

PERBANDINGAN KINERJA KERNEL LINEAR DAN RBF SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA KAI ACCESS PADA GOOGLE PLAY STORE

Eka Rizqi Mar'atus Sholihah, I Gede Susrama Mas Diyasa, Eva Yulia Puspaningrum

Program Studi Informatika S1, Fakultas Ilmu Komputer,
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jalan Raya Rungkut Madya Gunung Anyar Surabaya, Indonesia
19081010022@student.upnjatim.ac.id

ABSTRAK

Aplikasi KAI Access, yang merupakan *platform* pembelian tiket online milik PT Kereta Api Indonesia (Persero), dapat diunduh melalui berbagai platform seperti Google Play Store. Ulasan pengguna aplikasi KAI Access di Google Play Store akan membentuk persepsi pengguna terkait aplikasi. Untuk memahami ulasan atau mendeteksi sentimen yang diberikan pengguna terhadap aplikasi, penting dilakukan suatu teknik pengolahan data yang disebut sebagai teknik analisis sentimen. Analisis sentimen dengan menggunakan algoritma support vector machine (SVM), diterapkan pada ulasan KAI Access di Google Play Store dalam penelitian ini. Tujuan dari penelitian ini adalah mengukur performa algoritma SVM dengan kernel linear dan kernel RBF dalam mengklasifikasikan sentimen menjadi dua kelas, yaitu positif dan negatif. Sebanyak 10.000 data ulasan aplikasi KAI Access di Google Play Store digunakan dalam penelitian ini. Hasil pengujian menunjukkan bahwa SVM Kernel Linear mampu menghasilkan model analisis sentimen dengan akurasi tertinggi sebesar 83,1%, sementara SVM Kernel RBF mampu mencapai akurasi tertinggi sebesar 86,1%.

Kata kunci : tuliskan maksimum 6 kata kunci di sini – *Support Vector Machine, Analisis Sentimen, KAI Access, Kernel Linear, Kernel RBF*

1. PENDAHULUAN

Kemajuan teknologi informasi telah mendorong perkembangan aplikasi *mobile* di berbagai sektor, termasuk transportasi. Aplikasi *mobile ticketing* seperti KAI Access telah menjadi alternatif yang lebih efisien dan efektif untuk pembelian tiket kereta api [1]. KAI Access adalah aplikasi *mobile* resmi dari PT Kereta Api Indonesia (Persero) Aplikasi ini pertama kali dirilis pada tahun 2014 dan terus berkembang dengan menambahkan fitur-fitur baru, seperti pembatalan tiket, perubahan jadwal, dan pencetakan boarding pass [2]. Aplikasi ini tersedia di platform penyedia aplikasi, seperti Google Play Store. Pengguna aplikasi dapat memberikan ulasan di Google Play Store. Ulasan-ulasan yang diberikan pengguna akan membentuk persepsi pengguna terhadap aplikasi [3]. Hal tersebut akan mempengaruhi keputusan khalayak dalam mengunduh aplikasi ini. Oleh karena itu, penting memahami ulasan-ulasan pengguna untuk mengetahui sentimen yang terkandung dalam ulasan tersebut apakah positif atau negatif. Untuk memahami ulasan atau mendeteksi sentimen yang diberikan pengguna terhadap aplikasi, dapat dilakukan dengan teknik analisis sentimen. Analisis sentimen adalah salah satu teknik *text mining* yang digunakan untuk mengklasifikasikan kalimat atau opini menjadi sentimen positif dan negatif [4].

Penelitian sebelumnya oleh Jessica dan Yessica pada tahun 2021 tentang analisis sentimen berbasis aspek membandingkan Naive Bayes, SVM, dan K-NN untuk menganalisis komentar YouTube tentang Samsung Galaxy Z Flip 3. Dari hasil penelitian, SVM

