

## SENTIMEN ANALISIS APLIKASI DIGITALENT MOBILE MENGGUNAKAN NAÏVE BAYES DAN SVM DENGAN EKSTRAKSI FITUR TF-IDF

### *SENTIMEN ANALYSIS DIGITALENT MOBILE APPLICATION USING NAÏVE BAYES AND SVM WITH TF-IDF FITUR EXTRACTION*

**Jeremi Azero Putra<sup>1</sup>, Alexander Dharmawan<sup>2</sup>, Jutono Gondohanindijo<sup>3</sup>**  
<sup>1,2,3</sup>Universitas AKI (Abdi Karya Indonesia) Semarang, Indonesia  
223200010@student.unaki.ac.id

#### ABSTRACT

*This research discusses the application of sentiment analysis on mobile application reviews using the Naïve Bayes algorithm and Support Vector Machine (SVM) with TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction. Sentiment analysis is the process of extracting information from text to determine the opinions contained therein, which is useful for application developers to understand user feedback. A dataset of 378 mobile app reviews was collected and cleaned before feature extraction using the TF-IDF method, which measures the importance of a word in a document relative to the document set. Next, two machine learning algorithms, Naïve Bayes and SVM, were applied to build a sentiment classification model. The model performance is evaluated using accuracy, precision, recall, and F1-score metrics from the confusion matrix test results. The results showed that the Naïve Bayes algorithm achieved 85.71% accuracy, while SVM achieved 82.14% accuracy. The purpose of this research is to emphasize the importance of algorithm selection and feature extraction techniques in sentiment analysis of mobile applications, as well as providing information for developers in improving application quality based on user feedback.*

**Keyword:** *Sentiment Analysis, Mobile Application, Naïve Bayes, SVM, TF-IDF.*

#### ABSTRAK

Penelitian ini membahas penerapan analisis sentimen pada ulasan aplikasi mobile menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM) dengan ekstraksi fitur TF-IDF (Term Frequency-Inverse Document Frequency). Analisis sentimen adalah proses penggalian informasi dari teks untuk menentukan opini yang terkandung di dalamnya, yang berguna bagi pengembang aplikasi untuk memahami umpan balik pengguna. Dataset ulasan aplikasi mobile berjumlah 378 ulasan yang dikumpulkan dan dibersihkan sebelum diekstraksi fiturnya menggunakan metode TF-IDF, yang mengukur pentingnya sebuah kata dalam dokumen relatif terhadap kumpulan dokumen. Selanjutnya, dua algoritma pembelajaran mesin, Naïve Bayes dan SVM, diterapkan untuk membangun model klasifikasi sentimen. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score dari hasil pengujian confusion matrix. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mencapai akurasi 85.71%, sedangkan SVM mencapai akurasi 82.14%. Tujuan dari penelitian ini adalah menekankan pentingnya pemilihan algoritma dan teknik ekstraksi fitur dalam analisis sentimen aplikasi mobile, serta memberikan informasi bagi pengembang dalam meningkatkan kualitas aplikasi berdasarkan umpan balik pengguna.

**Kata Kunci:** Sentimen Analisis, Aplikasi Mobile, Naïve Bayes, SVM, TF-IDF.

#### PENDAHULUAN

Aplikasi *mobile* menjadi komponen penting dalam kehidupan sehari-hari di dunia yang semakin digital ini, terutama dalam hal pendidikan dan pengembangan keterampilan. Perkembangan teknologi sekarang menjadi bagian penting dari pendidikan (Fitri et al., 2020). Pelatihan online adalah salah satu metode pendidikan yang sedang populer dan sedang menjadi tren saat ini (Kasmanto, 2020). Salah satu aplikasi yang menarik

perhatian dalam penelitian ini adalah Digitalent Kominfo, sebuah platform yang disediakan oleh Kementerian Komunikasi dan Informatika Indonesia untuk pelatihan digital bagi masyarakat. Mengingat pentingnya umpan balik dari pengguna dalam meningkatkan kualitas layanan, analisis sentimen menjadi alat yang sangat berguna (Gunawan et al., 2018).

Analisis sentimen bertujuan untuk mengidentifikasi dan mengklasifikasikan opini pengguna menjadi kategori positif,

negatif, atau netral (Sihombing et al., 2021). Dalam penelitian ini penulis hanya menggunakan kategori positif dan negative. Dengan demikian, penyedia layanan dapat memahami persepsi pengguna dan bisa melakukan perbaikan yang diperlukan.

