Analisis Sentimen Ulasan Aplikasi Vidio pada *Play Store* Menggunakan Algoritma *Support Vector Machine*

Bima Santika*[1], Agung Kharisma Hidayah[2]
[1,2]Universitas Muhammadiyah Bengkulu; Jl. Bali, Kampung Bali, Teluk Segara, Kota Bengkulu, 38119
[3]Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu, Bengkulu
e-mail: *[1]**santikabima83@gmail.com**, [2]kharisma@umb.ac.id,

Abstrak

Vidio: Sports, Movies, Series adalah aplikasi streaming populer di Indonesia dengan rating 3,5 bintang di Play Store pada 2024. Meski banyak digunakan, sejumlah ulasan pengguna mencerminkan ketidakpuasan terhadap beberapa fitur dan performanya. Penelitian ini difokuskan menganalisis pendapat pengguna tentang aplikasi Vidio di Play Store dan mengevaluasi akurasi algoritma SVM dalam mengelompokkan sentimen menggunakan metode KDD. KDD diterapkan untuk menentukan pendekatan analisis dari data yang tersedia. Pada tahap transformasi data, data pelatihan dan pengujian dibagi dalam tiga rasio: 90:10, 80:20, dan 70:30. Kata-kata diberi bobot menggunakan teknik TF-IDF. Penelitian ini menguji performa SVM menggunakan empat kernel: Linear, RBF, Poly, dan Sigmoid. Hasil penelitian menunjukkan bahwa rasio 70:30 dengan kernel Sigmoid menghasilkan performa terbaik, yakni akurasi 83,33%, precision 86,13%, recall 83,33%, dan F1-score 76,25%. Model ini berhasil mencapai komposisi yang optimal antara precision dan recall, meskipun ada sedikit kompromi yang terlihat pada F1- score.

Kata kunci— *Analisis Sentimen, SVM, Vidio, Play Store*

Abstract

Vidio: Sports, Movies, Series is a popular streaming application in Indonesia with a 3.5 star rating on the Play Store in 2024. Although widely used, a number of user reviews reflect dissatisfaction with some of its features and performance. This research focuses on analyzing user opinions about the Vidio application on the Play Store and evaluating the accuracy of the SVM algorithm in grouping sentiments using the KDD method. KDD is applied to determine the analysis approach from the available data. In the data transformation stage, training and testing data are divided into three ratios: 90:10, 80:20, and 70:30. Words are weighted using the TF-IDF technique. This research tests SVM performance using four kernels: Linear, RBF, Poly, and Sigmoid. The research results show that a 70:30 ratio with the Sigmoid kernel produces the best performance, namely 83.33% accuracy, 86.13% precision, 83.33% recall, and 76.25% F1-score. This model succeeded in achieving an optimal composition between precision and recall, although there was a slight compromise seen in the F1- score.

Keywords— *Sentiment Analysis, SVM, Vidio, Play Store*

1. PENDAHULUAN

Kemajuan teknologi telah berkembang dengan sangat cepat, menjadikan teknologi memiliki peran yang krusial dalam memenuhi berbagai kebutuhan manusia [1]. Salah satu wujudnya adalah pertumbuhan pesat aplikasi *mobile* di tengah masyarakat Indonesia, seiring dengan kemajuan teknologi *smartphone*. Menurut data dari Menristekdikti, Jumlah pengguna ponsel pintar di Indonesia saat ini telah mencapai sekitar 167 juta orang, atau sekitar 89% dari total populasi, yang mencerminkan besarnya pengaruh teknologi ini dalam kehidupan sehari-hari[2]. Kemajuan teknologi informasi juga mendorong perubahan pola konsumsi media, dengan banyaknya penonton yang beralih dari televisi ke layanan *streaming* online, menunjukkan dampak teknologi pada cara manusia menikmati hiburan[3].

Google mengelola sebuah platform bernama *Play Store* yang menyediakan berbagai macam konten digital, seperti permainan, aplikasi, buku, film, dan musik, yang dibagi berdasarkan kategori tertentu[4]. Vidio adalah salah satu aplikasi streaming populer di Indonesia, *platform* hiburan untuk streaming film lokal, internasional, dan olahraga langsung. Sebagian besar kontennya memerlukan langganan Vidio *Premier*, dan aplikasi ini memiliki *rating* 3,5 bintang di *Play Store* pada tahun 2024[5]. Meskipun demikian, kualitas layanan aplikasi Vidio sering menjadi perdebatan di kalangan pengguna. Banyak pengguna yang memberikan ulasan negatif terkait dengan kinerja aplikasi, seperti masalah pada kecepatan streaming, kesulitan navigasi, dan ketidakstabilan aplikasi. Dengan demikian, analisis sentimen terhadap ulasan ini menjadi penting untuk memberikan masukan kepada pengembang dalam meningkatkan layanan yang ditawarkan.

Selain itu, *Google Play* memiliki fitur ulasan pengguna yang memungkinkan calon pengguna untuk mengetahui pengalaman dan opini orang lain mengenai aplikasi seperti Vidio. Fitur ini menjadi sarana yang praktis dan berguna bagi pengguna dalam memperoleh informasi terkait kualitas atau layanan suatu aplikasi[6]. Namun, meskipun ulasan tersebut dapat memberikan informasi yang sangat berguna, analisis manual terhadap ribuan ulasan tidaklah efisien dan rawan terhadap bias. Oleh karena itu, dibutuhkan penerapan teknik analisis sentimen otomatis untuk mengkategorikan ulasan pengguna ke dalam sentimen positif atau negatif, sehingga mempermudah pengembang maupun calon pengguna dalam mengevaluasi kualitas aplikasi.