menunjukkan performa terbaik dengan rata-rata akurasi 96,43%. Akurasi tertinggi diperoleh untuk aspek desain (94,40%), diikuti oleh harga (97,44%), spesifikasi (96,22%), dan citra merek (97,63%) [5]. Penelitian oleh Nuraeni, dkk, tentang analisis sentimen ulasan aplikasi zoom menggunakan algoritma Naive Bayes dan SVM diperoleh hasil penelitian yang menunjukkan bahwa SVM lebih unggul dari Naive Bayes, dengan akurasi 81,22% dan AUC 0,886, dibandingkan dengan akurasi 74,37% dan AUC 0,659 untuk Naive Bayes [6]. Lalu pada penelitian lainnya yang dilakukan Faizal dan rekan rekannya tentang analisis sentimen review aplikasi Ruang Guru menggunakan 3 kernel SVM, yaitu linear, RBF, dan polynomial. Didapatkan hasil penelitian yang menunjukkan bahwa kernel linear memiliki akurasi tertinggi, yaitu 0.897, diikuti oleh kernel RBF 0.865 dan polynomial 0.845 [7]. Dari penelitian-penelitian sebelumnya menunjukkan bahwa SVM memberikan akurasi yang baik dalam melakukan analisis sentimen.

Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, maka pada penelitian ini akan dilakukan perbandingan performa algoritma SVM kernel linear dan kernel RBF dalam melakukan analisis sentimen ulasan pengguna aplikasi KAI Access di Google Play Store.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah sebuah penelitian untuk menganalisis pendapat, perasaan, dan emosi yang disampaikan dalam teks. Metode ini dikenal juga

sebagai penambangan opini atau *opinion mining*, yang bertujuan untuk mengekstrak dan merangkum data yang terdapat dalam teks, lalu menganalisisnya menggunakan teknik *machine learning*. Analisis sentimen juga dapat digunakan untuk menentukan arah sikap terhadap suatu sentimen, yaitu sentimen positif atau negatif [6].

2.2. Scraping

Scraping adalah suatu teknik yang digunakan untuk mengambil data atau mengekstrak informasi, dan menganalisis data dari. Data dapat berupa teks, gambar, ataupun video. Data yang di ekstrak biasanya akan diubah ke format yang baru. Scraping data dapat dilakukan dengan menjalankan barisan kode atau juga dapat dilakukan dengan bantuan tools scraper data internet [8].

2.3. Preprocessing

Preprocessing data merupakan suatu proses dalam tahapan *text mining* yang bertujuan untuk membersihkan dan menyiapkan data agar siap untuk diolah. Tahapan pada *preprocessing* data antara lain, yaitu *case folding*, *cleaning*, *tokenizing*, *normalization*, *stop word removal*, dan *stemming*. Dengan melakukan *preprocessing data*, data mentah akan menjadi lebih terstruktur, bersih, dan relevan sehingga dapat membantu meningkatkan kualitas analisis dan pemodelan yang dilakukan pada data tersebut [9].

2.4. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF adalah metode untuk menghitung bobot kata (term) dalam sebuah dokumen. Bobot kata tersebut ditentukan oleh seberapa sering kata tersebut muncul dalam dokumen dan seberapa jarang kata tersebut muncul dalam kumpulan dokumen. Hasil dari perhitungan TF-IDF adalah bobot kata yang digunakan dalam analisis data selanjutnya [10]. Adapun rumus pembobotan kata menggunakan TF-IDF adalah sebagai berikut [11].

$$W = TF \times IDF = TF \times \log N/df \quad (1)$$

Keterangan:

W = Bobot TF-IDF
 TF = Jumlah frekuensi kata
 IDF = Jumlah inverse frekuensi dokumen tiap kata
 N = Jumlah keseluruhan dokumen yang ada
 df = Jumlah kemunculan kata pada semua dokumen

2.5. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode machine learning yang memisahkan data dengan dua atau lebih kelas. SVM membangun model dengan menemukan bidang pemisah yang memaksimalkan jarak antara kelas-kelas. SVM dapat menangani data non linear dengan memetakan data ke ruang dimensi

yang lebih tinggi menggunakan kernel [12]. Berikut adalah persamaan dari setiap kernel [9].

Kernel linear

$$K(x_i, x) = (x_i x) \quad (2)$$

Kernel RBF

$$K(x_i, x) = \exp \frac{-||x_i - x||^2}{2\sigma^2} \quad (3)$$

Kernel Polynomial

$$K(x_i, x) = (x_i x)^d \quad (4)$$

Kernel Sigmoid

$$K(x_i, x) = \tanh(\sigma(x_i x) + c) \quad (5)$$

2.6. Confusion matrix

Confusion matrix adalah tabel yang menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang salah diklasifikasikan. Matriks ini dapat digunakan untuk menilai kinerja model klasifikasi yaitu dengan menghitung nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Matriks ini dibagi menjadi empat sel, yang masing-masing mewakili kombinasi dua kemungkinan: kelas yang diprediksi benar dan kelas yang diprediksi salah [13].

