapat dihitung pada persamaan 9.

$$IDF(t_k) = \log \frac{D}{df(t_k)} \quad (9)$$

keterangan:

D : Total jumlah dokumen dalam korpus

$df(t_k)$: Jumlah dokumen yang mengandung *term* t_k

Sehingga TF-IDF dapat dirumuskan pada persamaan 10:

$$TF\ IDF(t_k, d_j) = TF(t_k, d_j) * ID. \quad (10)$$

5. K-Fold Cross Validation

K-Fold Cross Validation adalah teknik untuk membagi data latih dan data uji. Teknik ini mengurangi bias dalam model dengan memberi setiap data kesempatan untuk menjadi data tes. Nilai K menunjukkan jumlah pembagian data. Pada setiap iterasi, sebagian data digunakan untuk pelatihan, dan sisanya untuk pengujian, sehingga setiap subset data berfungsi sebagai data uji.

6. Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi kinerja algoritma klasifikasi dengan menampilkan perbandingan antara klasifikasi yang dihasilkan dan data sebenarnya. Elemen-elemen penting dalam *confusion matrix* meliputi:

- *True Positive* (TP): Jumlah label positif yang diprediksi benar.
- *True Negative* (TN): Jumlah label negatif yang diprediksi benar.
- *False Positive* (FP): Jumlah label negatif yang diprediksi positif.
- *False Negative* (FN): Jumlah label positif yang diprediksi negatif.

7. Visualisasi Word Cloud

Visualisasi *Word Cloud* menampilkan kata—kata dari teks dengan ukuran *font* yang menunjukkan frekuensi kemunculan. Kata-kata yang lebih sering muncul ukuran lebih besar, memungkinkan analisis lebih cepat terhadap konteks. Contohnya, dalam konteks wisata halal di pulau Lombok, kata “indah” muncul paling sering, menunjukkan fasilitas yang penting bagi wisatawan muslim, diikuti oleh kata-

[illegible]

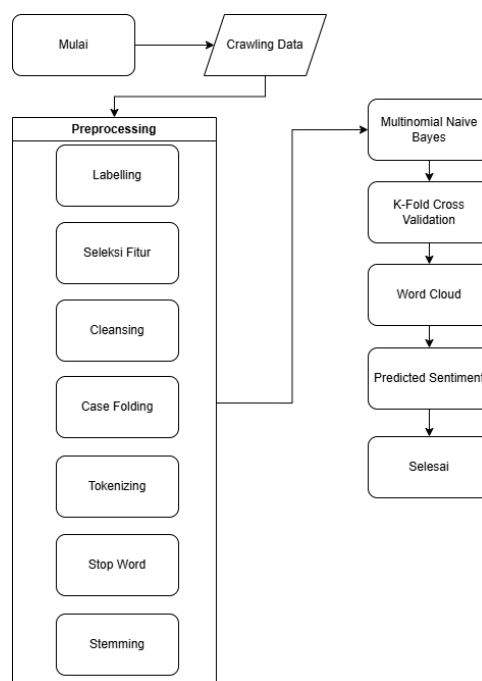
5

4. *Stopword*

Stopword yakni mengubah kata kebentuk dasarnya.

5. *Stemming*

Stemming yakni menghapus kata-kata yang tidak signifikan atau tidak mempengaruhi analisis.



Gambar 2. Alur Penelitian

4. HASIL DAN PEMBAHASAN

A. *Crawling Data*

Crawling data adalah proses pengumpulan informasi secara otomatis dari berbagai sumber di internet. Dalam penelitian ini, dilakukan crawling data sehingga diperoleh total 2000 data, yang terdiri atas 967 data dengan kelas positif dan 1033 data dengan kelas negatif. Hasil ini menunjukkan distribusi data yang seimbang untuk analisis sentimen berikut adalah datanya.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
0	c21c64d4-0a88-4d53-9ee4-45ad5f295519	Chelsea Machiko	lh.googleusercontent.com/a/ACg8oc...	Gambar nya jdi kaku, karakter nya gk gerak, cu...	3	33	6.5.4	2024-11-07 06:14:55	None	None
1	9cea2ddf-1174-47a1-8933-7610263af841	Rindi Antika	lh.googleusercontent.com/a/ACg8oc...	buat kalian yang lagi proses belajar ini recom...	5	123	6.7.3	2024-11-14 05:41:05	None	None
2	fa3bd275-a55e-4456-b3bc-e92ccf3f2553	Miharu 09	lh.googleusercontent.com/a/-ALV-U...	Jujur suka banget belajar bahasa kalau pakai D...	5	445	6.4.5	2024-11-03 07:23:41	None	None
3	074a28bd-b0a7-47c5-9c23-6e2bc0fb61a	Paima Rotua	lh.googleusercontent.com/a/ACg8oc...	Hallo Duolingo!, Saya sangat senang menggunaka...	5	30	6.7.3	2024-11-18 11:45:17	None	None
4	16c77e81-28c5-4977-87e8-598fcabe1065	Nazwa Aulya	lh.googleusercontent.com/a/ACg8oc...	kalo mau belajar bahasa rasanya jadi seru dan ...	5	261	6.5.3	2024-11-02 19:41:09	None	None