Sebuah metode klasifikasi dapat digunakan untuk melakukan analisis sentiment (Iskandar & Nataliani, 2021). *Naïve Bayes* dan *Support Vector Machine* (SVM), kedua algoritma atau metode ini sering digunakan untuk klasifikasi teks dalam analisis sentiment (Cindo & Rini, 2019). Kedua algoritma ini memiliki keunggulan masing-masing dalam menangani data teks yang kompleks.

*Naïve Bayes* merupakan salah satu algoritma yang dapat melakukan klasifikasi dengan cepat dan efektif. Selain itu, algoritma ini efisien dalam menganalisis data dalam skala besar (Sihombing et al., 2021). Dalam studi ini, algoritma *Support Vector Machine* (SVM) juga digunakan untuk membantu proses klasifikasi sentimen. Saat digunakan di berbagai industri, SVM menjadi metode pola klasifikasi yang efektif. Karena kemampuan SVM untuk menangani berbagai masalah pembelajaran, banyak anggota komunitas *machine learning* tertarik untuk mempelajari dan mengembangkannya (Suryati et al., 2023).

Penelitian sebelumnya yang dilakukan oleh (Alita & Shodiqin, 2023) dengan judul Sentimen Analisis Vaksin Covid-19 Menggunakan *Naïve Bayes* Dan *Support Vector Machine*, menghasilkan tingkat akurasi sebesar 77,88% untuk metode *Support Vector Machine* dan 72,88% untuk teknik *Naïve Bayes*. Dalam penelitian ini, data twitter dianalisis untuk mengukur opini masyarakat terhadap vaksinasi COVID-19. Klasifikasi positif dan negatif kemudian diidentifikasi menggunakan teknik *Naïve Bayes* dan *Support Vector Machine*.

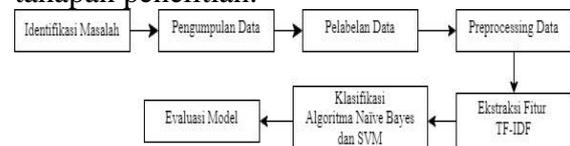
Dalam penelitian ini, *TF-IDF* adalah metode analisis yang berguna dan relevan. Metode statistik yang disebut *TF-IDF*

digunakan untuk mengevaluasi pentingnya sebuah kata di dalam dokumen dalam kaitannya dengan total kumpulan teks. Penerapan metodologi *TF - IDF* dalam analisis teks pasien depresi menghasilkan pemahaman yang lebih baik tentang isu dan pola yang terkandung dalam tulisan mereka, seperti yang ditunjukkan dalam penelitian sebelumnya yang berjudul Analisis Teknis *TF - IDF* dalam Mengidentifikasi Faktor Penyebab Depresi pada Individu (Annisa et al., 2024). Jadi dengan menggabungkan ekstraksi fitur *TF-IDF* dengan algoritma *Naïve Bayes* dan SVM, diharapkan metode ini dapat menghasilkan analisis yang lebih akurat dan informatif mengenai sentimen pengguna terhadap aplikasi Digitalent Kominfo.

Tujuan dari penelitian ini adalah untuk membandingkan kinerja analisis emosional kedua algoritma terhadap evaluasi pengguna aplikasi Digitalent Kominfo. Temuan penelitian ini dapat membantu peneliti mendapatkan pemahaman yang lebih baik tentang perspektif pengguna dan berfungsi sebagai panduan bagi pengembang aplikasi yang ingin meningkatkan pengalaman pengguna dan kualitas layanan.

## METODE

Aplikasi mobile Digitalent Kominfo digunakan dalam penelitian untuk melakukan analisis sentimen menggunakan metodologi yang menyeluruh. Penelitian ini menggunakan bahasa pemrograman *Python*, *Support Vector Machine* (SVM), metode *Naïve Bayes*, dan teknik ekstraksi fitur *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Gambar 1 menunjukkan tahapan penelitian.



Gambar 1. Tahapan Penelitian

## Identifikasi Masalah

Kualitas penelitian sangat dipengaruhi oleh masalah penelitian yang diangkat. Untuk menjelaskan masalah tersebut secara mendalam dan mengukur penjelasannya, perlu dilakukan identifikasi masalah. Oleh karena itu, identifikasi masalah adalah langkah penting dalam mendefinisikan masalah yang ada dalam sebuah penelitian (Wahyuni, 2022).