SVM merupakan teknik untuk mengklasifikasikan data yang menggunakan pendekatan pembelajaran terawasi, dan sering kali memberikan hasil yang lebih unggul dibandingkan metode pembelajaran mesin lainnya [7]. Keunggulan ini berasal dari kemampuannya memisahkan data non-*Linear* berdimensi besar secara *Linear* dengan bantuan fungsi *kernel*, sehingga membuat *SVM* sangat efektif dalam berbagai jenis analisis data[8]. Selain itu, *SVM* dapat dengan cepat menggeneralisasi data di luar pelatihan, meminimalkan kesalahan, dan tetap efektif tanpa terpengaruh jumlah atau dimensi variabel[9]. Hal ini menjadikan *SVM* salah satu metode terbaik dalam kategori klasifikasi, menghasilkan performa yang lebih baik dibandingkan banyak pendekatan lainnya[10].

Berdasarkan penelitian terdahulu, algoritma *SVM* dengan *kernel Linear* pada skenario 70:30 menunjukkan kinerja terbaik dengan akurasi 93,7% dan *F1-score* 92,3%. Hasil ini lebih unggul dibandingkan algoritma *Decision Tree*, yang hanya mencapai akurasi 83% dan *F1-score* 79,1%[11]. Penelitian lain juga mengonfirmasi keunggulan *SVM*, di mana akurasi yang dicapai adalah 88%, sedikit lebih tinggi dibandingkan *Naive Bayes* yang memperoleh akurasi 87%[12]. Adapun penelitian terkait lainnya, menunjukkan bahwa penggunaan algoritma *SVM* dengan Vader mencapai akurasi 59% dan *F1-score* 55%, yang lebih tinggi dibandingkan *SVM* dengan *RoBERTa* yang hanya mencapai akurasi 55% dan *F1-score* 52%. Sebaliknya, algoritma *Naive Bayes* menunjukkan performa lebih rendah, dengan akurasi 47% ketika menggunakan Vader dan 43% saat menggunakan *RoBERTa*.[13].

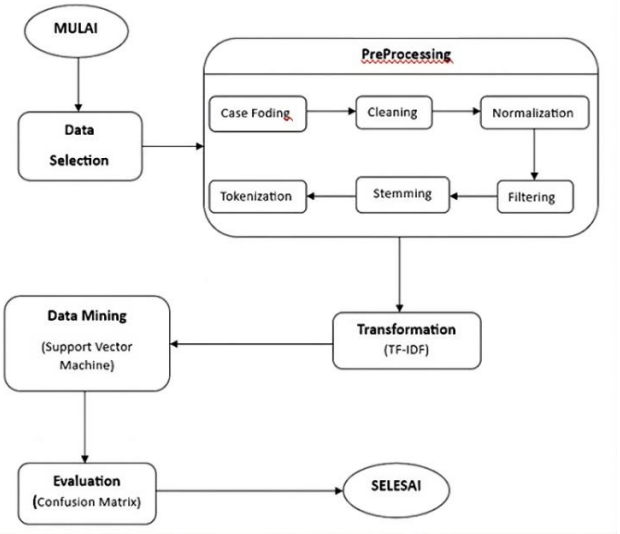
Adapun penelitian serupa lainnya, keunggulan *SVM* juga terlihat dalam klasifikasi ulasan

aplikasi BRImo, di mana *SVM* mencatatkan akurasi 97,69%, lebih tinggi dibandingkan *Naive Bayes* yang mencapai 96,53%[14]. Penelitian lain yang menggunakan model *CRISP-DM* menunjukkan bahwa algoritma *SVM* memiliki keunggulan dibandingkan *Naive Bayes* dan *k-NN*, dengan rata-rata akurasi mencapai 96,43%. Hasil menunjukkan desain 94,40%, harga 97,44%, spesifikasi 96,22%, dan citra merek 97,63%[15]. Berdasarkan temuan tersebut, dapat disimpulkan bahwa *SVM* secara konsisten menunjukkan kinerja yang lebih unggul dibandingkan algoritma lainnya dalam berbagai studi.

Studi ini mengarah pada analisis sentimen yang terdapat dalam ulasan pengguna aplikasi Vidio di *Play Store* dengan memanfaatkan metode algoritma *Support Vector Machine (SVM)*. Algoritma ini dipilih karena penelitian terdahulu menunjukkan performanya yang unggul dibandingkan metode lainnya. Selain itu, penting untuk memperhatikan kendala yang muncul dalam analisis ulasan, seperti keberagaman bahasa dan konteks dalam setiap ulasan yang dapat menyebabkan kesulitan dalam klasifikasi sentimen secara akurat. Untuk itu, Penelitian ini difokuskan untuk menganalisis sentimen ulasan pengguna aplikasi Vidio di *Play Store* dan mengevaluasi akurasi algoritma *SVM* efektif dalam mengklasifikasikan sentimen tersebut.

