

Perbandingan Algoritma *Naïve Bayes* dan *Support Vector Machine* dalam Analisis Sentimen Aplikasi Teman Bus

Fitri Aurellia Soraya¹, Aries Dwi Indriyanti²

^{1,3} Program Studi S1 Sistem Informasi, Universitas Negeri Surabaya

¹fitri.20060@mhs.unesa.ac.id

³ariesdwi@unesa.ac.id

Abstrak— Penelitian ini mengevaluasi performa dua algoritma klasifikasi, yakni *Naïve Bayes* dan *Support Vector Machine* (SVM), dalam analisis sentimen ulasan pengguna terhadap aplikasi Teman Bus di *Google Play store*. Meskipun kedua algoritma ini telah digunakan pada analisis sentimen pada penelitian sebelumnya, belum ada studi yang secara langsung membandingkannya dalam konteks aplikasi Teman Bus. Analisis sentimen ini melibatkan pemrosesan teks dan klasifikasi sentimen untuk menilai respons pengguna terhadap layanan bus tersebut.

Hasil perbandingan mengungkapkan perbedaan performa antara *Naïve Bayes* dan *Support Vector Machine* pada klasifikasi sentimen. SVM dengan kernel RBF menunjukkan akurasi sebesar 85%, lebih baik dalam menangani pola sentimen yang kompleks dan tidak linier dibandingkan dengan *Naïve Bayes* yang mencapai akurasi 82%, khususnya pada dataset dengan rasio pembagian data 30:70.

Penelitian ini memiliki tujuan agar memberikan informasi yang lebih mendetail mengenai evaluasi pengguna terhadap layanan Teman Bus menggunakan ulasan yang ditemukan di *Google Play store* serta memberikan wawasan berguna untuk memilih algoritma yang tepat dalam tugas serupa di masa depan. SVM dengan kernel RBF cenderung menjadi pilihan yang lebih baik untuk analisis sentimen aplikasi Teman Bus, meskipun pemilihan algoritma terbaik tetap bergantung pada konteks dan karakteristik data yang ada.

Kata Kunci— Analisis Sentimen, *Naïve Bayes*, *Support Vector Machine*, Aplikasi Teman Bus, Ulasan, *Google Play store*.

I. PENDAHULUAN

Kemacetan adalah keadaan di mana aliran lalu lintas terhambat karena jumlah kendaraan yang melampaui kapasitas jalan. Isu ini sering terjadi di kota-kota besar yang memiliki sistem transportasi publik yang kurang memadai, sehingga menyebabkan gangguan dalam produksi dan distribusi barang serta memengaruhi kesejahteraan fisik dan mental pekerja. Berdasarkan data dari Numbeo, Indonesia menempati peringkat ke-17 dalam daftar negara dengan tingkat kemacetan terburuk di dunia, dengan indeks sebesar 246,6.

Transportasi memegang peran penting dalam menunjang aktivitas sehari-hari dan pertumbuhan ekonomi. Di perkotaan, transportasi umum seperti bus sangat penting untuk mobilitas penduduk. Untuk mengurangi kemacetan, Kementerian Perhubungan RI mengembangkan sistem transportasi umum bus, termasuk aplikasi seluler Teman Bus yang memungkinkan pengguna melacak lokasi bus, mengakses jadwal perjalanan, dan melakukan pembayaran tanpa uang tunai. Namun, aplikasi ini sering mendapatkan keluhan,

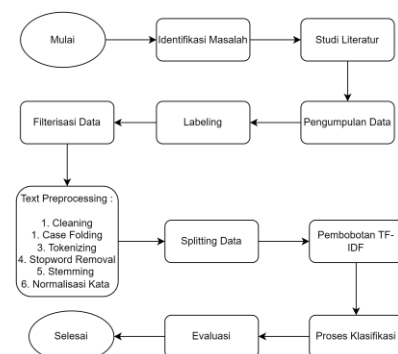
terutama terkait pelacakan dan ketepatan waktu bus, serta memiliki rating rendah di *Google Play store*, yaitu 2,7 dari 5.

Untuk memahami dan meningkatkan kualitas layanan Teman Bus, analisis ulasan pengguna di *Google Play store* sangat penting. Ulasan ini mencakup aspek-aspek seperti kenyamanan, keandalan, dan kecepatan layanan. Analisis sentimen ulasan pengguna dapat memberikan wawasan berharga bagi penyedia layanan mengenai aspek yang perlu diperbaiki[1].

Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma klasifikasi yang sering digunakan dalam analisis sentimen: *Naïve Bayes* dan *Support Vector Machine* (SVM). Meskipun kedua algoritma ini sering digunakan pada analisis sentimen, belum ada perbandingan kinerja dalam konteks ulasan pengguna Teman Bus. Output penelitian ini penulis berharap bisa memberi pemahaman lebih mendalam mengenai evaluasi pengguna terhadap Teman Bus dan memberikan wawasan dalam memilih algoritma yang tepat untuk tugas analisis sentimen serupa di masa depan, serta membantu penyedia layanan dalam meningkatkan kualitas layanannya

II. METODOLOGI PENELITIAN

Penelitian ini mengadopsi proses dan tujuan Knowledge Discovery in Database (KDD). KDD pada dasarnya merupakan proses yang melibatkan pengumpulan dan eksploitasi data historis guna mengidentifikasi pola dan keteraturan dalam volume besar database. Proses KDD dapat memanfaatkan metode pada statistik, kecerdasan buatan, matematika, dan pembelajaran mesin dalam mengungkap data atau informasi baru. [2].



Gbr 1 Alur Penelitian

A. Identifikasi Masalah

Dalam fase permulaan penelitian ini, peneliti mengenali masalah yang berkaitan dengan perbandingan teknik *Naïve Bayes* dan *Support Vector Machine* dalam menganalisa sentimen pemakai aplikasi Teman Bus. Dengan membandingkan kedua metode ini, diharapkan dapat meningkatkan efektivitas dalam menganalisis sentimen pengguna aplikasi Teman Bus.

B. Studi Literatur

Studi literatur ini memaparkan hasil dan kerangka konseptual dari penelitian sebelumnya. Peneliti meninjau jurnal, tesis, dan sumber-sumber relevan lainnya terkait aplikasi Teman Bus. Tujuan utama studi literatur ini adalah untuk mengidentifikasi tren, pemahaman, dan isu-isu penting yang telah dibahas dalam penelitian terdahulu.

C. Pengumpulan Data

Dalam tahap pengumpulan data, peneliti memanfaatkan teknik web scraping untuk mengambil ulasan yang diberikan dari pemakai terhadap aplikasi Teman Bus dari platform *Google Play store*. Data yang diambil disimpan dalam bentuk tabel atau spreadsheet pada file lokal atau database untuk diproses lebih lanjut. Gambar flowchart web scraping menggambarkan alur kerja dari proses pengumpulan data ulasan tersebut.

D. Labeling

Setelah tahap web scraping, dilakukan pelabelan data di mana ulasan pengguna aplikasi Teman Bus diberi label komentar, dan sentimen ulasan tersebut diberi label positif atau negatif.

E. Filterisasi Data

Filterisasi data adalah salah satu langkah dalam pre-processing yang bertujuan untuk mendapatkan teks yang relevan dan menghapus teks serta kalimat yang tidak relevan atau tidak diperlukan dalam analisis. [3].

F. Text Preprocessing

Pre-processing merupakan tahapan yang krusial dalam data mining, yang melibatkan konversi data mentah menjadi format yang sesuai dengan prosedur mining yang akan dijalankan. Tahapan ini dianggap sebagai langkah yang paling krusial dalam proses data mining.

1) *Cleansing*

merupakan tahapan yang dilakukan untuk menghilangkan noise atau elemen-elemen yang tidak dibutuhkan dalam suatu dokumen dengan tujuan untuk memperbaiki dan meningkatkan kualitas data [4].

2) *Case Folding*

menjadikan huruf semua huruf kecil, mengubah semua huruf besar menjadi huruf kecil. [5].

3) *Tokenizing*

tindakan membagi rangkaian karakter ke dalam bagian lebih kecil (kata) yang disebut token.

Tujuannya untuk mendapatkan fitur dari setiap dokumen setelah diberikan bobot nilai[6].

4) *Stopword Removal*

Stopword removal ialah tahap eliminasi kata-kata yang tak mempunyai dampak signifikan pada proses klasifikasi, contohnya kata depan, kata sambung, dan sejenisnya. [4].