### **Pengumpulan Data**

Dataset yang digunakan dalam penelitian ini berasal dari evaluasi pengguna Google Play Store terhadap aplikasi Digitalent Mobile. Informasi tersebut diperoleh melalui *scraping* halaman web Google Play Store dengan bahasa pemrograman Python.

### **Pelabelan Data**

Tahap klasifikasi sentimen pada data ulasan pengguna dilakukan dengan mengelompokkan ulasan menjadi dua kategori, yaitu sentimen positif dan negatif. Pelabelan ini dilakukan secara otomatis menggunakan pemrograman Python. Tahap ini bertujuan untuk melatih model yang akan digunakan pada tahap pelatihan data.

### **Preprocessing Data**

*Data preprocessing* adalah langkah pertama dalam pengolahan data yang mengubah data mentah yang sebelumnya diperoleh menjadi data yang lebih terstruktur yang disiapkan untuk analisis lebih lanjut (Rahman Salam et al., 2023). Adapun tahap yang dilakukan dalam preprocessing data, yaitu:

1. *Cleaning*, proses menghilangkan tanda baca dan karakter yang tidak berguna seperti tanda tanya, panggilan, koma, dan titik. URL, HTML, emoji, @/mention, hashtag, dan simbol yang tidak perlu juga dihilangkan (Amira Sumitro et al., n.d.).
2. *Case Folding*, proses isi mengubah semua huruf dalam teks ke dalam huruf kecil, sehingga lebih mudah untuk

membandingkan dan menganalisis teks (Rahman Salam et al., 2023).

3. *Stopword Removal*, proses ini berfungsi untuk menyingkirkan kata-kata yang muncul banyak tetapi tidak benar-benar menyiratkan apa-apa seperti dan, ke, atau, dari, yang, dll.. Menghapus kata-kata yang tidak berarti adalah tujuan utama penghapusan *stopword*, yang juga meningkatkan kinerja pemrosesan (Harmandini & Muslim, 2024).
4. *Tokenization*, proses memecah kalimat yang semula berupa kalimat menjadi potongan-potongan kata berdasarkan tiap kata yang menyusunnya (Suryati et al., 2023).
5. *Stemming*, bertujuan untuk menghilangkan imbuhan di awal dan akhir kata dengan memperhatikan awalan umum dan sufiks yaitu imbuhan yang melekat pada kata dasar (Ade Dwi Dayani et al., 2024).

### **Ekstraksi Fitur**

Metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) merupakan suatu teknik ekstraksi fitur yang memberikan bobot atau nilai pada setiap kata yang muncul dalam teks guna menjamin bahwa suatu istilah relevan dengan suatu dokumen atau frase (Khatib Sulaiman et al., n.d.). Fitur *TF-IDF* juga membantu mengoptimalkan analisis sentimen (Rahman Salam et al., 2023). Dalam pendekatan *TF-IDF*, *Inverse Document Frequencies* (*IDF*) lebih mementingkan seberapa jarang sebuah kata muncul di seluruh kumpulan dokumen, sedangkan *Term Frequencies* (*TF*) lebih mementingkan seberapa sering suatu istilah muncul dalam sebuah dokumen (Gunawan et al., 2018).

### **Klasifikasi**

Dalam pembelajaran mesin, klasifikasi adalah tugas untuk memprediksi label kelas dari sampel berdasarkan sejumlah fitur atau karakteristik yang dimilikinya (Gondohanindijo et al., 2019). Satu set

data pelatihan dan satu set data pengujian akan diekstraksi dari kumpulan data untuk menilai efektivitas algoritma klasifikasi. Model dilatih menggunakan data pelatihan, dan performanya diuji menggunakan data pengujian. Metrik termasuk akurasi, presisi, *recall*, dan *F1-score* yang mengukur seberapa efektif model memprediksi masa depan disertakan dalam *confusion matrix* yang digunakan untuk penilaian (Syahril Dwi Prasetyo et al., 2023).