2. METODE PENELITIAN

Penelitian ini menggunakan algoritma *Support Vector Machine (SVM)* dalam kerangka *Knowledge Discovery in Database (KDD)* untuk menganalisis data. *KDD* adalah proses ilmiah data mining yang mencakup analisis data, pembuatan model, dan identifikasi pola tersembunyi. Gambar 1 menunjukkan tahapan penelitian dengan metode *KDD*.



Gambar 1. Proses penelitian
(*sumber: dokumen pribadi)*

2.1 Data Selection

Pada langkah seleksi data dalam proses *KDD*, dilakukan pengambilan, pemilihan, dan pemberian label pada data. Ulasan pengguna aplikasi Vidio dikumpulkan dari *Google Play Store* selama periode 15 September hingga 21 November 2024. Data yang terkumpul disimpan dalam format .csv agar lebih mudah diakses. Pengumpulan data dilakukan dengan teknik *web scraping* menggunakan *library google-play-scraper* untuk memperoleh ulasan langsung dari *platform Google Play*.

2.2 PreProccesing

Proses *preprocessing* data bertujuan untuk mengubah data yang tidak terstruktur menjadi bentuk yang lebih terorganisir dan siap untuk dianalisis. Proses ini terdiri dari beberapa langkah:

1. *Case Folding*
2. *Cleaning*
3. *Normalization*
4. *Filtering*
5. *Stemming*
6. *Tokenization*

2.3 Transformation

Tahap transformasi merupakan langkah untuk mengubah data menjadi format yang siap diproses dalam tahap data mining. Pada tahap ini, data akan dibagi menjadi data *Training* dan data *Testing* dalam tiga skenario: 90:10, 80:20, dan 70:30. Selanjutnya, pembobotan kata dilakukan dengan metode *TF-IDF* untuk mengonversi data teks menjadi vektor dengan bobot. Perhitungan *TF-IDF* akan dijelaskan dengan rumus yang tertera berikut.

$$\omega_t = TF_{t,d} x IDF_t = TF_{t,d} \times log\left(\frac{N}{DF_t}\right) \quad (1)$$

2. 4 Data Mining

Pada tahap data mining, klasifikasi sentimen terhadap data ulasan dilakukan dengan menerapkan algoritma *Support Vector Machine (SVM)* yang menggunakan empat tipe *kernel*, yaitu *Linear, RBF, Sigmoid, dan Polynomial*. Rumus yang digunakan untuk setiap jenis *kernel* tersebut adalah sebagai berikut:

1. *Linear*, rumusnya: $k(x_i, x) = x_i \cdot x$ (2)
2. *RBF*, rumusnya: $k(x_i, x) = exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$ (3)
3. *Polynomial*, rumusnya: $k(x_i, x) = (x_i \cdot x)^d$ (4)
4. *Sigmoid*, rumusnya: $k(x_i, x) = tanh(\sigma(x_i \cdot x) + c)$ (5)

2. 5 Evaluation

Pada tahap evaluasi, performa setiap *kernel* dari algoritma *SVM* dinilai dengan menggunakan confusion matrix. Proses ini menghasilkan metrik evaluasi yang mencakup akurasi, presisi, *recall*, serta *F1-score*.

3. HASIL DAN PEMBAHASAN

3.1 Data Selection

Pada tahap pengumpulan data ulasan aplikasi vidio, dilakukan dengan menggunakan teknik *web scraping* dengan memanfaatkan *google-play-scraper*. Alat ini memudahkan proses pengambilan data dari *Google Play Store* dengan menyediakan *API* dan tersedia sebagai pustaka dalam bahasa pemrograman *python*. Pada tahap ini peneliti mengambil sebanyak 2000 ulasan yang di ambil dari ulasan terakhir, dari tanggal 15 september sampai 21 november 2024.

userName	rating	tanggal	ulasan
Idan Rodiaman	1	21/11/2024 15:07	Metode pembayaran menggunakan Telkomsel ditolak tanpa penjelas. Sangat menjengkelkan.
hendrik malonda	5	21/11/2024 14:46	Bagus
Aqmal Nurfadilah	5	21/11/2024 14:17	Apk streaming terbaik menurut saya selain yt
Mutiara Syifa Cindaka	2	21/11/2024 14:17	Banyak iklan dan apk nya jadi jelek
Andi Whisnu Syam (E	4	21/11/2024 12:47	Sisa 1 ðŸŒŸ klu bagus
Febri Anto	1	21/11/2024 12:05	Sudah beelangganan sejak lama, dan otomatis perpanjang tetapi justru saat ini paketnya selalu minta beli ulang padahal sudah ada emai
Abang eddy	1	21/11/2024 11:55	Gak pernah pake apk ini, terpaksa download cuma buat nonton byon doang sih hehe
Airen Nadir	5	21/11/2024 11:11	Nice
Abizard Steampunk	5	21/11/2024 10:17	Mines gk bisa hapus history,waduh makin kesini makin banyak vareasi filmnya baguss pools developer nyaa manataps
aditya rizky ramadhar	2	21/11/2024 09:51	Mau nonton liga 1 video nya patahÃ² atau nge leg parah , padhal pakai wifi dan sdh di uninstall ulang tetap sama pdahal sudah bayar jdi
Jesica Chyta	1	21/11/2024 09:24	apk kntl lemot ck dnck
bagus kristiawan	1	21/11/2024 09:01	sering lag
Bujang Minang Adven	1	21/11/2024 09:00	Saya kecewa Sudah isi pulsa tapi ditolak mau beli paket platinum
Nurfadillahwtp233	1	21/11/2024 08:28	gimana inii mohon maaf,katanya tunggu kode dulu tapi kode nya nda di kirim udah cape" nunggu,mohon tolong di perbaiki yah
Oozy Shop	5	21/11/2024 07:12	Okay
Wahyu Pranata	2	21/11/2024 07:06	Lemot apk nya