5) *Stemming*

Tahap ini bertujuan dalam mengurangi total entitas berbeda pada dataset dan mengelompokkan kata-kata dasar dengan makna serupa namun dengan variasi bentuk karena imbuhan yang berbeda. [5].

G. Splitting Data

Splitting Data, juga dikenal sebagai pemisahan data, adalah langkah penting dalam proses pengembangan model. Tujuan pemisahan data adalah untuk membagi dataset menjadi dua atau lebih subset yang berbeda untuk keperluan pelatihan dan pengujian model[7].

H. Normalisasi Kata

Normalisasi kata adalah proses koreksi atau penggantian kata-kata yang salah pengejaan atau disingkat ke dalam format standar. Tujuannya adalah untuk mengurangi jumlah dimensi kata yang berbeda, yang dapat terjadi jika kesalahan ejaan atau singkatan tidak diperbaiki. Meskipun kata-kata tersebut memiliki arti yang sama, jika tidak dinormalisasi, dalam matriks, mereka dianggap sebagai entitas yang berbeda[8].

I. Pembobotan TF-IDF

Dalam penelitian ini, pembobotan kata diterapkan dengan memberikan nilai atau bobot pada setiap kata. Metode yang digunakan adalah TF-IDF karena dianggap lebih efektif dalam perhitungan. TF-IDF merupakan proses pembobotan kata yang mengalikan nilai frekuensi term (TF) dengan nilai inverse document frequency (IDF). [10].

J. Proses Klasifikasi

Penelitian ini menerapkan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) untuk menguji performa klasifikasi dalam dua kelas memakai data latih dan data pengujian. *Naïve Bayes* termasuk dalam kategori supervised learning karena memerlukan bimbingan dalam pelatihan data. Kinerja *Naïve Bayes* diakui memiliki waktu klasifikasi yang cepat, yang membantu meningkatkan efisiensi sistem analisis sentimen. Rumus *Naïve Bayes* adalah sebagai berikut [11] :

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan :

X = Data kelas tidak diketahui

H = Data hipotesis X adalah kelas yang terpisah

P(C|X) = Probabilitas hipotesis H berdasarkan kondisi

P(H) = Probabilitas hipotesis H

P(X|C) = Probabilitas X berdasarkan kondisi pada hipotesis

P(X) = Probabilitas X

SVM dipakai dalam mengenali pola pada data yang sudah dilabeli, membuat model pembelajaran mesin, dan mengaplikasikannya pada data tambahan yang tidak dilabeli, seperti tweet tambahan, untuk meramal sentimen apakah negatif atau positif.. [11].

K. Evaluasi

Dalam evaluasi, digunakan instrumen bernama *Confusion Matrix* dalam memperhitungkan nilai akurasi, presisi, recall, dan nilai f-score. Ini dilakukan untuk menguji kinerja model. True Positive (TP) menggambarkan jumlah prediksi positif yang benar dari kelas yang faktanya positif. False Positive (FP) menggambarkan jumlah perkiraan positif yang salah dari kelas yang faktanya negatif. True Negative (TN) menggambarkan jumlah perkiraan negatif yang benar dari kelas yang faktanya negatif. False Negative (FN) menggambarkan jumlah prediksi negatif yang salah dari kelas yang sebenarnya positif. [11].

TABEL I
 CONFUSION MATRIX

		Prediksi	
		Negatif	Positif
Akurasi	Negatif	True Negative (TN)	False Positive (FP)
	Positif	False Negative (FN)	True Positive (TP)

Area Under Curve (AUC) merupakan metrik yang dipakai dalam memperhitungkan kinerja model pada membedakan atau mendiskriminasi antara kelas-kelas yang berbeda menggunakan perkiraan probabilitas hasil dari sampel yang tentukan secara acak dari populasi positif dan negatif. Nilai AUC ada pada range dari 0 hingga 1, yang mana nilai 1 menjelaskan kelompok yang sempurna, dan nilai 0.5 menjelaskan kelompok yang acak. Dalam konteks ini, klasifikasi dianggap baik jika nilai AUC semakin mendekati 1, yang menjelaskan bila model mempunyai kecakapan yang baik untuk membedakan antara kelas-kelas yang berbeda. [12].