Gambar 3. Hasil Crawling Data

B. *Labelling*a. *VADER (Valanced Aware Dictionary Sentiment Reasoner)*

VADER adalah metode analisis sentimen berbasis pendekatan leksikal yang efektif untuk teks berbahasa alami. Metode ini menggunakan kamus leksikon dengan daftar kata dan skor sentimen yang telah ditentukan sebelumnya. Pendekatannya berpusat pada kebijaksanaan dan penilaian manusia.

Prosesnya melibatkan pencocokan kata dalam teks dengan kamus untuk menghitung skor sentimen keseluruhan tanpa memerlukan konsultasi dengan sumber lain (Ramadhanu dkk., 2023).

Validasi *labelling* dilakukan dengan validasi manual terhadap 10% sampel data. Hasil validasi menunjukkan tingkat kesesuaian sebesar 95%, mengindikasikan bahwa metode VADER cukup andal untuk tugas ini. Validasi ini penting untuk meminimalkan kesalahan klasifikasi awal yang dapat mempengaruhi performa model secara keseluruhan.

Adapun hasil dari *labelling* adalah sebagai berikut:

	content	score	\
0	keren😎 sangat bagus untuk menambah wawasan dan...	5	
1	(Edit)Mohon maaf bintangnya saya kurangi lagi ...	1	
2	dengan duo linggo aku belajar nya lebih nyaman...	5	
3	Ini bukan nya tambah pintar malah jadi tambah ...	1	
4	kan disuruh ngomong gitu kan aku udah ngomong ...	1	

	vader_sentiment_score	sentiment
0	0.836	positive
1	0.000	negative
2	0.000	positive
3	0.000	negative
4	0.000	negative

Gambar 4. Hasil *labelling*

C. Preprocessing

Pada tahapan ini proses membersihkan dan mempersiapkan data teks sebelum analisis, mencakup langkah seperti *cleansing*, *case folding*, *tokenizing*, *stopword*, dan *stemming* untuk meningkatkan akurasi model. Adapun hasil dari *preprocessing* adalah sebagai berikut:

score		content	label	vader_sentiment_score	sentiment	newcontent	TOKENIZATION	STOP_REMOVAL	STEMMED	STEMMING
0	5	keren😎 sangat bagus untuk menambah wawasan dan...	1	0.836	positive	keren sangat bagus untuk menambah wawasan dan ...	['keren', 'sangat', 'bagus', 'untuk', 'menamba...']	['keren', 'sangat', 'bagus', ' ', 'menambah', ' ', '...']	keren sangat bagus untuk tambah wawas dan ilmu...	['keren', 'sangat', 'bagus', 'untuk', 'tambah', 'wawas', 'dan', 'ilmu', '...']
1	1	(Edit)Mohon maaf bintangnya saya kurangi lagi ...	0	0.000	negative	(Edit)Mohon maaf bintangnya saya kurangi lagi ...	['(', 'edit', 'mohon', 'maaf', 'bintangnya', 'saya', 'kurangi', 'lagi', '...', ')']	['(', 'edit', 'mohon', 'maaf', 'bintangnya', ' ', '...']	edit mohon maaf bintang saya kurang lagi 4 kar...	['edit', 'mohon', 'maaf', 'bintang', 'saya', 'kurang', 'lagi', ' ', '...']
2	5	dengan duo linggo aku belajar nya lebih nyaman...	1	0.000	positive	dengan duo linggo aku belajar nya lebih nyaman...	['dengan', 'duo', 'linggo', 'aku', 'belajar', 'nya', 'lebih', 'nyaman', '...']	[' ', 'duo', 'linggo', 'aku', 'belajar', 'nya', ' ', '...']	dengan duo linggo aku ajar nya lebih nyaman bi...	['dengan', 'duo', 'linggo', 'aku', 'ajar', 'nya', ' ', '...']
3	1	Ini bukan nya tambah pintar malah jadi tambah ...	0	0.000	negative	Ini bukan nya tambah pintar malah jadi tambah ...	['ini', 'bukan', 'nya', 'tambah', 'pinter', 'malah', 'jadi', 'tambah', '...']	[' ', 'bukan', 'nya', 'tambah', 'pinter', 'malah', ' ', '...']	ini bukan nya tambah pintar malah jadi tambah ...	['ini', 'bukan', 'nya', 'tambah', 'pinter', 'malah', 'jadi', 'tambah', '...']
4	1	kan disuruh ngomong gitu kan aku udah ngomong ...	0	0.000	negative	kan disuruh ngomong gitu kan aku udah ngomong ...	['kan', 'disuruh', 'ngomong', 'gitu', 'kan', 'aku', 'udah', 'ngomong', ' ', '...']	['kan', 'disuruh', 'ngomong', 'ngomong', 'gitu', 'kan', 'aku', 'udah', 'ngomong', ' ', '...']	kan suruh ngomong gitu kan aku udah ngomong ya...	['kan', 'suruh', 'ngomong', 'ngomong', 'gitu', 'kan', 'aku', 'udah', 'ngomong', ' ', '...']