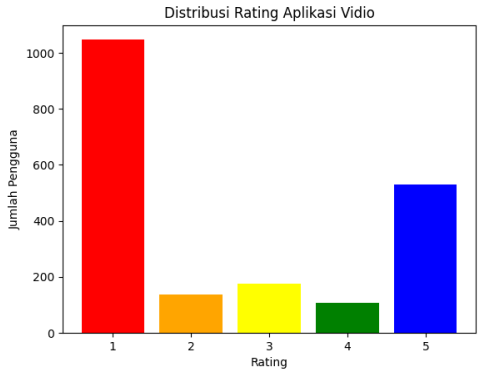
Gambar 2. Hasil Scrapping Data Ulasan Aplikasi Vidio

Pada data ulasan aplikasi vidio terdapat 4 atribut, yaitu atribut "*userName*","*rating*","tanggal" dan "ulasan. Namun untuk proses pelabelan dalam penelitian ini, peneliti menggunakan *rating* yang di berikan pengguna melalui ulasan aplikasi vidio di *Google Play Store*. Dimana proses pelabelan ini di bagi menjadi dua kategori: label negatif diberikan untuk *rating* dari 1 hingga 3, sementara label positif diberikan untuk *rating* 4 dan 5.

Tabel 1. Jumlah Ulasan Tiap *Rating*

Rating	**Jumlah**
1	1047
2	137
3	176
4	109
5	531
Total	2000

Menurut Tabel 1. jumlah ulasan setiap *rating* menunjukkan bahwa *rating* 1 memiliki jumlah terbanyak, yaitu 1047 ulasan, selanjutnya yang memiliki jumlah terbanyak, yaitu *rating* 5 dengan 531 ulasan, dan ranting 4 memiliki jumlah paling sedikit, yakni 109.



Gambar 3. Grafik *Rating* Ulasan

Sentimen negatif dalam dataset ini terdiri dari 1360 ulasan, yang diidentifikasi berdasarkan pemberian *rating* 1 hingga 3. Sementara itu, sentimen positif mencakup 640 ulasan, yang ditandai dengan pemberian *rating* 4 dan 5. Oleh karena itu, lebih banyak ulasan yang memiliki sentimen negatif dibandingkan dengan yang positif.

3.2. Preprocessing

- *Case Folding*

  Tahap ini melibatkan proses standarisasi format teks dengan mengubah semua huruf dalam dokumen menjadi huruf kecil. Tujuannya adalah untuk menghindari kesalahan

identifikasi kata yang seharusnya sama, tetapi berbeda akibat penggunaan huruf kapital atau huruf kecil.

Tabel 2. Sebelum dan Sesudah Case Folding

Sebelum Case Folding	Sesudah Case Folding
Apa coba buffering, lag" padahal sinyal baik" Aja Rusak apk lu bng!! anehh sia" 20k gw mending beli naspad knyang seharian, dari pada kesal sebulan gra" Apk gjls"	apa coba buffering, lag" padahal sinyal baik" aja rusak apk lu bng!! anehh sia" 20k gw mending beli naspad knyang seharian, dari pada kesal sebulan gra" apk gjls"
Alhamdulillah Lumayan Buat Sementara... Moga Bisa Ke Beli TV....	alhamdulillah lumayan buat sementara... moga bisa ke beli tv....
Saya suka berlangganan aplikasi Vidio	saya suka berlangganan aplikasi vidio

- *Cleaning*

Pada tahap ini bertujuan untuk menyaring elemen-elemen yang tidak berkontribusi pada analisis, memastikan data ulasan hanya terdiri dari karakter *alfabet* yang relafan. Dimana pada tahap ini melibatkan berbagai proses pembersihan data, seperti menghapus karakter special, angka, nilai kosong (*missing values*), symbol symbol-simbol seperti '!'#%&*'=-()/?<>+?[@]^$, serta atribut yang tidak relavan seperti tanda baca, spasi kosong, *URL*, *hashtag* dan emoji.

Tabel 3. Sebelum dan Sesudah Data *Cleaning*

Sebelum *Cleaning*	**Sesudah *Cleaning***
apa coba buffering, lag" padahal sinyal baik" aja rusak apk lu bng!! anehh sia" 20k gw mending beli naspad knyang seharian, dari pada kesal sebulan gra" apk gjls"	apa coba buffering lag padahal sinyal baik aja rusak apk lu bng anehh sia 20k gw mending beli naspad knyang seharian dari pada kesal sebulan gra apk gjls
alhamdulillah lumayan buat sementara... moga bisa ke beli tv....	alhamdulillah lumayan buat sementara moga bisa ke beli tv
saya suka berlangganan aplikasi vidio	saya suka berlangganan aplikasi vidio

- *Normalization*

Pada tahap ini bertujuan untuk memperbaiki ejaan yang salah, termasuk kesalahan pada kata, agar kata-kata dengan makna yang sama dapat disamakan. Pada tahap ini juga, setiap kata singkatan di ubah menjadi bentuk lengkapnya, seperti "bgt" yang disesuaikan menjadi "banget", "tdk" menjadi "tidak", dan singkatan lainnya.