TABEL III
 NILAI AREA UNDER CURVE

Nilai	Keterangan
0,91 – 1,00	Klasifikasi Sangat Baik
0,81 – 0,90	Klasifikasi Baik
0,71 – 0,80	Klasifikasi Cukup
0,61 – 0,70	Klasifikasi Buruk
<0,6	Klasifikasi Salah

Pada bagian ini, peneliti nantinya menjelaskan hasil dari penelitian yang sudah dilaksanakan. Pembahasan ini, akan diketahui ulasan dari aplikasi Teman Bus pada *Google Play store* memiliki komentar terbanyak negatif atau positif. Selain itu, peneliti akan merekomendasikan metode terbaik dari kedua tekni yakni *Naïve Bayes* dan *Support Vector Machine*..

A. Pengumpulan Data

Data yang dipakai dalam penelitian ini berasal dari ulasan aplikasi Teman Bus yang terdapat di *Google Play store*. Peneliti menggunakan teknik web scraping menggunakan library Pandas di lingkungan pemrograman *Python*, khususnya di platform *Google Colab*.

	userName	score	at	content
198	Christian Hadinata	2	2023-07-03 11:01:14	Di regis kartu khusus buat pelajar kasih opsi ...
197	Cjshs Rd	5	2023-07-03 13:21:24	Sangat puas
196	Adit Akbar	3	2023-07-04 00:51:03	Kan katanya pelajar lansia sama apa ya lupa po...
195	Afian	1	2023-07-04 13:37:11	Tak berfungsi sama sekali, setiap pilih kota s...
194	Arif Julianto Prasetyo	1	2023-07-05 00:58:57	Kota nya dikit aplikasi ga jelas

Gbr 2 Hasil Komentar

B. Pelabelan Data

Pada tahap ini nantinya dilaksanakan pelabelan dari data tahap sebelumnya. Data dapat dilabeli dengan cara melihat dari rating yang diberikan oleh pengguna pada aplikasi Teman Bus. Rating dari pengguna dapat dijadikan pedoman untuk melabeli data

≤ 3	Negatif
4	Positif
5	Positif

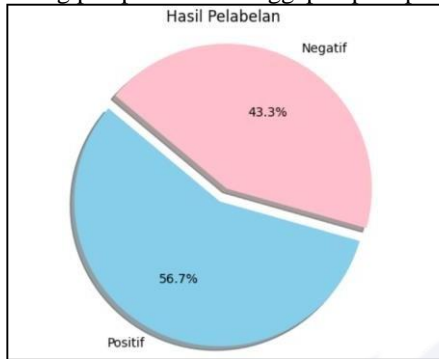
Dengan tabel 2 akan didapatkan hasil komentar seperti Gambar

content	Label
Di regis kartu khusus buat pelajar kasih opsi ...	Negatif
Sangat puas	Positif
Kan katanya pelajar lansia sama apa ya lupa po...	Positif
Tak berfungsi sama sekali, setiap pilih kota s...	Negatif
Kota nya dikit aplikasi ga jelas	Negatif
Good	Positif
Apk tidak sinkron	Negatif
Good	Positif
BAGUS Banget..	Positif
Jelek banget ui ux nya	Negatif
KENAPA YAH SUDAH BANYAK KELUHAN TAPI MASIH TID...	Negatif
Teman bus ada jadwal istirahatnya yaa?setiap N...	Negatif
Posisi kendaraan tidak sesuai dengan info di a...	Negatif
Banyak kekurangan	Negatif
Sebenarnya teman bus ini sangat membantu sekal...	Negatif

Gbr 3 Pelabelan

III. HASIL DAN PEMBAHASAN

Dari hasil diatas didapatkan dapat dilihat analisis visual yang disajikan dalam Gambar 4 memberikan pemahaman yang lebih mendalam tentang perspektif dan tanggapan para pengguna.



Gbr 4 Visualisasi Komentar Positif dan Negatif

C. Filterisasi Data

Beberapa komentar telah disaring karena tidak relevan, seperti komentar yang berhubungan dengan antarmuka pengguna, bug, atau kesalahan, serta kalimat yang tidak sesuai dengan keluhan atau sentimen pengguna.

D. Text Preprocessing

Dengan melakukan ini, kita memastikan ketersediaan alat yang diperlukan untuk mengoptimalkan dan mempermudah.