Tabel 1. Confusion matrix

Correct classification	Classified as	
	Predicted “+”	Predicted “-”
Actual “+”	True positive	False negatives
Actual “-”	False positive	True negatives

Keterangan:

- True Positive (TP) yaitu jumlah data uji positif yang diklasifikasikan sebagai positif.
- False Positive (FP) yaitu jumlah data uji negatif yang diklasifikasikan sebagai positif
- False negatives (FN) yaitu jumlah data uji positif yang diklasifikasikan sebagai negatif.
- True negatives (TN) yaitu jumlah data uji negatif yang diklasifikasikan sebagai negatif.

Adapun rumus metrik evaluasi yang digunakan dalam perhitungan yang melibatkan confusion matrix antara lain adalah sebagai berikut:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$precision = \frac{TP}{TP+FP} \quad (7)$$

$$recall = \frac{TP}{TP+FN} \quad (8)$$

$$f1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (9)$$

3. METODE PENELITIAN

Metode penelitian yang diterapkan pada penelitian ini terdiri dari beberapa tahapan antara lain:

3.1. Pengumpulan Data

Dalam tahap ini dilakukan pengumpulan data ulasan pengguna serta seleksi atribut data yang akan digunakan pada penelitian. Pada tahap ini juga akan

dilakukan pelabelan data menggunakan skor atau *rating* aplikasi yang diberikan pengguna [14].

3.2. Preprocessing Data

Dalam tahap preprocessing ini dilakukan proses pembersihan data untuk menyiapkan data agar siap dianalisis. Proses ini meliputi beberapa tahap, yaitu [9]:

- *Case folding*: mengubah semua huruf menjadi huruf kecil atau besar.
- *Cleaning*: menghilangkan kesalahan ketik, simbol, dan kata yang tidak relevan.
- *Tokenizing*: memecah data teks menjadi kata-kata atau frasa.
- *Normalization*: mengubah kata-kata ke bentuk baku.
- *Stop word removal*: menghilangkan kata-kata yang umum dan tidak relevan.
- *Stemming*: mengubah kata-kata ke akar kata.

3.3. Transformation

Pada tahap *transformation*, data teks diubah menjadi data numerik menggunakan metode TF-IDF. Hal ini dilakukan agar data mampu dianalisis oleh algoritma klasifikasi.

3.4. Implementasi Algoritma

Pada tahap ini, dilakukan proses klasifikasi oleh algoritma terhadap data. Algoritma yang diaplikasikan pada proses ini adalah *support vector machine* (SVM).

3.5. Evaluasi Model

Setelah model analisis sentimen berhasil dibuat, langkah selanjutnya adalah evaluasi model menggunakan confusion matrix. Hasil dari proses ini adalah berupa *accuracy*, *presicion*, *recall*, dan *f1-score*.

4. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan menggunakan bahasa pemrograman python dengan mengimplementasikan algoritma klasifikasi *Support Vector Machine* (SVM) kernel linier dan kernel RBF untuk membuat model analisis sentimen ulasan pengguna aplikasi KAI Access pada Google Play Store.

4.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah ulasan pengguna aplikasi KAI Access di Google Play Store yang diambil pada tanggal 20 Oktober 2023. Ulasan tersebut dikumpulkan menggunakan teknik *scraping* dengan library *google-play-scraper* python. Jumlah ulasan yang digunakan adalah 10.000, dan ulasan tersebut diurutkan berdasarkan urutan terbaru. Proses labeling data dilakukan untuk memberi label sentimen pada data ulasan aplikasi KAI Access. Label sentimen dibagi menjadi dua, yaitu positif dan negatif. Pemberian label dilakukan menggunakan *rating*, *rating* 1 dan 2 dikategorikan sebagai sentimen negatif, sedangkan *rating* 3, 4, dan 5 dikategorikan sebagai

sentimen positif. Notasi labeling untuk sentimen positif adalah 1, sedangkan untuk sentimen negatif adalah 0.