Tabel 4. Sebelum dan Sesudah *Normalization*

Sebelum *Normalization*	**Sesudah *Normalization***
apa coba buffering lag padahal sinyal baik aja rusak apk lu bng anehh sia 20k gw mending beli naspad knyang seharian dari pada kesal sebulan gra apk gjls	apa coba buffering lag padahal sinyal baik aja rusak aplikasi lu bang aneh sial 20 ribu saya mending beli nasi padang kenyang seharian dari pada kesal sebulan gara aplikasi ga jelas
alhamdulillah lumayan buat sementara moga bisa ke beli tv	alhamdulillah lumayan buat sementara semoga bisa ke beli tv
saya suka berlangganan aplikasi vidio	saya suka berlangganan aplikasi vidio

- *Filtering*

Pada tahap *Filtering* ini bertujuan untuk memproses penyaringan data dengan menghapus kata-kata yang tidak memiliki nilai informasi signifikan atau tidak relevan untuk analisis. Kata-kata tersebut biasanya termasuk *stopwords*, yaitu merupakan kata-kata yang umum seperti "di", "yang", "dan" serta kata-kata umum lainnya.

Tabel 5. Sebelum dan Sesudah *Filtering*

Sebelum *Filtering*	**Sesudah *Filtering***
apa coba buffering lag padahal sinyal baik aja rusak aplikasi lu bang aneh sial 20 ribu saya mending beli nasi padang kenyang seharian dari pada kesal sebulan gara aplikasi ga jelas	coba buffering lag padahal sinyal baik aja rusak aplikasi lu bang aneh sial 20 ribu mending beli nasi padang kenyang seharian kesal sebulan gara aplikasi ga jelas
alhamdulillah lumayan buat sementara semoga bisa ke beli tv	alhamdulillah lumayan buat sementara semoga bisa beli tv
saya suka berlangganan aplikasi vidio	suka berlangganan aplikasi vidio

- *Stemming*

Pada tahap *Stemming* ini, kata-kata yang mengandung imbuhan, baik pada kata awalan maupun akhir, akan diubah menjadi kata baku atau bentuk dasar, untuk menghasilkan kata yang lebih sederhana dan sesuai dengan kaidah Bahasa. Proses ini umumnya dilakukan dengan memanfaatkan pustaka *Python Sastrawi*, yang dirancang khusus untuk pemrosesan bahasa Indonesia.

Tabel 6. Sebelum dan Sesudah *Stemming*

Sebelum *Stemming*	**Sesudah *Stemming***
coba buffering lag padahal sinyal baik aja rusak aplikasi lu bang aneh sial 20 ribu mending beli nasi padang kenyang seharian kesal sebulan gara aplikasi ga jelas	coba buffering lag padahal sinyal baik aja rusak aplikasi lu bang aneh sial 20 ribu mending beli nasi padang kenyang hari kesal bulan gara aplikasi ga jelas
alhamdulillah lumayan buat sementara semoga bisa beli tv	alhamdulillah lumayan buat sementara moga bisa beli tv
suka berlangganan aplikasi vidio	suka langgan aplikasi vidio

- *Tokenization*

Pada tahap terakhir ini, yaitu tahap *Tokenization* atau tokenisasi Dimana pada tahap ini bertujuan untuk memecah kalimat menjadi kata-kata yang terpisah dengan membedakan karakter pemisah seperti tanda baca atau spasi, sehingga pada setiap kata dalam teks dapat dianalisis sebagai unit terpisah. Pada proses ini bertujuan untuk memudahkan analisis teks karena memungkinkan system untuk menganalisis atau mengidentifikasi setiap kata secara individu.

Tabel 7. Sebelum dan Sesudah *Tokenization*

Sebelum *Tokenization*	**Sesudah *Tokenization***
coba buffering lag padahal sinyal baik aja rusak aplikasi lu bang aneh sial 20 ribu mending beli nasi padang kenyang hari kesal bulan gara aplikasi ga jelas	['coba', 'buffering', 'lag', 'padahal', 'sinyal', 'baik', 'aja', 'rusak', 'aplikasi', 'lu', 'bang', 'aneh', 'sial', '20', 'ribu', 'mending', 'beli', 'nasi', 'padang', 'kenyang', 'hari', 'kesal', 'bulan', 'gara', 'aplikasi', 'ga', 'jelas']
alhamdulillah lumayan buat sementara moga bisa beli tv	['alhamdulillah', 'lumayan', 'buat', 'sementara', 'moga', 'bisa', 'beli', 'tv']

suka langgan aplikasi vidio	['suka', 'langgan', 'aplikasi', 'vidio']