1) *Cleansing*, bertujuan untuk menghapus kata berulang atau kata dasar

content	Label	text_cleaning
Baru di install kok error	Negatif	Baru di install kok error
Nice	Positif	Nice
Kok gak bs di daftar ya	Negatif	Kok gak bs di daftar ya
Oke	Positif	Oke
Karna teman bus .saya bisa ke bandara Adisucipt...	Positif	Karna teman bus saya bisa ke bandara Adisucipt...
Terima kasih sudah menghadirkan transportasi u...	Positif	Terima kasih sudah menghadirkan transportasi u...
Menyenangkan	Positif	Menyenangkan
Good	Positif	Good
Hai teman bus cuma kasih saran untuk pembayara...	Negatif	Hai teman bus cuma kasih saran untuk pembayara...
idak bisa baca lokasi padahal GPS aktif	Negatif	Tidak bisa baca lokasi padahal GPS aktif

Gbr 5 Cleaning Data

2) *Case Folsing*, teknik menyamaratakan huruf kapital atau setiap huruf atau teks menjadi huruf kecil, dengan huruf "a" sampai "z" yang diizinkan.

text_cleaning	text_casfolding
Baru di install kok error	baru di install kok error
Nice	nice
Kok gak bs di daftar ya	kok gak bs di daftar ya
Oke	oke
Karna teman bus saya bisa ke bandara Adisucipt...	karna teman bus saya bisa ke bandara adisucipt...
Terima kasih sudah menghadirkan transportasi u...	terima kasih sudah menghadirkan transportasi u...
Menyenangkan	menyenangkan
Good	good

Gbr 6 Case Folding

3) *Stopword Removal*, memperkuat kemampuan dan kecepatan analisis sentimen dengan meminimalisir jumlah data yang perlu diproses

text_casfolding	text_stopword
baru di install kok error	install error
nice	nice
kok gak bs di daftar ya	gak bs daftar ya
oke	oke
karna teman bus saya bisa ke bandara adisucipta...	karna teman bus bandara adisucipto yogyakarta ...
terima kasih sudah menghadirkan transportasi u...	terima kasih menghadirkan transportasi bali sa...

Gbr 7 Stopword Removal

4) *Tokenizing*, mendekonstruksi teks menjadi elemen-elemen dasar seperti kata-kata atau frasa, memudahkan pelaksanaan analisis sentiment

text_stopword	text_token
install error	[install, error]
nice	[nice]
gak bs daftar ya	[gak, bs, daftar, ya]
oke	[oke]
karna teman bus bandara adisucipto yogyakarta ...	[karna, teman, bus, bandara, adisucipto, yogya...]

Gbr 8 Tokenizing

5) *Stemming*, menyederhanakan teks menjadi bentuk dasar kata, meningkatkan kemampuan mengelola variasi kata tanpa mengurangi keakuratan analisis sentiment

text_stopword	text_token	text_steamindo
install error	['install', 'error']	install error
nice	['nice']	nice
gak bs daftar ya	['gak', 'bs', 'daftar', 'ya']	gak bs daftar ya
oke	['oke']	oke
karna teman bus bandara adisucipto yogyakarta ...	['karna', 'teman', 'bus', 'bandara', 'adisucipto', 'yogyakarta']	karna teman bus bandara adisucipto yogyakarta ...

Gbr 9 Stemming

6) *Splitting Data*

Dilakukan *splitting data* dengan rasio 70:30 dan 80:20 karena sebelumnya telah dicoba *splitting data* dengan rasio lainnya, yaitu 60:40. Namun, hasil dari *splitting data* dengan rasio 60:40 kurang maksimal. Oleh karena itu, diputuskan untuk menggunakan rasio 70:30 dan 80:20 guna mendapatkan hasil yang lebih optimal.

E. *Normalisasi Kata*

Mengubah kata-kata menjadi bentuk standar atau konsisten agar lebih mudah diproses

Data Awal	Hasil Normalisasi
yg	yang
tdk	tidak
bs	bisa
apk	aplikasi
bgt	banget
bgs	bagus
udh	udah
knp	kenapa
klo	kalau
gk	tidak
tp	tapi

Gbr 10 Kamus Normalisasi

F. *Pembobotan TF-IDF*

Dalam modul pembobotan kata TF-IDF, kita akan menghitung nilai atau bobot untuk setiap kata atau fitur. Proses implementasi pembobotan kata TF-IDF dilakukan dengan menghitung nilai term frequency (TF) dikalikan nilai inverse document frequency (IDF). TF menghitung kemunculan kata dalam sebuah dokumen, sementara IDF mengukur seberapa penting kata ini pada segala koleksi dokumen. Dengan mengalikan kedua nilai ini, kita mendapatkan bobot yang menggambarkan pentingnya kata tersebut dalam dokumen spesifik dan koleksi dokumen secara keseluruhan.