Gambar 5. Hasil *Preprocessing*

D. Visualisasi Word Cloud

Visualisasi *Word Cloud* adalah representasi grafis dari frekuensi kemunculan kata dalam teks, di mana ukuran huruf menunjukkan frekuensi kata. Kata yang sering muncul ditampilkan lebih besar dan menonjol, sedangkan kata dengan frekuensi lebih rendah ditampilkan lebih kecil (Alida, 2020). Hasil dari visualisasi *word cloud* adalah sebagai berikut:



```

Menghitung P(D) untuk kata 'ajar':
P(D) = (P(ajar | Positif) * P(Positif)) + (P(ajar | Negatif) * P(Negatif))
P(D) = (0.03986532 * 0.4835) + (0.01805511 * 0.5165)
P(D) = 0.01927488 + 0.00932546 ≈ 0.02860034

Menghitung Posterior untuk Positif untuk kata 'ajar':
P(Positif | D) = P(ajar | Positif) * P(Positif) / P(D)
P(Positif | D) = 0.01927488 / 0.02860034 ≈ 0.67393881

Menghitung Posterior untuk Negatif untuk kata 'ajar':
P(Negatif | D) = P(ajar | Negatif) * P(Negatif) / P(D)
P(Negatif | D) = 0.00932546 / 0.02860034 ≈ 0.32606119

```

Gambar 8. Hasil Perhitungan *Posterior Probability*

Pada perhitungan *posterior probability* nilai yang perlu dicari adalah *prior probability* dan *likelihood*, *prior probability* mencerminkan peluang awal suatu dokumen termasuk dalam kelas tertentu sebelum mempertimbangkan fitur lainnya. Dalam penelitian ini, *prior probability* ditentukan berdasarkan distribusi data dengan nilai positif sebesar 48.35% dan negatif sebesar 51.65%. Pemilihan ini memastikan bahwa model mempertimbangkan proporsi sebenarnya dari semua kelas, sehingga menghasilkan prediksi yang lebih akurat.

Likelihood dalam algoritma ini dihitung berdasarkan frekuensi relatif suatu kata dalam kelas positif dan negatif. Pemilihan *likelihood* dipengaruhi oleh penggunaan *smoothing* untuk mengatasi kemungkinan kata dengan frekuensi 0, pendekatan ini menjamin bahwa semua kata mendapatkan nilai probabilitas, bahkan jika kata tersebut jarang muncul atau tidak ada. *Likelihood* ini sangat penting untuk menentukan *posterior probability* yang digunakan dalam klasifikasi, berikut adalah *prior* dan *likelihood*.

```

Prior Probability P(Positif) = 0.4835
Prior Probability P(Negatif) = 0.5165