4.2. Preprocessing Data

Preprocessing dilakukan agar data menjadi lebih terstruktur dan konsisten. Hal ini dikarenakan data yang diperoleh belum siap diproses langsung karena masih terlalu banyak *noise* serta fitur kata yang kurang bermakna. Oleh karena itu, pada proses *preprocessing data* ini, akan dilakukan beberapa tahapan untuk mengatasi hal tersebut. Contoh hasil dari proses *preprocessing* dapat dilihat pada tabel berikut:

Tabel 1. Contoh Hasil Preprocessing

Proses	Hasil
Content Ulasan	Saat perubahan jadwal maupun perubahan no kursi udh bagus, ada biaya 25%. tp bagus proses cepat. Tp knp saat pembatalan tiket proses pengembalian dananya sangat lama..?? itu saja..
Case Folding	saat perubahan jadwal maupun perubahan no kursi udh bagus, ada biaya 25%. tp bagus proses cepat. tp knp saat pembatalan tiket proses pengembalian dananya sangat lama..?? itu saja..
Cleaning	saat perubahan jadwal maupun perubahan no kursi udh bagus ada biaya tp bagus proses cepat tp knp saat pembatalan tiket proses pengembalian dananya sangat lama itu saja
Tokenizing	['saat', 'perubahan', 'jadwal', 'maupun', 'perubahan', 'no', 'kursi', 'udh', 'bagus', 'ada', 'biaya', 'tp', 'bagus', 'proses', 'cepat', 'tp', 'knp', 'saat', 'pembatalan', 'tiket', 'proses', 'pengembalian', 'dananya', 'sangat', 'lama', 'itu', 'saja']
Normalization	['saat', 'perubahan', 'jadwal', 'maupun', 'perubahan', 'no', 'kursi', 'sudah', 'bagus', 'ada', 'biaya', 'tapi', 'bagus', 'proses', 'cepat', 'tapi', 'kenapa', 'saat', 'pembatalan', 'tiket', 'proses', 'pengembalian', 'dananya', 'sangat', 'lama', 'itu', 'saja']
Stop Word Removal	['perubahan', 'jadwal', 'perubahan', 'no', 'kursi', 'bagus', 'biaya', 'bagus', 'proses', 'cepat', 'pembatalan', 'tiket', 'proses', 'pengembalian', 'dananya']
Stemming	['ubah', 'jadwal', 'ubah', 'no', 'kursi', 'bagus', 'biaya', 'bagus', 'proses', 'cepat', 'batal', 'tiket', 'proses', 'kembali', 'dana']

4.3. Transformation

Gambar 1. Matriks hasil proses TF-IDF

Data yang telah diproses sebelumnya hingga tahap *stemming*, kemudian diubah kedalam bentuk numerik menggunakan TF-IDF.

Dari gambar diatas dapat diketahui bahwa seluruh data yang telah di proses pada proses TF-IDF menghasilkan matriks 10000 x 8453. Hal ini menandakan bahwa fitur kata yang dihasilkan sebanyak 8453 dari 10000 data.

4.4. Implementasi Algoritma

Pada tahap ini dilakukan klasifikasi sentimen dengan mengimplementasikan algoritma SVM sebagai algoritma klasifikasinya. Klasifikasi sentimen dilakukan dengan membagi data menjadi dua skenario, yaitu 90% data latih dengan 10% data uji dan 80% data latih dengan 20% data uji. Selain itu juga dilakukan beberapa skenario pengujian menggunakan nilai hyperparameter untuk mencari nilai akurasi yang terbaik. Skenario pengujian selengkapnya sebagai berikut:

Tabel 3. Skenario Pengujian

No	Model klasifikasi	Data train	Data test	Hyperparameter
1	SVM Kernel Linear	80%	20%	C= 1
2		80%	20%	C= 10
3		90%	10%	C= 1
4		90%	10%	C= 10
5	SVM Kernel RBF	90%	10%	C=1, $\gamma= 0,1$
6		90%	10%	C=10, $\gamma= 0,1$
7		90%	10%	C=1, $\gamma= 1$
8		90%	10%	C=10, $\gamma= 1$
9		80%	20%	C=10, $\gamma= 0,1$
10		80%	20%	C=1, $\gamma= 0,1$
11		80%	20%	C=1, $\gamma= 1$
12		80%	20%	C=10, $\gamma= 1$