Tabel III menjelaskan bahwa "term" mengacu pada kata-kata yang sering kali muncul dalam ulasan komentar, sementara D1, D2, dan D3 menunjukkan seberapa pentingnya setiap kata tersebut dalam komentar.

TABEL III
TF-IDF

Term	W = TF * IDF				
	D1	D2	D3	D4	D5
Aplikasi	0	0,2267	0	0	0,1125
Bus	0	0	0	0,1641	0
Teman	0	0	0	0	0
Bagus	0	0	0	0	0,1338
Bantu	0	0	0,2598	0	0
Baik	0	0	0	0	0,7458
Buka	0	0,3667	0	0	0
Halte	0,5311	0	0	0,2545	0,1600
Nyaman	0	0	0	0	0,66611
Tolong	0	0	0	0	0,1644
Jadwal	0	0	0	0	0



Pada Gambar 11, tampak hasil visualisasi WordCloud yang dibentuk berdasarkan bobot Term Frequency-Inverse Document Frequency (TF-IDF) yang dihasilkan dari tahap vektorisasi teks. WordCloud ini memberi deskripsi visual yang menarik mengenai frekuensi kemunculan kata-kata pada teks, yang mana ukuran kata-kata dalam WordCloud mencerminkan bobot TF-IDF mereka. Hal ini menggambarkan pentingnya kata-kata tersebut dalam dokumen atau korpus teks yang digunakan, di mana kata-kata yang lebih sering muncul dan mempunyai nilai TF-IDF yang lebih tinggi akan dimunculkan dengan ukuran yang lebih besar dalam WordCloud.

G. Klasifikasi

Pada proses ini, dilaksanakan pengelompokan memakai dua algoritma, yaitu *Support Vector Machine* dan *Naïve Bayes*, untuk menganalisis pola data yang diperoleh.

Peneliti menguji tiga kernel yang berbeda untuk mengamati pengaruhnya terhadap klasifikasi data dalam algoritma *Support Vector Machine*. Kernel yang diujikan yakni :

TABEL IVV
 KERNEL YANG DIUJIKAN

Kernel yang diujikan	Radial Basis Function (RBF)
	Sigmoid

H. Evaluasi

Pada tahap evaluasi algoritma *Naïve Bayes* dan *Support Vector Machine*, dilaksanakan analisis memakai metrik evaluasi seperti *Confusion Matrix*, F1 score, dan akurasi. Dengan memperhatikan metrik-metrik ini, kinerja kedua algoritma dievaluasi dan dibandingkan untuk memilih algoritma yang paling sesuai dengan kebutuhan analisis data.

1) Naïve Bayes

Untuk rasio 70:30, hasil dari percobaan tersebut ditampilkan dalam V. Tabel V menampilkan

Confusion Matrix yang menggambarkan jumlah prediksi yang benar dan salah untuk setiap kelas pada model yang diuji menggunakan rasio pembagian 70:30.

TABEL V
 CONFUSION MATRIX NAIVE BAYES RASIO 70:30

	Negatif	Positif
Negatif	166	41
Positif	50	251

Tabel VI memperlihatkan hasil evaluasi metrik seperti akurasi, presisi, recall, dan F1-Score dari Algoritma *Naïve Bayes*.

TABEL VI
 CONFUSION MATRIX NAIVE BAYES RASIO 70:30

Rasio	Naive Bayes			
	Akurasi	Presisi	Recall	F1-Score
30 : 70	82%	77%	82%	79%

2) *Support Vector Machine Kernel Sigmoid*

Untuk rasio pembagian 70:30 dan penggunaan kernel sigmoid, hasil percobaan ditampilkan dalam Tabel VII yang menggambarkan *Confusion Matrix*. *Confusion Matrix* ini mencatat jumlah prediksi yang tepat dan yang salah untuk setiap kelas pada model yang diuji dengan rasio 70:30.