```

Gambar 9. Nilai *Prior Probability*

```

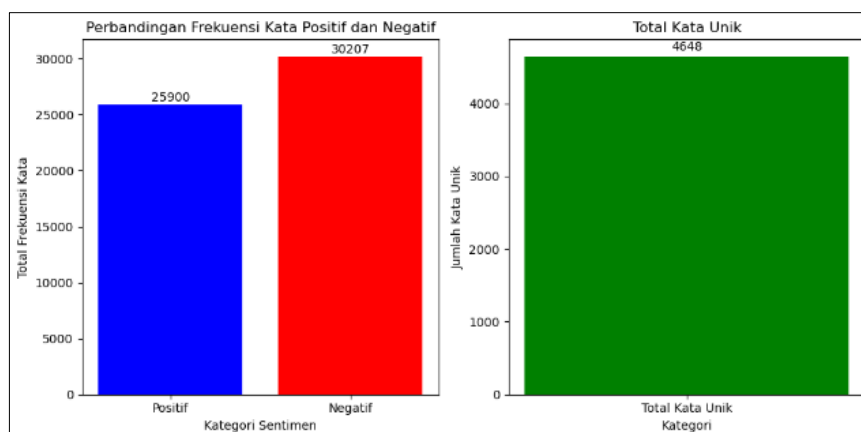
Kata: ajar
Likelihood Positif = 0.688277
Likelihood Negatif = 0.311723

Kata: ajarin
Likelihood Positif = 0.632018
Likelihood Negatif = 0.367982

```

Gambar 10. Nilai *Likelihood*

Pada perhitungan *likelihood* melibatkan frekuensi kata positif, frekuensi kata negatif dan total kata unik. Berikut adalah hasilnya:



Gambar 11. Frekuensi Kata dan Total Kata Unik

b. MAP (*Maximum a Posteriori*)

Maximum a Posteriori (MAP) adalah suatu pendekatan dalam statistik dan pembelajaran mesin yang digunakan untuk memperkirakan parameter suatu model berdasarkan data yang diamati, dengan memaksimalkan probabilitas *posterior*. Pendekatan ini menggabungkan informasi dari data (*likelihood*) dan pengetahuan sebelumnya tentang distribusi parameter (*prior*) untuk menghasilkan estimasi parameter yang paling mungkin. Dalam MAP, *probabilitas posterior* dihitung menggunakan rumus Bayes, yang menggabungkan *prior* dan *likelihood*, dan kemudian nilai parameter yang memaksimalkan *posterior* tersebut dipilih sebagai estimasi terbaik.

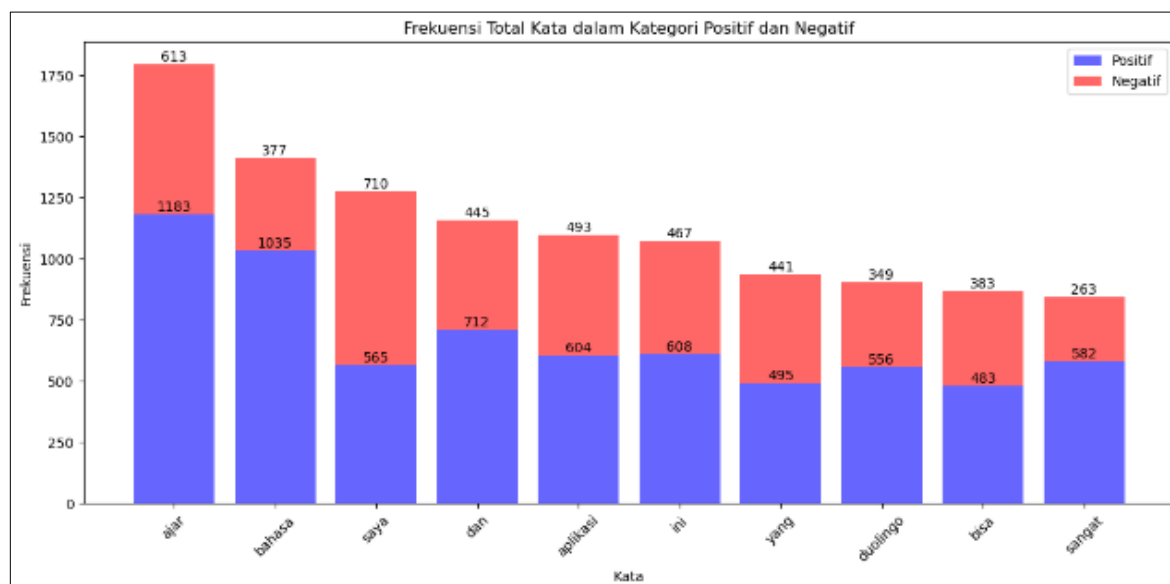
```
Perhitungan C_map:

Kata: ajar
Prior Positif: P(Positif) = 0.4835 -> log(P(Positif)) = -0.7267
Prior Negatif: P(Negatif) = 0.5165 -> log(P(Negatif)) = -0.6607
P(Positif|D) = 0.6732 -> log(P(Positif|D)) = -0.3958
P(Negatif|D) = 0.3268 -> log(P(Negatif|D)) = -1.1183
C_map Positif = log(P(Positif)) + log(P(Positif|D)) = -0.7267 + -0.3958 = -1.1225
C_map Negatif = log(P(Negatif)) + log(P(Negatif|D)) = -0.6607 + -1.1183 = -1.7790
Klasifikasi: Positif
```

