

ANALISIS SENTIMEN PADA ULASAN APLIKASI THREADS DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Farhan Nufairi¹⁾, Nunik Pratiwi*²⁾, Fariz Herlando³⁾

1. Universitas Muhammadiyah Prof Dr.Hamka, Indonesia
2. Universitas Muhammadiyah Prof Dr.Hamka, Indonesia
3. Universitas Muhammadiyah Prof Dr.Hamka, Indonesia

Article Info

Kata Kunci: *Support Vector Machine, Threads, Analisis Sentimen*

Keywords: *Support Vector Machine, Threads, Sentiment Analysis*

Article history:

Received 1 December 2023

Revised 15 December 2023

Accepted 29 December 2023

Available online 1 March 2024

DOI :

<https://doi.org/10.29100/jipi.v9i1.4929>

* Corresponding author.

Nunik Pratiwi

E-mail address:

npratiwi@uhamka.ac.id

ABSTRAK

Threads merupakan sebuah aplikasi media sosial yang dirilis oleh *Instagram* pada tahun 2023 bulan Juli lalu. Di *Google Play Store* sendiri aplikasi *Threads* telah menjadi populer, dengan lebih dari 50 juta unduhan dan rating rata-rata 4,5. Terkadang, isi rating dan ulasan suatu produk atau layanan dapat berbeda. Dalam mengembangkan aplikasi, perlu memperhatikan pendapat dari pengguna, bukan hanya jumlah rating. Oleh karena itu, penelitian ini bertujuan untuk mengetahui sentimen pengguna aplikasi *Threads* di *Google Play Store*, baik positif maupun negatif, dan untuk mengetahui tingkat akurasi algoritma *Support Vector Machine* (SVM) dalam menganalisis sentimen tersebut. Metode penelitian melibatkan pengumpulan ulasan pengguna *Threads*, *preprocessing*, pelabelan, dan penerapan algoritma SVM untuk mengklasifikasikan sentimen ulasan menjadi positif dan negatif. Setelah melakukan implementasi algoritma, pengujian evaluasi merupakan tahap terakhir dari proses implementasi algoritma. Data yang digunakan dalam penelitian ini adalah data ulasan aplikasi *Threads* di *Google Play Store*, yang berjumlah 1429 data. Dari jumlah tersebut, 1103 merupakan ulasan positif dan 326 merupakan ulasan negatif. Hasil pengujian evaluasi menunjukkan bahwa algoritma SVM dapat mengklasifikasi ulasan aplikasi *Threads* dengan akurasi sebesar 88%. Akurasi ini diperoleh dengan menggunakan rasio pembagian data sebesar 80:20.

ABSTRACT

Threads is an application released by *Instagram* on July 2023. *Threads* has already become a well-known application in *Google Play Store*, with more than 50 million downloads, and rated 4,5 on average. Occasionally, the content of a rating and review on a product or service, could be various. It needs review from users, not only ratings, to develop an application. In accordance with the statement mentioned before, the purpose of this research is to find positive and negative sentiments towards *Threads* in *Google Play Store*, and to find the algorithm accuracy of *Support Vector Machine* (SVM) on analyzing those sentiments. Methods used in this research are collecting reviews from *Threads* users, *preprocessing*, labeling, and application of SVM algorithm to classify positive and negative reviews. After the implementation of the algorithm, the final step of the algorithm implementation is evaluation test activity. Datas used in this research are reviews of *Threads* application in *Google Play Store*, in the quantity of 1429 datas. Out of 1429 datas mentioned before, 1103 are positive reviews and 326 are negative reviews. The result of evaluation test activity is SVM algorithm is able to classify *Threads* reviews with 88% accuracy. The accuracy is obtained by using data division in 80:20 ratio.