TABEL VII
 CONFUSION MATRIX SUPPORT VECTOR MACHINE SIGMOID RASIO 70:30

	Negatif	Positif
Negatif	262	56
Positif	77	367

3) *Support Vector Machine Kernel RBF*

Untuk rasio pembagian 70:30 dan penggunaan kernel RBF, hasil percobaan ditampilkan dalam Tabel VIII yang menggambarkan *Confusion Matrix*. *Confusion Matrix* ini mencatat jumlah prediksi yang tepat dan yang salah untuk setiap kelas pada model yang diuji dengan rasio 70:30.

TABEL VIII
 CONFUSION MATRIX SUPPORT VECTOR MACHINE RBF RASIO 70:30

	Negatif	Positif
Negatif	176	31
Positif	44	257

TABEL IX
 HASIL PERBANDINGAN KERNEL YANG DIUJI

Kernel	Support Vector Machine			
	Akurasi	Presisi	Recall	F1-Score
Sigmoid	82%	77%	81%	79%
RBF	85%	81%	85%	83%
Sigmoid	81%	75%	80%	77%
RBF	85%	80%	85%	82%

Hasil eksperimen menunjukkan bahwa kinerja *Support Vector Machine* (SVM) bergantung pada jenis kernel. Kernel RBF memberikan akurasi tertinggi untuk data kompleks dan non-linear. Kernel polinomial cukup baik untuk data dengan struktur polinomial, sedangkan kernel sigmoid lebih cocok untuk data yang dapat dipisahkan secara linier tetapi kurang efektif untuk data kompleks. Pemilihan kernel yang tepat sangat penting berdasarkan sifat data dan kompleksitas masalah.

I. Perbandingan Metode *Naïve Bayes* dan *Support Vector Machine*

TABEL X
 HASIL PERBANDINGAN ALGORITMA

Rasio	Naive Bayes			Support Vector Machine		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
30 : 70	82%	77%	82%	85%	81%	85%

Dari Tabel X, dapat dilihat bahwa pada rasio pembagian data 30:70, baik *Naive Bayes* maupun *Support Vector Machine* mempunyai akurasi yang tinggi, dengan *Support Vector Machine* sedikit lebih unggul dalam hal akurasi (85% vs. 82%). Namun, *Naive Bayes* menunjukkan presisi yang lebih baik (77% vs. 81%), sementara recall-nya sama dengan *Support Vector Machine* (82%).

Sementara pada rasio pembagian data 20:80, keduanya memiliki akurasi yang sama (85%), namun, *Support Vector Machine* memiliki presisi yang sedikit lebih tinggi (80% vs. 76% untuk *Naive Bayes*), sementara recall-nya sama-sama 85%.

IV. KESIMPULAN

1. Proses analisis sentimen ulasan Aplikasi Teman Bus melibatkan penerapan Algoritma *Naïve Bayes* dan *Support Vector Machine*. Melalui penggunaan kedua algoritma tersebut, dilakukan pemrosesan teks dan klasifikasi sentimen untuk mengevaluasi tanggapan pengguna terhadap aplikasi tersebut.
2. Hasil perbandingan antara pemakaian algoritma *Naïve Bayes* dan *Support Vector Machine* menunjukkan variasi pada performa klasifikasi sentimen. *Support Vector Machine*, khususnya dengan kernel RBF mendapatkan akurasi 85%, cenderung memberikan hasil yang lebih baik dalam menangani pola sentimen yang kompleks dan tidak linier dibandingkan dengan *Naïve Bayes* yakni 82%, terutama pada dataset dengan rasio pembagian data 30 : 70.
3. Penentuan metode terbaik dalam menganalisis sentimen aplikasi Teman Bus memperhitungkan variasi performa antara *Naïve Bayes* dan *Support Vector Machine*. *Support Vector Machine* dengan kernel RBF cenderung menjadi pilihan yang lebih unggul 3 % yakni dengan akurasi 85% dalam menangani pola sentimen yang kompleks, namun pemilihan metode harus disesuaikan dengan karakteristik khusus dari dataset yang digunakan, seperti rasio pembagian data dan struktur pola sentimen yang ada.