Gambar 12. Hasil Perhitungan MAP

c. Frekuensi Kata Positif dan Negatif

Diagram ini menggambarkan perbandingan frekuensi kata-kata teratas dalam dua kategori sentimen, yaitu positif dan negatif. Setiap kata yang ditampilkan mewakili kata-kata dengan frekuensi tertinggi dalam dataset, dan batang berwarna biru menunjukkan frekuensi kata dalam kategori positif, sementara batang berwarna merah menunjukkan frekuensi kata dalam kategori negatif. Visualisasi ini memudahkan untuk melihat kata-kata mana yang paling sering muncul dalam masing-masing kategori sentimen, serta bagaimana kedua kategori tersebut saling berhubungan dalam hal penggunaan kata.



Gambar 13. Frekuensi Total Setiap Kata

d. TF-IDF dan Bobot TF-IDF MNB

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik transformasi data yang memberikan nilai numerik pada setiap kata dalam teks ulasan. Teknik ini memberi bobot pada setiap kata berdasarkan frekuensinya dalam dokumen dan pentingnya kata tersebut di seluruh koleksi dokumen, membantu sistem untuk memahami konteks dan relevansi kata dalam dokumen. Pembobotan ini sering diwakili dalam bentuk nilai biner untuk menandai keberadaan kata dalam teks.

	word	tfidf	Bobot_MNB_TF-IDF
125	ajar	0.064755	0.012429
346	bahasa	0.061845	0.011870
2956	saya	0.052686	0.010112
2936	sangat	0.049301	0.009463
237	aplikasi	0.047985	0.009210
764	dan	0.046745	0.008972
975	duolingo	0.043249	0.008301
1425	ini	0.043072	0.008267
3726	yang	0.040364	0.007747
1418	inggris	0.040109	0.007698

Gambar 14. Hasil TF-IDF dan Bobot TF-IDF MNB

e. *K-Fold Cross Validation*

K-Fold Cross Validation adalah teknik untuk membagi data menjadi set data latih dan data uji, metode ini digunakan untuk mengurangi bias dalam evaluasi model dengan membagi dataset kedalam $K = 5$. Dalam penelitian ini, nilai K ditetapkan sebesar 5. Pemilihan $K = 5$ bertujuan untuk menjaga keseimbangan antara ukuran data latih dan data uji disetiap iterasi, sekaligus meminimalkan resiko *obverfitting* dan *underfitting*. $K = 5$ sering digunakan dalam praktek *Machine Learning* karena memberikan keseimbangan yang baik antara bias dan variasi model. Selain itu jumlah iterasi yang relatif rendah dengan $K = 5$ juga memastikan efisiensi waktu komputasi dibandingkan nilai K yang lebih besar. Teknik ini juga memastikan bahwa setiap data diuji setidaknya sekali, sehingga hasil evaluasi lebih konsisten (Chugani, 2024).

- Skenario Pertama

Evaluasi Model Skenario 1				
Pada skenario 1, dilakukan pengujian klasifikasi menggu				
Berikut merupakan hasilnya :				
Tabel 1. Hasil Confusion Matrix Skenario 1				
Total Data: 320				
Data Salah: 34				
Akurasi: 89.38%				
Classification Report:				
	precision	recall	f1-score	support
Negatif	0.94	0.87	0.90	183
Positif	0.84	0.93	0.88	137
accuracy			0.89	320
macro avg	0.89	0.90	0.89	320
weighted avg	0.90	0.89	0.89	320

Gambar 15. Hasil Skenario Pertama

- Skenario Kedua

Evaluasi Model Skenario 2				
Pada skenario 2, dilakukan pengujian klasifikasi menggu				
Berikut merupakan hasilnya :				
Tabel 1. Hasil Confusion Matrix Skenario 2				
Total Data: 320				
Data Salah: 38				
Akurasi: 88.12%				
Classification Report:				
	precision	recall	f1-score	support
Negatif	0.90	0.87	0.88	165
Positif	0.86	0.90	0.88	155
accuracy			0.88	320
macro avg	0.88	0.88	0.88	320
weighted avg	0.88	0.88	0.88	320

Gambar 16. Hasil Skenario Kedua

- Skenario Ketiga

Evaluasi Model Skenario 3
Pada skenario 3, dilakukan pengujian klasifikasi menggunakan
Berikut merupakan hasilnya :