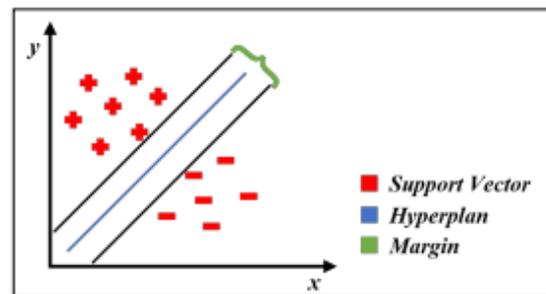
### ***Naïve Bayes***

Teknik klasifikasi langsung yang disebut *Naïve Bayes* menentukan probabilitas dengan memanfaatkan teorema *Bayes* dan frekuensi nilai yang ada dalam database. Dengan menggunakan pengalaman sebelumnya, teorema *Bayes* dapat digunakan untuk memperkirakan peluang di masa depan (Kartika & Gondohanindijo, n.d.). Teknik ini digunakan dalam analisis sentimen untuk memperkirakan sentimen suatu teks, baik positif maupun negatif, berdasarkan karakteristik tertentu. Prosesnya melibatkan mengumpulkan data teks berlabel sentimen dan mengonversi kata-kata dalam teks menjadi fitur, seperti frekuensi kata atau skor TF-IDF. Model *Naïve Bayes* kemudian menghitung probabilitas setiap kata muncul dalam teks dengan sentimen tertentu. *Naïve Bayes* efisien dan efektif untuk analisis sentimen pada dataset besar (Syahril Dwi Prasetyo et al., 2023).

### ***Support Vector Machine (SVM)***

Model klasifikasi yang kedua menggunakan *Support Vector Machine (SVM)* untuk mengelompokkan data. SVM adalah metode dalam pembelajaran mesin yang memisahkan data menjadi dua atau lebih kelompok. Cara kerjanya adalah dengan mencari garis pemisah terbaik, yang disebut *hyperplane*, yang memisahkan data dengan jarak terjauh dari titik-titik terdekat di masing-masing kelompok, yang disebut *support vector*.

Jarak ini disebut *margin*. Tujuan SVM adalah menemukan *hyperplane* yang memberikan *margin* terbesar untuk membuat pengelompokan data lebih akurat. SVM menggunakan tiga jenis kernel: linier, *Radial Basis Function (RBF)*, dan *polynomial* untuk membantu menemukan garis pemisah ini (Khatib Sulaiman et al., n.d.).



**Gambar 2. Klasifikasi SVM**

Sumber : (Khatib Sulaiman et al., n.d.)

### **Evaluasi**

Dalam penelitian ini, untuk mengevaluasi dan memvalidasi kinerja algoritma yang digunakan (Widayati et al., 2024). Pengujian melibatkan metode evaluasi yaitu *confusion matrix*. Pendekatan ini sangat bermanfaat dalam analisis kualitas classifier. Setelah menggunakan *confusion matrix*, nilai-nilai seperti akurasi, *recall*, presisi, dan *f1-score* dapat dihitung dan dipresentasikan dalam bentuk persentase. (Harmandini & Muslim, 2024).

Konsep seperti *True Positive (TP)* adalah hasil klasifikasi yang benar dan *True Negative (TN)* adalah hasil klasifikasi yang tidak akurat dapat digunakan untuk menghasilkan *confusion matrix* (Gondohanindijo et al., n.d.). *False negative (FN)* adalah klasifikasi yang benar yang dianggap salah, sedangkan *False Positive (FP)* adalah hasil klasifikasi yang salah yang dianggap benar (Iskandar & Nataliani, 2021).

Akurasi adalah proporsi data yang diklasifikasikan dengan benar selama pengujian. Nilai akurasi mencerminkan seberapa tepat prediksi model terhadap data yang sudah diketahui labelnya. *Recall* mengukur keakuratan model dalam

mengidentifikasi data positif yang sebenarnya. Rasio positif sejati ( $TP$ ) terhadap semua data positif aktual ( $TP+FN$ ) digunakan untuk menghitung perolehan kembali. Presisi menggambarkan ketepatan model dalam mengklasifikasikan data positif ( $TP$ ) dan negatif ( $FP$ ) dengan benar. Nilai presisi dihitung sebagai rasio antara true positive ( $TP$ ) dengan total prediksi positif yang dilakukan model ( $TP+FP$ ).  $F1$ -Score adalah ukuran rata-rata harmonis dari presisi dan recall (Matheos Sarimole & Ihsan, 2023).