V. SARAN

Berlandaskan kesimpulan sebelumnya, berikut merupakan saran agar dapat mengembangkan penelitian lebih lanjut:

1. Perluasan Kamus Normalisasi : Disarankan untuk memperluas kamus normalisasi dengan kata-kata yang umum digunakan dalam media sosial untuk meningkatkan akurasi normalisasi teks.
2. Eksplorasi Metode Klasifikasi Alternatif : Disarankan untuk mengeksplorasi metode klasifikasi lainnya selain *Naive Bayes* dan *Support Vector Machine* dalam menerima wawasan tambahan tentang analisis sentimen.
3. Validasi Hasil dengan Pengujian Lebih Lanjut : Penting untuk melakukan validasi hasil dengan pengujian lebih lanjut menggunakan dataset yang lebih besar dan beragam untuk mengonfirmasi konsistensi dan generalisabilitas hasil.
4. Kolaborasi dengan Pihak Terkai t: Saran untuk melakukan kolaborasi dengan Kementerian Perhubungan RI atau Pemerintah Kota Surabaya untuk memastikan responsifitas aplikasi "Teman Bus" terhadap kebutuhan pengguna.
5. Evaluasi Terhadap Pengguna Aktif : Disarankan untuk melakukan evaluasi langsung terhadap pengguna aktif aplikasi "Teman Bus" untuk memahami persepsi dan kebutuhan mereka secara lebih baik.

REFERENSI

- [1] Ashari, S. A., Saputra, M. W. A., Larosa, E., & Rijal, B. S. (2023). Analisis sentimen pada aplikasi Translate Google menggunakan metode SVM (studi kasus: Komentar pada Playstore). *Jurnal Teknik*, 21(2). Universitas Negeri Gorontalo.
- [2] Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika*. *Jurnal Edik Informatika*, 2(2), 213–219.
- [3] Nurdin, N., Suhendri, M., Afrilia, Y., & Rizal, R. P. (2021). Klasifikasi karya ilmiah (tugas akhir) mahasiswa menggunakan metode Naive Bayes Classifier (NBC). *SISTEMASI*, 10(2), 268.
- [4] Analisis Sentimen Aplikasi Investasi Reksa Dana (Bibit) Menggunakan Metode Support Vector Machine. (N.D.).
- [5] Tineges, R. (2021). Tahapan Text Preprocessing dalam Teknik Pengolahan Data. <https://dqlab.id/Tahapan-Text-Preprocessing-Dalam-Teknik-Pengolahan-Data>.
- [6] Setiadi, K. (2023). Analisis Sentimen Pelanggan Terhadap Layanan Shopeefood Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine (SUPPORT VECTOR MACHINE). *KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika*, 12(1).
- [7] Trivusi. (2022). Data Splitting: Pengertian, Metode, dan Kegunaannya. <https://howdoi.blog/Contoh-Pengambilan-Acak-Berdasarkan-Lapisan-Stratified-Random-Sampling>
- [8] Khotimah, N. (2019). Analisis Sentimen Terhadap Review E-Commerce Dengan Metode Stochastic Gradient Descent. Sarjana / Sarjana Terapan (S1/D4) thesis, Muhammadiyah University, Semarang. J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Renocongestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. hal. 99-02, 1999.
- [9] Tsani, M., Rupaka, A., Asmoro, L., & Pradana, B. (2020). Analisis Sentimen Review Transportasi Menggunakan Algoritma Support Vector Machine Berbasis Chi Square. *Smart Comp :Jurnalnya Orang Pintar Komputer*, 9(1), 35–39. <https://doi.org/10.30591/smartcomp.v9i1.1817>
- [10] Yunus, M. (2020, April 30). TF-IDF (Term Frequency-Inverse Document Frequency): Representasi vector data text. *Medium*. <https://yunusmuhammad007.medium.com/tf-idf-term-frequency-inverse-document-frequency-representasi-vector-data-text-2a4eff56cda>
- [11] Rahayu, A. S., Fauzi, A., & Rahmat, R. (2022). Komparasi Algoritma Naive Bayes Dan Support Vector Machine (SUPPORT VECTOR MACHINE) Pada Analisis Sentimen Spotify. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 349.
- [12] Nabila, A. (2022). Analisis sentimen ulasan produk toner pada beauty brand "The Body Shop" menggunakan metode Naive Bayes Classifier dan Support Vector Machine: Studi kasus di Female Daily (Laporan Tugas Akhir, Universitas Islam Indonesia). Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia, Yogyakarta.