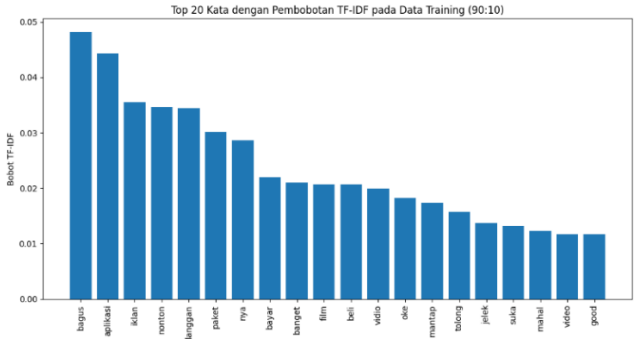
3.3 Transformation

Pada tahap transformasi ini, proses dimulai dengan membagi dataset menjadi dua bagian utama: data *Training* dan *Testing*, yang dilakukan dengan tiga skenario yang berbeda. Untuk memastikan konsistensi dalam pembagian data pada setiap percobaan, kami memanfaatkan fungsi *train_test_split* dengan parameter *random_state=42*. Ini menjamin bahwa pembagian data akan tetap serupa setiap kali eksperimen dilakukan. Pembagian data yang dilakukan tercatat pada Tabel 8, yang menunjukkan hasil untuk masing-masing skenario. Pembagian ini dilakukan untuk menguji kinerja model terhadap data yang berbeda-beda

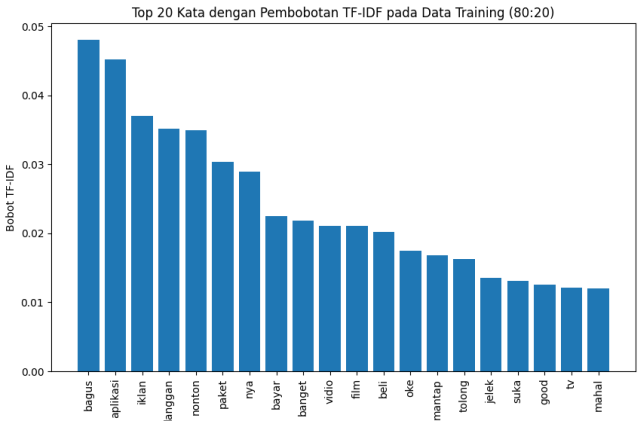
Tabel 8. Pemisahan Data

Rasio Pembagian		Jumlah Data *Training*	Jumlah Data *Testing*
Data *Training*	Data *Testing*		
90%	10%	1800	200
80%	20%	1600	400
70%	30%	1400	600

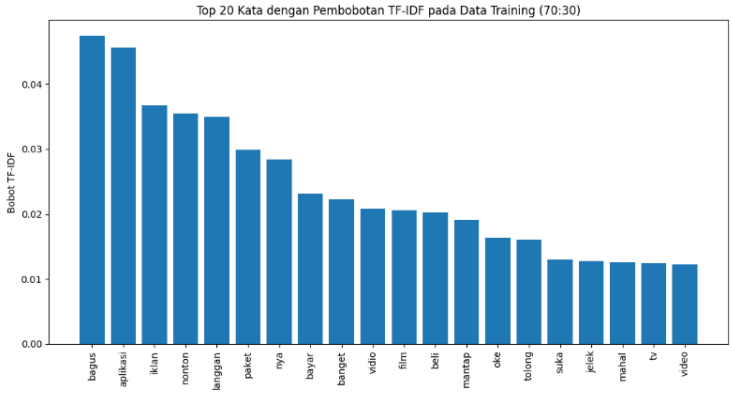
Setelah data dipisahkan ke dalam kelompok *Training* dan *Testing*, langkah berikutnya adalah melakukan transformasi dengan menerapkan metode *TF-IDF* untuk pembobotan kata. Pada proses ini mengubaha kata-kata menjadi bentuk *numerik*, yang menghitung frekuensi token dalam dokumen dan menghasilkan bobot untuk analisis lebih lanjut. Pada gambar hasil pembobotan data dibawah ini, menggunakan 20 kata yang memiliki nilai tertinggi atau yang paling menonjol berdasarkan kriteria tertentu.



Gambar 4. Contoh penerapan metode *TF-IDF* (90:10)



Gambar 5. Contoh penerapan metode *TF-IDF* (80:20)



Gambar 6. Contoh penerapan metode *TF-IDF* (70:30)

3.4. Data Mining

Pada tahap ini, proses klasifikasi sentimen dilakukan dengan membagi data menjadi tiga skenario yang berbeda yaitu 90:10, 80:20, 70:30, dan menggunakan empat jenis karnel (*Linear*, *RBF*, poly dan *Sigmoid*) pada algoritma *SVM* untuk mengevaluasi performa model. Hasil dari tahap klasifikasi setiap karnel dan skenario disajikan pada Tabel 9.

Tabel 9. Hasil klasifikasi *SVM*

Skenario	***Kernel***	***Accuracy***
90:10	*Linear*	78.12%
90:10	*RBF*	79.69%
90:10	*poly*	79.69%
90:10	*Sigmoid*	78.12%
80:20	*Linear*	82.03%
80:20	*RBF*	82.03%
80:20	*poly*	81.25%
80:20	*Sigmoid*	82.03%
70:30	*Linear*	82.81%
70:30	*RBF*	82.29%
70:30	*poly*	82.29%
70:30	*Sigmoid*	83.33%

3.5. *Evaluation*

Evaluation merupakan tahap akhir penelitian, Pada langkah ini, dilakukan proses evaluasi terhadap kinerja model untuk menilai efektivitasnya yang dibagun menggunakan *SVM* (*Support Vector Machine*) melalui metode *confusion matrix*. Pada tahap Evaluasi ini membandingkan kinerja model dengan empat *kernel SVM* yang digunakan dalam berbagai skenario data mining.