## HASIL DAN PEMBAHASAN

### Pengumpulan Data

Setelah melakukan *scrapping*, data yang didapat sebanyak 376 yang berisi tentang username, at (tanggal), content (isi review), dan score (bintang) dalam rentang waktu bulan Februari 2023 sampai dengan Mei 2024. Data tersebut memiliki nilai positif sebanyak 330 data atau sebesar 87.77%, sedangkan yang memiliki nilai negatif sebanyak 46 data atau 12.23% sebelum dilakukannya pelabelan data.

**Tabel 1. Hasil Pengumpulan Data**

userName	at	content	score
Oktavia	5/18/2024	Niceee	5
Oktavia	4:18:05		
eti	5/8/2024	Baru nyoba	4
mawangsah	13:44:22	pakai, kasih bintang empat dulu	
Nizar	1/24/2024	Bagus dan	5
Islami	12:41:16	keren, bermanfaat untuk generasi masa kini untuk mengikuti pelatihan yg di minat	
Maulida	10/3/2023	Gabisa masuk.	1
Firdausi	23:32:41	Padahal baru aja daftar. Keteranganannya "email atau password salah" mulu	
Rahmat	6/16/2023	Banyak bug	2
Alamsyah	4:20:38	nya, sering logout sendiri dan request time out, saya harap dapat diperbaiki	

### Pelabelan Data

Pengelompokkan kelas atau pelabelan data menjadi positif dan negatif. Data positif berdasarkan komentar yang memiliki score (bintang) 4 sampai 5, sedangkan untuk data negatif adalah berdasarkan score (bintang) 1 sampai 2. Komentar yang memiliki score (bintang) 3 tidak akan dipakai.

### Preprocessing Data

Pada tahap ini, proses dilakukan menggunakan pustaka atau library dalam bahasa pemrograman Python. Preprocessing data melibatkan serangkaian langkah seperti pembersihan (*cleaning*), penyesuaian huruf (*case folding*), penghapusan kata-kata tidak penting (*stopwords removal*), *tokenization*, dan *stemming*. Komentar yang memiliki skor (rating) 3 akan dihapus dari dataset. Kolom username, at, dan score juga akan dihapus untuk menyederhanakan analisis.

Pada tahap ini data yang tidak memiliki nilai atau *null* juga akan dihapus menjadi 338 data. Dikarenakan ketidakseimbangan data antara data positif dan negatif, maka dilakukan pengurangan data sebanyak 200 data supaya data seimbang antara data positif dan negatif. Sehingga dataset berjumlah 138 data yang berisi data positif berjumlah 92 data atau 66.67% dan data negatif berjumlah 46 data atau 33.33%.

**Tabel 3 Hasil Cleaning dan Case Folding**

content	label	Cleaning dan Case Folding
Niceee	Positif	niceee
Baru nyoba pakai, kasih bintang empat dulu	Positif	baru nyoba pakai kasih bintang empat dulu
Bagus dan keren, bermanfaat untuk generasi masa kini untuk mengikuti pelatihan yg di minat	Positif	bagus dan keren, bermanfaat untuk generasi masa kini untuk mengikuti pelatihan yg di minat
Gabisa masuk. Padahal baru aja daftar. Keteranganannya "email atau password salah" mulu	Negatif	gabisa masuk padahal baru aja daftar keteranganannya email atau password salah mulu

Banyak bug nya, sering logout sendiri dan request time out, saya harap dapat diperbaiki	Negatif	banyak bug nya sering logout sendiri dan request time out saya harap dapat diperbaiki
---	---------	---

**Tabel. 4 Hasil Stopword Removal**

<i>Cleaning dan Case Folding</i>	<i>Stopword Removal</i>
niceee	niceee
baru nyoba pakai kasih bintang empat dulu	nyoba pakai kasih bintang empat dulu
bagus dan keren bermanfaat untuk generasi masa kini untuk mengikuti pelatihan yg di minat	bagus keren bermanfaat generasi mengikuti pelatihan yg minat
gabisa masuk padahal baru aja daftar keterangannya email atau password salah mulu	gabisa masuk aja daftar keterangannya email password salah mulu
banyak bug nya sering logout sendiri dan request time out saya harap dapat diperbaiki	bug nya logout request time out harap diperbaiki