4.5. Evaluasi Model

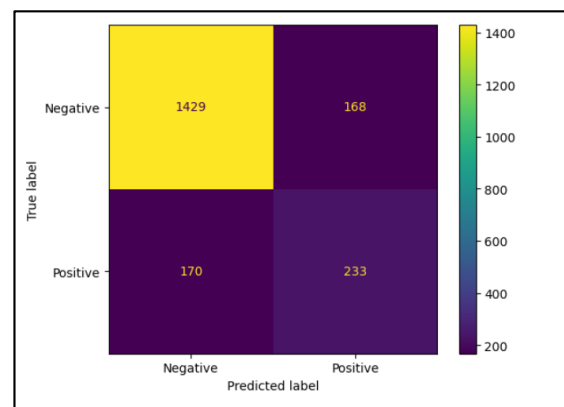
Pada tahapan ini, model klasifikasi yang telah dibuat dievaluasi. Metode evaluasi model yang digunakan yaitu confusion matrix untuk mengetahui nilai *True Positive*, *True Negative*, *False positive* dan *False Negative*. Dari nilai nilai tersebut nantinya akan dapat digunakan untuk mengetahui nilai *Accuracy*, *Precision*, *Recall*, dan *F-1 score*. Hasil evaluasi model tersebut tersaji pada tabel berikut:

Tabel 4. Hasil Pengujian

No.	Sentimen	p	r	f-s	Accuracy
1	1	0,59	0,5	0,54	82,95%
	-1	0,88	0,91	0,9	
2	1	0,58	0,58	0,58	83,10%
	-1	0,89	0,89	0,89	
3	1	0,59	0,5	0,54	83%
	-1	0,88	0,91	0,9	
4	1	0,56	0,57	0,57	82,50%
	-1	0,89	0,89	0,89	
5	1	0,84	0,38	0,52	86,10%
	-1	0,86	0,98	0,92	
6	1	0,99	0,31	0,47	85,90%
	-1	0,85	1	0,92	
7	1	0,87	0,31	0,46	85%
	-1	0,85	0,99	0,91	

No.	Sentimen	p	r	f-s	Accuracy
8	1	0,89	0,32	0,47	85,60%
	-1	0,85	0,99	0,92	
9	1	0,96	0,24	0,39	85%
	-1	0,84	1	0,91	
10	1	0,82	0,36	0,5	85%
	-1	0,86	0,98	0,92	
11	1	0,85	0,3	0,44	85%
	-1	0,85	0,99	0,91	
12	1	0,91	0,31	0,47	85,55%
	-1	0,85	0,99	0,92	

Dari tabel yang tersaji diatas, pengujian SVM kernel linear berada pada pengujian ke-1 sampai 4 dan SVM kernel RBF berada pada pengujian ke-5 hingga 12. Dari 4 pengujian yang diuji cobakan pada algoritma SVM kernel linear, didapatkan hasil akurasi terbaik sebesar 83,1%. Hasil tersebut didapatkan dari pengujian ke-2 yang dilakukan dengan skenario pembagian data 80% data training dan 20% data testing serta penggunaan hyperparameter C dengan nilai 10. Visualiasi *confusion matrix* yang dihasilkan adalah sebagai berikut.



Gambar 2. *Confusion matrix* pengujian ke-2

Dari pengujian skenario 2 tersebut, model berhasil memprediksi 1662 data secara benar dari 2000 data secara keseluruhan. Terdapat 1429 data label negatif yang terprediksi benar sebagai negatif dan 233 data label positif yang terprediksi benar sebagai positif. Sedangkan jumlah data yang salah diprediksi adalah 338 data. *Classification report* untuk pengujian tersebut adalah sebagai berikut.

	precision	recall	f1-score	support
-1	0.89	0.89	0.89	1597
1	0.58	0.58	0.58	403
accuracy			0.83	2000
macro avg	0.74	0.74	0.74	2000
weighted avg	0.83	0.83	0.83	2000