Skenario	Kernel	Accuracy	Precision	Recall	F1-Score
90:10	linear	78.12%	61.04%	78.12%	68.53%
90:10	rbf	79.69%	83.88%	79.69%	72.05%
90:10	poly	79.69%	83.88%	79.69%	72.05%
90:10	sigmoid	78.12%	61.04%	78.12%	68.53%
80:20	linear	82.03%	79.45%	82.03%	75.87%
80:20	rbf	82.03%	85.29%	82.03%	74.66%
80:20	poly	81.25%	75.79%	81.25%	74.21%
80:20	sigmoid	82.03%	85.29%	82.03%	74.66%
70:30	linear	82.81%	77.46%	82.81%	75.96%
70:30	rbf	82.29%	68.5%	82.29%	74.77%
70:30	poly	82.29%	68.5%	82.29%	74.77%
70:30	***sigmoid***	***83.33%***	***86.13%***	***83.33%***	***76.25%***

Gambar 7. Perbandingan Hasil *Evaluation*

Gambar 6 menunjukkan bahwa scenario 3 dengan *kernel Sigmoid* memberikan hasil yang terbaik, dengan menghasilkan *accuracy* 83.33%, *precision* 86.13%, *recall* 83.33%, dan *F1-score*

76.25%. Model ini berhasil mencapai komposisi yang optimal antara *precision* dan *recall*, meskipun ada sedikit kompromi yang terlihat pada *F1-score*. Secara keseluruhan, performa model ini menunjukkan hasil yang solid, dengan kinerja yang baik dalam mengklasifikasikan data negatif maupun positif.

Untuk mengidentifikasi kata-kata yang dominan, fungsi *wordcloud* dapat digunakan, yang membantu dalam menemukan tema utama atau kata kunci dalam *dataset*, mempermudah pemahaman terhadap sentimen yang diungkapkan oleh pengguna. Dengan menggunakan wordcloud, kata-kata yang dominan dalam kemunculannya dalam teks akan ditampilkan, di mana kata-kata dengan frekuensi tinggi akan terlihat lebih besar.



Gambar 8. Ulasan *wordcloud* positif



Gambar 9. Ulasan *wordcloud* negatif

Berdasarkan gambar 8 dan 9, penerapan analisis visual menggunakan wordcloud berhasil mengidentifikasi kata-kata utama dalam ulasan pengguna di *Play Store*. Pada ulasan positif, kata-kata seperti “bagus”, “aplikasi”, dan “nonton” sering muncul, menunjukkan pengalaman pengguna yang baik terhadap aplikasi. Sebaliknya, ulasan negatif didominasi oleh kata-kata seperti “tidak”, “aplikasi”, “nonton”, dan “langganan”, yang mencerminkan area yang perlu diperbaiki oleh pengembang aplikasi.

Secara keseluruhan, Penelitian ini berhasil menunjukkan efektivitas algoritma *SVM* dengan *kernel sigmoid* dalam mengidentifikasi sentimen pengguna. Penelitian ini juga menyajikan wawasan berharga tentang pola sentimen pengguna, yang dapat dimanfaatkan oleh pengembang untuk meningkatkan dan mengembangkan aplikasi di masa depan.

4. KESIMPULAN

Dari hasil penelitian yang dilakukan, kesimpulannya adalah sebagai berikut:

1) Penelitian ini menggunakan algoritma *Support Vector Machine (SVM)* untuk menganalisi 2000 ulasan pada aplikasi vidio pada *Play Store*. Kemudian dilakukan proses *PreProccesing* , pembobotan menggunakan *TF-IDF*, dan klasifikasi menggunakan algoritma *Support*

Vector Machine (SVM), dengan menggunakan tiga skanario pembagian data yaitu 90:10, 80:20, 70:30, dan menggunakan empat jenis karnel (linier, *RBF*, *Polynomial* dan *Sigmoid*).

2) Pada hasil penelitian menunjukkan bahwa scenario 70:30 dengan *kernel Sigmoid* memberikan hasil yang terbaik, dengan menghasilkan *accuracy* 83.33%, *precision* 86.13%, *recall* 83.33%, dan *F1-score* 76.25. Model ini berhasil mencapai komposisi yang optimal antara *precision* dan *recall*, meskipun ada sedikit kompromi yang terlihat pada *F1-score*. Penelitian ini menegaskan keunggulan *SVM* sebagai algoritma yang efektif dalam analisis sentimen.

5. SARAN

Adapun sejumlah saran untuk penelitian mendatang dapat dijelaskan sebagai berikut:

1) Untuk pengembangan model, Uji parameter *kernel SVM* seperti C, gamma, dan derajat *Polynomial*, serta coba metode embedding teks seperti *Word2Vec* atau *BERT*. Perluas dataset dengan ulasan dari aplikasi lain dan lakukan validasi silang untuk meningkatkan akurasi model.
2) Selanjutnya, Gunakan metrik tambahan seperti AUC-ROC untuk hasil yang lebih akurat, dan manfaatkan analisis sentimen untuk membantu pengembang aplikasi Vidio dalam meningkatkan kualitas layanan mereka.