Tabel 1. Hasil Confusion Matrix Skenario 3
Total Data: 320
Data Salah: 38
Akurasi: 88.12%

Classification Report:

	precision	recall	f1-score	support
Negatif	0.91	0.84	0.88	159
Positif	0.86	0.92	0.89	161
accuracy			0.88	320
macro avg	0.88	0.88	0.88	320
weighted avg	0.88	0.88	0.88	320

Gambar 17. Hasil Skenario Ketiga

- Skenario Keempat

Evaluasi Model Skenario 4
Pada skenario 4, dilakukan pengujian klasifikasi menggunakan
Berikut merupakan hasilnya :

Tabel 1. Hasil Confusion Matrix Skenario 4
Total Data: 320
Data Salah: 33
Akurasi: 89.69%

Classification Report:

	precision	recall	f1-score	support
Negatif	0.91	0.88	0.89	153
Positif	0.89	0.92	0.90	167
accuracy			0.90	320
macro avg	0.90	0.90	0.90	320
weighted avg	0.90	0.90	0.90	320

Gambar 18. Hasil Skenario Keempat

- Skenario Kelima

Evaluasi Model Skenario 5
Pada skenario 5, dilakukan pengujian klasifikasi menggunakan
Berikut merupakan hasilnya :

Tabel 1. Hasil Confusion Matrix Skenario 5
Total Data: 320
Data Salah: 34
Akurasi: 89.38%

Classification Report:

	precision	recall	f1-score	support
Negatif	0.89	0.91	0.90	163
Positif	0.90	0.88	0.89	157
accuracy			0.89	320
macro avg	0.89	0.89	0.89	320
weighted avg	0.89	0.89	0.89	320

Gambar 19. Hasil Skenario Kelima

f. Evaluasi Model

Pada tahapan ini evaluasi model merupakan langkah penting dalam analisis untuk mengukur performa algoritma dalam memprediksi data. Pada model *Multinomial Naïve Bayes*, evaluasi dilakukan dengan memanfaatkan beberapa metrik yang sering digunakan dalam pembelajaran mesin, seperti *accuracy*, *precision*, *recall*, dan *f1-score*.

- *Accuracy* menunjukkan proporsi data yang berhasil diklasifikasikan dengan benar terhadap keseluruhan data.

- *Precision* mengukur sejauh mana prediksi model untuk kelas tertentu benar-benar relevan, terutama dalam mendeteksi data positif.
- *Recall*, atau sensitivitas, menunjukkan kemampuan model untuk mendeteksi seluruh data positif dalam dataset.
- *F1-Score* adalah rata-rata harmonik antara *precision* dan *recall* yang memberikan keseimbangan antara kedua metrik tersebut.

Selain itu, *Confusion Matrix* digunakan untuk memberikan gambaran lebih rinci mengenai hasil prediksi, dengan membagi hasil klasifikasi menjadi empat kategori: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Akurasi Model Multinomial Naive Bayes: 0.89				
Laporan Klasifikasi:				
	precision	recall	f1-score	support
negatif	0.92	0.87	0.89	210
positif	0.86	0.91	0.89	190
accuracy			0.89	400
macro avg	0.89	0.89	0.89	400
weighted avg	0.89	0.89	0.89	400
Confusion Matrix:				
[[183 27]				
[17 173]]				

Gambar 20. Hasil Evaluasi Model MNB

Tabel 4. 1 Nilai Confusion Matrix

Prediksi	Negatif	Positif	Total
Negatif	183(TN)	27(FP)	210
Positif	17(FN)	173(TP)	190
Total	200	200	400

g. *Predicted Sentiment*

Berikut adalah hasil analisis sentimen menggunakan model yang telah dilatih sebelumnya untuk mengklasifikasikan teks ulasan ke dalam dua kategori sentimen, yaitu positif dan negatif. Analisis ini dilakukan dengan cara memasukkan teks baru sebagai input, yang kemudian diproses oleh model untuk menentukan sentimen berdasarkan pola yang telah dipelajari. Pada contoh berikut, model mampu mengklasifikasikan teks ulasan dengan akurat, memberikan wawasan tentang pandangan atau opini pengguna terhadap suatu produk atau layanan.

```

new_text = input("\nMasukkan Teks Baru: ")
new_text_vec = vectorizer.transform([new_text])
predicted_sentimen = naive_bayes.predict(new_text_vec)

if predicted_sentimen[0] == 1:
    sentiment_label = "positif"
elif predicted_sentimen[0] == 0:
    sentiment_label = "negatif"

print("Hasil Analisis Sentimen untuk Teks Baru :",sentiment_label)

Masukkan Teks Baru: aplikasinya bagus untuk belajar sehari hari
Hasil Analisis Sentimen untuk Teks Baru : positif