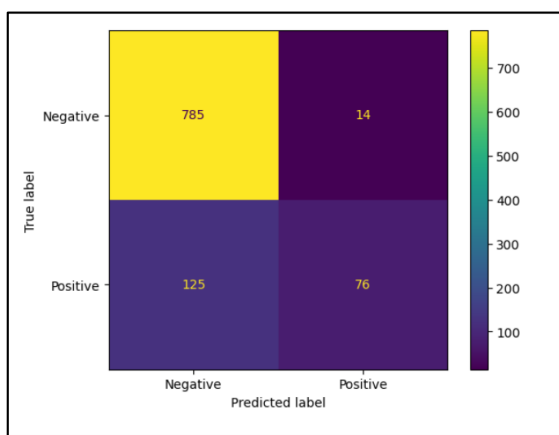
Gambar 3. *Classification report* pengujian ke-2

Dari *classification report* yang tersaji diatas dapat dilihat bahwa nilai presisi untuk sentimen negatif mencapai 89%, menunjukkan bahwa 89% dari prediksi yang dinyatakan sebagai sentimen negatif

adalah akurat. Sebaliknya, presisi untuk sentimen positif hanya mencapai 58%, yang berarti bahwa dari prediksi sentimen positif, hanya 58% yang sesuai. Secara bersamaan, nilai recall untuk sentimen negatif dan positif juga adalah 89% dan 58% masing-masing. Ini mengindikasikan bahwa model berhasil memprediksi 89% dari seluruh data uji yang benar-benar sentimen negatif, sementara hanya 58% dari data uji yang benar-benar sentimen positif yang berhasil diprediksi oleh model. Terakhir, nilai f1-score, yang merupakan nilai harmonic mean dari presisi dan recall, adalah 89% untuk sentimen negatif dan 58% untuk sentimen positif. Ini memberikan gambaran holistik tentang keseimbangan antara presisi dan recall untuk kedua kelas sentimen. Perhitungan manual akurasi dari pengujian tersebut adalah sebagai berikut.

$$\begin{aligned} \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\ &= \frac{233+1429+168+170}{233+1429+168+170} \times 100\% \\ &= \frac{1662}{2000} \times 100\% \\ &= 83,1\% \end{aligned}$$

Lalu pada pengujian yang dilakukan pada algoritma SVM kernel RBF, dari 8 pengujian yang diuji cobakan, didapatkan hasil akurasi terbaik 86,1%. Hasil ini diperoleh dari skenario pengujian ke-5 yang menggunakan pembagian data training 90% dan data testing 10% serta menggunakan hyperparameter C dengan nilai 1 dan hyperparameter gamma dengan nilai 0,1. Visualisasi *confusion matrix* yang dihasilkan adalah sebagai berikut.



Gambar 4. *Confusion matrix* pengujian ke-5

Dari pengujian skenario 5 tersebut, model berhasil memprediksi 861 data secara benar dari 1000 data secara keseluruhan. Terdapat 785 data label negatif yang terprediksi benar sebagai negatif dan 76 data label positif yang terprediksi benar sebagai positif. Sedangkan jumlah data yang salah diprediksi adalah 139 data. *Classification report* untuk pengujian tersebut adalah sebagai berikut.

	precision	recall	f1-score	support
-1	0.86	0.98	0.92	799
1	0.84	0.38	0.52	201
accuracy			0.86	1000
macro avg	0.85	0.68	0.72	1000
weighted avg	0.86	0.86	0.84	1000

Gambar 5. *Classification report* pengujian ke-5

Dari *classification report* diatas menunjukkan kinerja model dalam memprediksi sentimen dengan nilai presisi sebesar 86% untuk sentimen negatif dan 84% untuk sentimen positif. Dengan kata lain, dari semua prediksi yang dikategorikan sebagai sentimen negatif, 86% di antaranya benar-benar mencerminkan sentimen negatif, sementara 84% dari keseluruhan prediksi sentimen positif sesuai dengan kenyataan. Untuk nilai recall, model berhasil memprediksi 93% dari seluruh data uji yang merupakan sentimen negatif, menandakan kemampuannya dalam mengenali sentimen negatif dengan akurat. Namun, pada sentimen positif, hanya 38% dari data uji yang benar-benar sentimen positif yang dapat diprediksi oleh model. Selanjutnya, nilai f1-score, yang merupakan nilai harmonic mean dari presisi dan recall, menunjukkan keseimbangan antara kinerja presisi dan recall untuk kedua kelas sentimen. Untuk sentimen negatif, f1-score mencapai 92%, sementara untuk sentimen positif mencapai 52%. Perhitungan manual akurasi dari pengujian tersebut adalah sebagai berikut.