**Tabel 5. Hasil Tokenization**

<i>Stopword Removal</i>	<i>Tokenization</i>
niceee	['niceee']
nyoba pakai kasih bintang empat dulu	['nyoba', 'pakai', 'kasih', 'bintang', 'empat', 'dulu']
bagus keren bermanfaat generasi mengikuti pelatihan yg minat	['bagus', 'keren', 'bermanfaat', 'generasi', 'mengikuti', 'pelatihan', 'yg', 'minat']
gabisa masuk aja daftar keterangannya email password salah mulu	['gabisa', 'masuk', 'aja', 'daftar', 'keterangannya', 'email', 'password', 'salah', 'mulu']
bug nya logout request time out harap diperbaiki	['bug', 'nya', 'logout', 'request', 'time', 'out', 'harap', 'diperbaiki']

**Tabel 6. Hasil Stemming**

<i>Tokenization</i>	<i>Stemming</i>
['niceee']	niceee
['nyoba', 'pakai', 'kasih', 'bintang', 'empat', 'dulu']	nyoba pakai kasih bintang empat
['bagus', 'keren', 'bermanfaat', 'generasi', 'mengikuti', 'pelatihan', 'yg', 'minat']	bagus keren manfaat generasi ikut latihan yg minat

['gabisa', 'masuk', 'aja', 'daftar', 'keterangannya', 'email', 'password', 'salah', 'mulu']	gabisa masuk aja daftar terang email password salah mulu
['bug', 'nya', 'logout', 'request', 'time', 'out', 'harap', 'diperbaiki']	bug nya logout request time out harap baik

**Ekstraksi Fitur**

Setelah *preprocessing data*, akan dilakukan proses penghitungan bobot setiap kata. Metode yang digunakan adalah TF-IDF. Untuk penghitungan bobot akan dilakukan secara otomatis menggunakan *library* pada Bahasa pemrograman python yaitu *scikitlearn* dan yang akan divisualisasikan ada 10 kata dengan nilai tertinggi pada table 7.

**Tabel 7. Hasil Penghitungan Bobot Kata TF-IDF**

<b>Kata</b>	<b>Score</b>
good	0.085661
bagus	0.064559
mantap	0.048634
ok	0.045070
aplikasi	0.038296
bantu	0.035572
manfaat	0.034718
keren	0.029548
baik	0.028861
latih	0.025968

**Klasifikasi Naïve Bayes dan SVM**

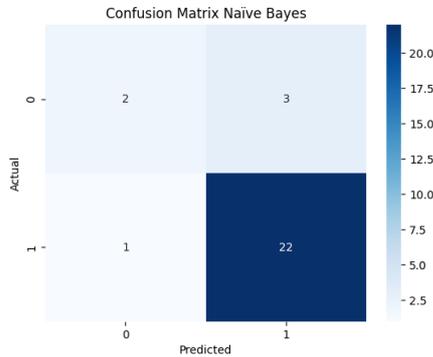
Pada proses ini akan dilakukan pemodelan klasifikasi *Naïve Bayes* dan SVM dengan menggunakan bahasa pemrograman python. Data latih dan uji akan dipisahkan dari dataset dengan perbandingan 80% untuk data pelatihan dan 20% untuk data pengujian. Data pengujian akan digunakan untuk menilai model setelah dilatih menggunakan data pelatihan selama fase pemodelan. Tingkat akurasi, presisi, *recall*, dan *F1-Score* ditampilkan dalam hasil klasifikasi model data pengujian, yang dievaluasi menggunakan matriks konfusi.

**Evaluasi**

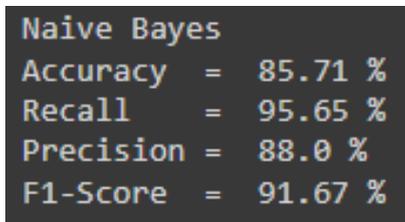
Setelah proses klasifikasi model, hasil dari pengujian model akan dilakukan menggunakan *confussion matrix* untuk

mengetahui berapa data yang berhasil memprediksi sentimen terhadap data yang data uji yang sudah berlabel. Lalu akan menghitung nilai akurasi, presisi, recall, dan *F1-Score* dari *confussion matrix*.