I. PENDAHULUAN

Perkembangan teknologi dalam lingkup internet telah mengalami kemajuan yang sangat cepat, mengenai berbagai aspek kehidupan manusia, mulai dari transformasi dalam sektor perdagangan yang melibatkan *e-commerce* hingga transformasi dalam metode pendidikan dengan penggunaan pembelajaran online, bahkan hingga pengaruhnya yang signifikan dalam mengubah cara manusia berinteraksi dan berkomunikasi secara sosial melalui berbagai platform dan media sosial yang ada. Di Indonesia, pertumbuhan pengguna internet telah mencapai

tingkat yang signifikan dalam dekade terakhir. Pada tahun 2013, jumlah pengguna internet di Indonesia hanya 71,19 juta. Namun, jumlah tersebut meningkat secara signifikan menjadi 215,63 juta pada tahun 2023, menurut survei APJII [1]. Berdasarkan data dari GoodStats, Indonesia memiliki jumlah pengguna internet terbanyak ke-4 di dunia [2]. Indonesia memiliki jumlah pengguna *smartphone* aktif yang sangat besar, mencapai 73 juta orang, sehingga berada di posisi keenam dalam jajaran negara dengan pengguna *smartphone* terbanyak di dunia [3]. Bahkan dibandingkan negara lain di kawasan Asia Tenggara, kecepatan internet di beberapa kota besar Indonesia tergolong rendah [4]. Sebagai negara dengan populasi yang sangat besar dan pertumbuhan konektivitas internet yang pesat, Indonesia menyajikan sebuah konteks yang sangat menarik untuk memahami dampak yang ditimbulkan oleh media sosial terhadap budaya, perilaku, dan dinamika sosial masyarakat.

Threads, platform media sosial terbaru yang dirilis pada tahun 2023 bulan Juli lalu oleh *Instagram*, telah mendapatkan popularitas di kalangan pengguna muda. Mark Zuckerberg, sebagai pemimpin *Meta*, telah menyatakan bahwa tujuan dirilisnya aplikasi *Threads* adalah untuk mengungguli *Twitter*. Pengguna *Threads* juga bisa memposting hingga 500 karakter dan memiliki fitur yang serupa dengan *Twitter* [5]. Secara keseluruhan, *Threads* memiliki konsep yang mirip dengan *Twitter* dalam hal memberikan pengguna platform untuk berbagi unggahan berbasis teks. *Threads* dan *Twitter* memiliki kesamaan, tetapi juga memiliki perbedaan. *Threads* terintegrasi dengan *Instagram*, memiliki fitur *Hidden Words* untuk memblokir kata atau frasa tertentu, tetapi belum memiliki fitur *direct messaging* (DM), trending, dan iklan. *Threads* juga menyertakan fitur seperti kemampuan melihat profil, memeriksa mention, dan beberapa fungsi lainnya [6]. Dalam penelitian ini, ulasan pengguna aplikasi *Threads* di *Google Play Store* digunakan. Data yang telah terkumpul akan diproses untuk diklasifikasikan menggunakan algoritma *Support Vector Machine*. Ada beberapa kasus di mana rating tidak sesuai dengan ulasanya [7]. Aplikasi *Threads* di *Google Play Store* mendapatkan ulasan sangat beragam dan tidak teratur, oleh karena itu diperlukan pendekatan analisis sentimen untuk memahami ulasan tersebut [8]. Dalam proses mengembangkan aplikasi, jumlah ulasan dan komentar pengguna harus diperhatikan. Dalam mengembangkan sebuah aplikasi, ada faktor lain yang juga perlu dipertimbangkan, selain jumlah rating, yaitu konten ulasan pengguna. Sehingga, pengembang dapat menjelaskan setiap ulasan pengguna.

Analisis sentimen adalah metode untuk memahami persepsi individu terhadap suatu subjek, dengan hasil akhir yang diklasifikasikan sebagai positif atau negatif [9]. Analisis sentimen juga merupakan proses otomatis untuk mengekstraksi data teks, dengan tujuan memperoleh informasi yang dilakukan terhadap suatu objek, termasuk dalam bentuk ulasan positif dan negatif. Sentimen masyarakat dapat menjadi acuan dalam proses pengambilan keputusan terkait suatu produk [10].