$$\begin{aligned} \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\ &= \frac{76+785}{76+785+14+125} \times 100\% \\ &= \frac{861}{1000} \times 100\% \\ &= 86,1\% \end{aligned}$$

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, diperoleh hasil bahwa tingkat akurasi penggunaan algoritma SVM Kernel linear dalam melakukan analisis sentimen pada ulasan aplikasi KAI Access di Google Play Store mencapai akurasi terbaik sebesar 83,1% yang diperoleh dari uji coba pembagian data 80/20 dengan penggunaan hyperparameter C=10. Selanjutnya, tingkat akurasi penggunaan algoritma SVM Kernel RBF dalam melakukan analisis sentimen pada ulasan aplikasi KAI Access di Google Play Store mencapai tingkat akurasi terbaik sebesar 86,1% yang diperoleh dari uji coba pembagian data 90/10 dengan penggunaan hyperparameter C=1 dan gamma=0,1. Dari temuan ini, dapat disimpulkan bahwa performa SVM Kernel RBF dalam melakukan analisis sentimen ulasan pengguna KAI Access memberikan model yang memiliki akurasi lebih baik dari pada SVM Kernel linear. Saran penelitian yang dapat diterapkan pada penelitian selanjutnya yaitu dapat menggunakan pelabelan data menggunakan metode lain serta dapat menambahkan seleksi fitur seperti *Particle Swarm Optimization* (PSO).

DAFTAR PUSTAKA

- [1] K. R. Hadi, H. M. Az-Zahra dan L. Fanani, "Analisis Dan Perbaikan Usability Aplikasi Mobile KAI Access Dengan Metode Usability Testing Dan Use Questionnaire," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. II, no. 9, pp. 2742-2750, September 2018.
- [2] S. D. Pratama dan E. Syaodih, "Analisis Perilaku Konsumen dalam Memanfaatkan Aplikasi KAI Access," *Service Management Triangle: Manajemen Jasa*, pp. 20-27, 2021.
- [3] M. Izunahdi, G. Aburrahman dan A. E. Wardoyo, "Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google Play Store Menggunakan Metode Multinomial Naive Bayes," *Jurnal Smart Teknologi*, vol. 2, no. 4, pp. 192-198, 2023.
- [4] R. Wahyudi dan G. Kusumawardhana, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *Jurnal Informatika*, pp. 200-207, September 2021.
- [5] J. W. Iskandar dan Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, pp. 1120-1126, 2021.
- [6] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata dan Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal of Computer Engineering System and Science)*, vol. 5, no. 2, pp. 293-295, Juli 2020.
- [7] F. F. Irfani, M. Triyanto, A. D. Hartanto dan Kusnawi, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *Jurnal Bisnis, Manajemen dan Informatika*, vol. 16, no. 3, pp. 258-266, Februari 2020.
- [8] M. A. Adli dan L. Firgia, "Rancang Bangun Web Scraping Pada Media Online Berita Nasional," *Jurnal ENTER*, pp. 118-128, 2018.
- [9] M. D. Hendriyanto, A. A. Ridha dan U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *Journal Of Information Technology and Computer Science (IntecomS)*, pp. 1-7, 2022.
- [10] B. Herwijayanti, D. E. Ratnawati dan L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, pp. 306-312, 2017.
- [11] J. A. Septian, T. M. Fahrudin dan A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*, no. 1, pp. 43-49, 2019.
- [12] S. Sutharan, "Support Vector Machine," dalam *Machine Learning Models and Algorithms for Big Data Classification. Integrated Series in Information Systems*, Boston, Springer, 2016, pp. 207-235.
- [13] M. F. Rahman, I. M. Darmawidjaja dan D. Alamsah, "Klasifikasi Untuk Diganosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RNN)," *Jurnal Informatika*, vol. 11, pp. 36-45, Januari 2017.
- [14] V. O. Tama, Y. Sibaroni dan Adiwijaya, "Labeling Analysis in the Classification of Product Review Sentiments by using Multinomial Naive Bayes Algorithm," *Journal of Physics: Conference Series*, vol. 1192, no. 2, pp. 1-11, 2019.