```

Gambar 21. Prediksi Sentimen Positif

```
new_text = input("\nMasukkan Teks Baru: ")
new_text_vec = vectorizer.transform([new_text])
predicted_sentimen = naive_bayes.predict(new_text_vec)

if predicted_sentimen[0] == 1:
    sentiment_label = "positif"
elif predicted_sentimen[0] == 0:
    sentiment_label = "negatif"

print("Hasil Analisis Sentimen untuk Teks Baru :", sentiment_label)

Masukkan Teks Baru: aplikasinya jelek, ada lgbtnya lagi aneh
Hasil Analisis Sentimen untuk Teks Baru : negatif
```

Gambar 22. Prediksi Sentimen Negatif

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Duolingo di *Play Store* menggunakan algoritma *Multinomial Naïve Bayes* yang diimplementasikan dengan teknik *K-Fold Cross Validation*. Berdasarkan hasil evaluasi model, yang menggunakan pembagian data 80% untuk pelatihan dan 20% untuk pengujian, didapatkan hasil yang cukup memuaskan. Nilai akurasi yang diperoleh sebesar 89%, dengan *precision* 86%, *recall* 91%, dan *f1score* 89%. Kinerja terbaik dicapai pada *fold* ke-4, yang menunjukkan bahwa model ini mampu mengklasifikasikan sentimen ulasan dengan tingkat akurasi yang tinggi. *Confusion Matrix* juga menunjukkan hasil yang konsisten dengan nilai akurasi dan metrik lainnya, yang memperkuat temuan tersebut. Hasil ini mengacu pada standar evaluasi, yang menyatakan bahwa akurasi mendekati 90% dapat dianggap sangat baik, sebagaimana dikemukakan oleh Widiastuti dkk (2024).

Mengingat hasil yang diperoleh, algoritma *Multinomial Naïve Bayes* yang dipadukan dengan teknik evaluasi *K-Fold Cross Validation* terbukti efektif dalam menganalisis sentimen dari teks ulasan pengguna aplikasi Duolingo. Algoritma ini berhasil mengidentifikasi pola-pola penting dalam ulasan yang dapat menunjukkan apakah ulasan tersebut bernada positif atau negatif. Hal ini menjadi bukti bahwa model berbasis probabilistik seperti *Naïve Bayes* sangat relevan untuk pengklasifikasian teks yang bersifat kategorikal, terutama pada aplikasi berbasis analisis sentimen yang membutuhkan keakuratan tinggi.

Namun, untuk penelitian selanjutnya, penulis menyarankan adanya pengembangan pada metode *labelling* yang lebih baik. Model *machine learning* seperti *Naïve Bayes* membutuhkan data latih yang terlabel dengan benar untuk menghasilkan model yang lebih akurat dan dapat diandalkan. Data yang tidak terlabel dengan baik atau tidak representatif dapat mempengaruhi kualitas hasil klasifikasi. Oleh karena itu, memperbaiki proses *labelling* data akan sangat penting untuk meningkatkan kualitas model.

Selain itu, di masa depan, peneliti dapat mengeksplorasi penggunaan algoritma lain yang lebih kompleks seperti *Support Vector Machine* (SVM) atau *Random Forest*. Kedua algoritma ini memiliki potensi untuk meningkatkan akurasi dan memberikan performa yang lebih baik dalam menangani data teks dengan jumlah kata yang besar dan kompleksitas yang tinggi. Penambahan dataset yang lebih variatif dan representatif juga dapat memperbaiki kinerja model dan memungkinkan model untuk lebih adaptif terhadap berbagai jenis ulasan dan konteks aplikasi.

Secara keseluruhan, meskipun penelitian ini sudah menghasilkan model dengan kinerja yang baik, masih terdapat banyak ruang untuk perbaikan dan pengembangan lebih lanjut, terutama dalam hal peningkatan kualitas data dan eksplorasi algoritma lainnya. Diharapkan penelitian selanjutnya dapat memberikan kontribusi lebih besar dalam pengembangan model klasifikasi sentimen yang lebih akurat dan relevan di berbagai domain aplikasi.