### Naïve Bayes

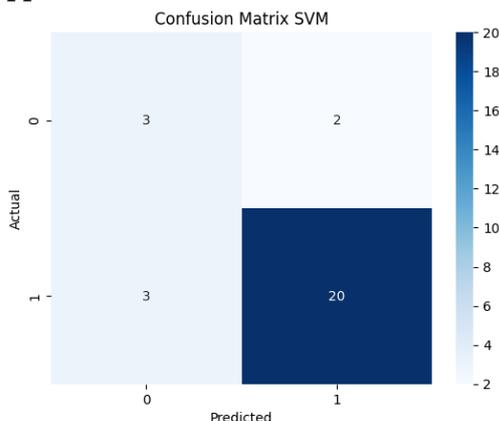


Gambar 3. Hasil *Confussion Matrix Naïve Bayes*

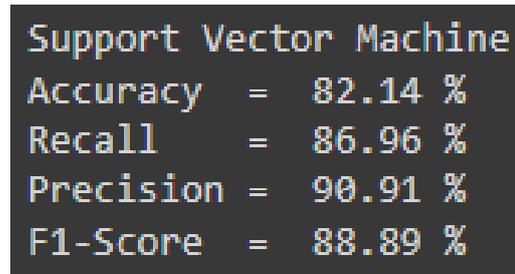


Gambar 4. *Classification Report Naïve Bayes*  
 Berdasarkan hasil dari *confussion matrix* pada gambar 3, model *Naïve Bayes* dapat memprediksi 22 data positif dan 2 data negatif dengan benar dari 28 data uji. Berdasarkan gambar 4, model *Naïve Bayes* memiliki tingkat akurasi sebesar 85.71%.

### Support Vector Machine



Gambar 5. Hasil *Confussion Matrix SVM*



Gambar 6. *Classification Report SVM*

Hasil dari *confussion matrix* pada gambar 5, model *SVM* dapat memprediksi 20 data positif dan 3 data negatif dengan benar dari 28 data uji. Berdasarkan gambar 6, model *SVM* memiliki tingkat akurasi sebesar 82.14%.

### Wordcloud

*Wordcloud* adalah gambar data teks yang membuat visualisasi bebas teks menjadi lebih mudah. Istilah-istilah yang paling sering dimasukkan dalam kumpulan data akan ditampilkan lebih menonjol dibandingkan istilah-istilah lainnya (Suryati et al., 2023). Dalam studi ini, *wordcloud* digunakan untuk menampilkan kata-kata yang diklasifikasikan sebagai positif dan negatif setelah melalui proses *stemming* dalam *preprocessing data*.



Gambar 7. *Wordcloud Berlabel Positif*



Gambar 8. *Wordcloud Berlabel Negatif*

### SIMPULAN

Dari penelitian yang sudah dilakukan dalam sentimen analisis dengan

algoritma *Naive Bayes* dan *SVM* menggunakan ekstraksi fitur *TF-IDF*, dapat diambil kesimpulannya sebagai berikut:

1. Sentimen *review* pengguna, dari 378 *review* ada 87.77% mereview positif yang berisikan seperti bermanfaat, sangat membantu, bagus, dll. Untuk *review* negatif ada 46 *review* atau 12.23% yang berisikan keluhan pengguna seperti *error* dan *bug*.
2. Metode *Naive Bayes* menghasilkan akurasi sebesar 85,7%, *F1-score* sebesar 91,67%, *recall* sebesar 95,65%, dan presisi sebesar 88%.
3. Algoritma *SVM (Support Vector Machine)* mencapai tingkat akurasi sebesar 82.14%, *recall* sebesar 86.96%, presisi sebesar 90.91%, dan *F1-score* sebesar 88.89%.
4. Berdasarkan hasil pengujian dua algoritma diatas terhadap *review* pengguna aplikasi *mobile* Digitalent, algoritma *Naive Bayes* memiliki nilai yang lebih tinggi dibandingkan algoritma *SVM*

#### DAFTAR PUSTAKA

Ade Dwi Dayani, Yuhandri, & Widi Nurcahyo, G. (2024). Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. *Jurnal KomtekInfo*, 1–10. <https://doi.org/10.35134/komtekinfo.v1i1.439>

Alita, D., & Shodiqin, R. A. (2023). Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 1–12. <https://doi.org/10.58602/jaiti.v1i1.20>

Amira Sumitro, P., Iskandar Mulyana, D., Saputro, W., Teknologi Informasi, J., Cipta Karya Informatika, S., Teknik Informatika, J., & Eresha, S. (n.d.). *Terbit online pada laman web*