UCAPAN TERIMA KASIH

Saya mengucapkan rasa syukur kepada Tuhan Yang Maha Esa atas segala rahmat dan kekuatan yang diberikan selama penelitian ini. Terima kasih yang mendalam saya sampaikan kepada orang tua saya atas dukungan dan doa yang tak pernah putus. Saya juga ingin mengucapkan terima kasih kepada dosen pembimbing yang dengan penuh kesabaran telah memberikan arahan dan masukan yang sangat berarti dalam penyusunan penelitian ini. Selain itu, terima kasih saya haturkan kepada sahabat-sahabat saya yang senantiasa memberikan dukungan dan bantuan sepanjang proses penelitian ini.

DAFTAR PUSTAKA

[1] A. A. Muhammad, Ermatita, and D. S. Prasvita, 2022, "ANALISIS SENTIMEN PENGGUNA APLIKASI DANA BERDASARKAN ULASAN PADA GOOGLE PLAY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE Prodi S1 Informatika / Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, no. 2, vol. 3 pp. 194–204, https://conference.upnvj.ac.id/index.php/senamika/article/view/2171.

[2] N. Lestari, E. Haerani, and R. M. Candra, 2023, "Analisa Sentimen Ulasan Aplikasi Wetv Untuk Peningkatan Layanan Menggunakan Metode Naïve Bayes," *J. Inf. Syst. Res.*, no. 3, vol. 4, pp. 874–882, doi: 10.47065/josh.v4i3.3355.

[3] M. Yusuf Rismanda Gaja, I. Maulana, and O. Komarudin, 2024, "Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan Playstore Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, no. 4, vol. 7, pp. 2767–2774, doi: 10.36040/jati.v7i4.7197.

[4] A. Bagas Pranata, A. R. Abdillah, and F. Irwiensyah, 2024, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Pengguna Aplikasi Netflix Pada Google Play Menggunakan Algoritma Naïve Bayes," *Media Online)*, no. 6, vol. 4, pp. 3091–3098, doi: 10.30865/klik.v4i6.1964.

[5] M. Fudhail Ferio Supeli and S. Setiaji, 2023, "Klasifikasi Sentimen Positif Dan Negatif Pada Aplikasi Vidio Dengan Algoritma K-Nearest Neighbor," *Indones. J. Comput. Sci.*,

no. 1, vol. 2, pp. 7–15, doi: 10.31294/ijcs.v2i1.1874.

[6] S. A. Aaputra, Didi Rosiyadi, Windu Gata, and Syepry Maulana Husain, 2019, "Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, no. 3, vol. 3, pp. 377–382, doi: 10.29207/resti.v3i3.1118.

[7] M. D. Hendriyanto, A. A. Ridha, and U. Enri, 2022, "Analisis Sentimen Ulasan Aplikasi Mola Pada *Google Play Store* Menggunakan Algoritma Support Vector Machine," *INTECOMS J. Inf. Technol. Comput. Sci.*, no. 1, vol. 5, pp. 1–7, doi: 10.31539/intecoms.v5i1.3708.

[8] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, 2023, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, no. 2, vol. 10, pp. 176–184, doi: 10.37373/tekno.v10i2.419.

[9] N. Pratiwi and Y. Setyawan, 2021, "Analisis Akurasi Dari Perbedaan Fungsi *Kernel* Dan Cost Pada Support Vector Machine Studi Kasus Klasifikasi Curah Hujan Di Jakarta," *J. Fundam. Math. Appl.*, no. 2, vol. 4, pp. 203–212, doi: 10.14710/jfma.v4i2.11691.

[10] F. Bei and S. Sudin, 2021 "Analisis Sentimen Aplikasi Tiket Online Di *Play Store* Menggunakan Metode *Support Vector Machine (SVM)*," *Sismatik*, no. 01, vol. 01, pp. 91–97, https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/13/12.

[11] N. Zelina and A. Afiyati, 2024, "Analisis Sentimen Ulasan Pengguna Aplikasi M-Banking Menggunakan Algoritma Support Vector Machine dan Decision Tree," *J. Linguist. Komputasional*, no. 1, vol. 7, pp. 31–37, doi: 10.26418/jlk.v7i1.169.

[12] U. Kusnia, F. Kurniawan, and S. Artikel, 2022, "Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (*SVM*) Dan Naive Bayes INFO ARTIKEL ABSTRAK," *J. Keilmuan dan Apl. Tek. Inform.*, no. 36, vol. 14(1), pp. 24–25, [Online]. Available: https://doi.org/10.35891/explorit

[13] D. Atmajaya, A. Febrianti, and H. Darwis, 2023, "Metode *SVM* dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter," *Indones. J. Comput. Sci.*, no. 4, vol. 12, pp. 2173–2181, doi: 10.33022/ijcs.v12i4.3341.

[14] A. Puji Astuti, S. Alam, and I. Jaelani, 2022, "Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo," *J. Bangkit Indones.*, no. 2, vol. 11, pp. 1–6, doi: 10.52771/bangkitindonesia.v11i2.196.

[15] J. W. Iskandar and Y. Nataliani, 2021, "Perbandingan Naïve Bayes, *SVM*, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, no. 6, vol. 5, pp. 1120–1126, doi: 10.29207/resti.v5i6.3588.