DAFTAR PUSTAKA

- Ajay. (2023). Demystifying Naïve Bayes: Log Probability and Laplace Smoothing, from, <https://medium.com/@dancerworld60/demystifying-na%C3%AFve-bayes-log-probability-and-laplace-smoothing-d6da61b0e70b>
- Alida. (2022). Kelebihan dan Kekurangan Word Cloud Sebagai Visualisasi, from, https://www-alida-com.translate.goog/the-alida-journal/the-pros-and-cons-of-word-clouds-as-visualizations?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc

- Anderson, B. (2023). Probabilitas Posterior: desfinisi + contoh, from <https://statorials.org/id/probabilitas-posterior/>
- Gunawan, Y., ER, N., Mahendra, I., Widiartha, I. M., Putra, I. G., & Kadyanan, I. G. (2022). “Analisis Sentimen Ulasan Aplikasi Transportasi Online Menggunakan Multinomial Naïve Bayes dan Query Expansion Ranking”. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 11(1), 121. <https://doi.org/10.24843/jlk.2022.v11.i01.p13>.
- Chugani, V. (2024). A Comprehensive Guide to K-Fold Cross Validation, from, https://www.datacamp.com/tutorial/k-fold-cross-validation?utm_source=chatgpt.com
- Nashirul. (2024). Rusia Tuduh Duolingo Sebarkan Propaganda LGBT, from <https://hidayatullah.com/berita/2024/02/15/267890/rusia-tuduh-duolingo-sebarkan-propaganda-lgbt.html>
- Mulyawan, R. (2023). *Text Preprocessing*: Pengertian, Apa itu *NLTK Library*? Macam Tahapan Basic (Dasar) serta Contoh Simple dan Kodenya, from <https://rifqimulyawan.com/blog/text-preprocessing/>
- Pratama, D., Munandar, T., & Ramdhania, K. (2024). “Multinomial Naive Bayes Algorithm for Indonesian language Sentiment Classification Related to Jakarta International Stadium (JIS)”. *International Journal of Information Technology and Computer Science Applications*, 2(1), 12–22. <https://doi.org/10.58776/ijitcsa.v2i1.118>.
- Ramadhanu, A., Mahessya, R, A., Zaky, M, R., & Isra, M. (2023). “Penerapan Teknologi Machine Learning Dengan Metode Vader Pada Aplikasi Sentimen Tamu Di Hotel Dymens”. *JOISIE (Journal of Information System and Informatics Engineering)*, 7(1), 165-173.
- Sabrani, A., Gede Putu Wirarama Wedashwara, I. W., & Bimantoro, F. (2020). “Metode Multinomial Naive Bayes Untuk Klasifikasi Artikel Online Tentang Gempa di Indonesia”. *JTIKA*, 2(1). <http://jtika.if.unram.ac.id/index.php/JTIKA/>.
- Sayogo, D., Irawan, B., & Bahtiar, A. (2024). “Analisis Sentimen Ulasan Instagram di Google Play Store Menggunakan Algoritma Naïve Bayes”. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3314–3319. <https://doi.org/10.36040/jati.v7i6.8178>.
- Sholekhah, F., Putri, A. D., Rahmaddeni, R., & Efrizoni, L. (2024). “Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom”. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 507–514. <https://doi.org/10.57152/malcom.v4i2.1249>.
- Starmer, J. (2020) *Naive Bayes, Clearly Explained*. Amerika Serikat. StatQuest with John Starmer, from <https://youtu.be/O2L2Uv9pdDA?si=TMEKpTU9KIX3t3cE>
- Suhendra, T., Intan, B., & Martadinata, A. (2024). “Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore Menggunakan Metode Naïve Bayes”. *Jurnal Ilmiah Teknik dan Ilmu Komputer*, 2(3), 1011–1022. <https://doi.org/10.55123/storage.v2i3.2333>.
- Wibowo, G., Widiastuti, S., Muratno, Lolang, E., & Soraya. (2023). “Penerapan Metode Teorema Bayes Dalam Mendiagnosa Penyakit Tuberculosis”. *Building of Informatics, Technology and Science (BITS)*, 4(4), 1782–1788. <https://doi.org/10.47065/bits.v4i4.3035>
- Widiastuti, L., Nurlaela, D., Surniandari, D., Utami, L. (2024). “Penerapan Komparasi Algoritma Klasifikasi Pada Analisis Sentimen Aplikasi Spotify”. *Jurnal Sistem Komputer Musi Rawas (Jusikom)*, 9(1), 23-33.
- Zulfikar, W. B., Atmadja, A. R., & Pratama, S. F. (2023). “Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes”. *Scientific Journal of Informatics*, 10(1), 25–34. <https://doi.org/10.15294/sji.v10i1.39952>.