*jurnal:*

<https://ejournalunsam.id/index.php/jicom/> Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based. <https://developer.twitter.com>

Annisa, L., Kalifia, A. D., Bisnis, F., Humaniora, D., & Yogyakarta, U. T. (2024). Volume 2; Nomor 1. Januari, 302–307. <https://doi.org/10.59435/gjmi.v2i1.249>

Cindo, M., & Rini, D. P. (2019). Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Literatur Review: Metode Klasifikasi Pada Sentimen Analisis. In Januari. <https://seminar-id.com/semnas-sainteks2019.html>

Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *TRANSFORMTIKA*, 18(1), 71–80. [www.nusamandiri.ac.id](http://www.nusamandiri.ac.id),

Gondohanindijo, J., Noersasongko, E., Pujiono, Muljono, Fanani, A. Z., Affandy, & Basuki, R. S. (2019). Comparison Method in Indonesian Emotion Speech Classification. *2019 International Seminar on Application for Technology of Information and Communication (ISEMANTIC)*, 230–235. <https://doi.org/10.1109/ISEMANTIC.2019.8884298>

Gondohanindijo, J., Noersasongko, E., & Rosal Moses Setiadi, D. (n.d.). Multi-Features Audio Extraction for Speech Emotion Recognition Based on Deep Learning. In *IJACSA International Journal of Advanced Computer Science and Applications* (Vol. 14, Issue 6). [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)

Gunawan, B., Sasty, H., #2, P., Esyudha, E., & #3, P. (2018). *JEPIN (Jurnal*

- Edukasi dan Penelitian Informatika) Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes.* 4(2), 17–29. [www.femaledaily.com](http://www.femaledaily.com)
- Harmandini, K. P., & Muslim, K. (2024). Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment. *Jurnal Dan Penelitian Teknik Informatika*, 8(2). <https://doi.org/10.33395/v8i2.13376>
- Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1120–1126. <https://doi.org/10.29207/resti.v5i6.3588>
- Kartika, E., & Gondohanindijo, J. (n.d.). *RANCANG BANGUN MODEL SENTIMEN ANALISIS REVIEW PRODUK PADA TOKO ONLINE MENGGUNAKAN NAIVE BAYES.*
- Kasmanto, R. (2020). Analisis pelatihan online teknis big data menggunakan data logger Moodle. *Jurnal Penelitian Ilmu Pendidikan*, 13(2), 137–146. <https://doi.org/10.21831/jpipfip.v13i2.29419>
- Khatib Sulaiman, J., Antibiotik di Indonesia Herdianti Darwis, P., Wanaspati, N., Anraeni, S., & Artikel Abstrak, I. (n.d.). Support Vector Machine untuk Analisis Sentimen Masyarakat Terhadap. *Indonesian Journal of Computer Science Attribution*, 12(4), 2196.
- Matheos Sarimole, F., & Ihsan, A. N. (2023). ANALISIS SENTIMEN TWITTER TERHADAP UU CIPTA KERJA DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE TWITTER SENTIMENT ANALYSIS OF THE CIPTA KERJA LAW USING THE NAÏVE BAYES ALGORITHM AND SUPPORT VECTOR MACHINE. *Journal of Information Technology and Computer Science (INTECOMS)*, 6(2).
- Rahman Salam, R., Fajri Jamil, M., Ibrahim, Y., Rahmaddeni, Soni, & Herianto. (2023). *MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine.* 3, 27–35.
- Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 233–242. <https://doi.org/10.29408/edumatic.v5i2.4089>
- Suryati, E., Ari Aldino, A., Penulis Korespondensi, N., & Suryati Submitted, E. (2023). *Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM).* 4(1), 96–106. <https://doi.org/10.33365/jtsi.v4i1.2445>
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 1–7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- Wahyuni, W. (2022). Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode Naive Bayes. *Jurnal Informatika Ekonomi Bisnis*, 148–153. <https://doi.org/10.37034/infeb.v4i4.162>

Widayati, Y. T., Widjaja, S., Wicaksono, A. P., Gondohanindijo, J., & Putri, C. C. (2024). Decision Tree Implementation in IT Job Recommendation System. *Jurnal Transformatika*, 21(2), 84. <https://doi.org/10.26623/transformatika.v21i2.8328>