Algoritma *Support Vector Machine* (SVM) merupakan salah satu metode yang paling sering digunakan untuk klasifikasi teks. SVM adalah metode klasifikasi data yang menggunakan *supervised learning*. SVM menunjukkan kinerja yang lebih baik daripada metode *machine learning* lainnya [11]. Keunggulan SVM adalah dapat memisahkan data input yang bersifat *nonlinear* dan memiliki dimensi yang besar secara efektif, dengan memanfaatkan fungsi kernel [12]. SVM dapat memprediksi data baru dengan cepat dan akurat, bahkan jika data tersebut tidak termasuk dalam data pelatihan. Kelebihan lainnya adalah proses prediksi tersebut tidak dipengaruhi oleh jumlah variabel yang dimiliki objek yang diamati [13]. Dalam beberapa penelitian, SVM terbukti memiliki akurasi yang lebih tinggi daripada metode lain. Sebagai contoh, dalam penelitian [14], k-NN memiliki akurasi sebesar 83,3%, sedangkan SVM memiliki akurasi sebesar 87,5%. Hasil penelitian lain [15] menunjukkan bahwa algoritma SVM memiliki akurasi sebesar 83%, *Random Forest* sebesar 82%, dan *Naïve Bayes* sebesar 74%. Hal ini menunjukkan bahwa SVM dapat mengklasifikasikan data dengan akurasi yang lebih tinggi daripada algoritma klasifikasi lainnya.

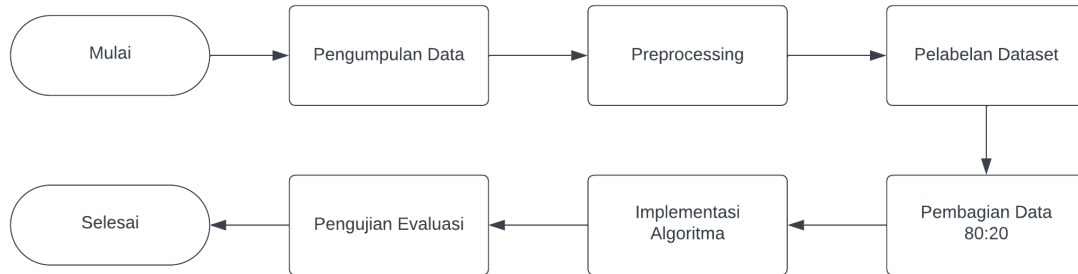
SVM juga merupakan algoritma yang dikenal luas dengan akurasi klasifikasi yang tinggi. Pada penelitian [16] terkait akurasi sebelumnya adalah algoritma SVM untuk analisis sentimen pada aplikasi *Truecaller* dan *Getcontact*, dan hasilnya membuktikan bahwa model klasifikasi SVM menunjukkan hasil terbaik, dengan akurasi 88,20% untuk *Truecaller* dan 87,90% untuk *Getcontact*. Hasil penelitian lain [17] terkait metode klasifikasi SVM untuk analisis sentimen menunjukkan bahwa pengklasifikasi SVM mencapai akurasi sebesar yaitu 95%. Bahkan dalam penelitian yang dilakukan oleh Primandani Arsi dan Retno Waluyo, *Support Vector Machine* berhasil mencapai tingkat akurasi sekitar 96,68% [18].

Berdasarkan informasi yang telah disampaikan, SVM dapat menghasilkan akurasi yang lebih tinggi daripada NBC dan k-NN dalam mengklasifikasikan sentimen pada ulasan aplikasi. Oleh karena itu, penelitian ini memutuskan untuk menggunakan algoritma *Support Vector Machine* untuk mengklasifikasikan ulasan aplikasi *Threads* berdasarkan sentimennya. Penelitian ini bertujuan untuk memahami pendapat pengguna tentang aplikasi *Threads* dengan menganalisis ulasan di *Google Play Store*. Tujuannya adalah untuk mengetahui apakah pengguna cenderung memberikan ulasan positif atau negatif, serta untuk mengevaluasi kinerja algoritma *Support Vector*

Machine dalam memprediksi sentimen pengguna. Hasil penelitian ini dapat membantu pengguna lain untuk membuat keputusan yang lebih tepat sebelum mengunduh atau menggunakan aplikasi tersebut.

II. METODE PENELITIAN

Alur skema penelitian yang akan dilakukan pada penelitian ini dapat dilihat pada gambar 1.



Gambar 1 Alur Penelitian

A. Pengumpulan Data

Tahap ini menggunakan teknik *crawling* dari *library google-play-scraper* untuk mengumpulkan data. Data dari *Google Play Store* digunakan untuk analisis sentimen, yang dapat dimanfaatkan oleh penyedia aplikasi untuk mengetahui kualitas aplikasi mereka berdasarkan ulasan pengguna [8]. Berdasarkan hal tersebut, banyak penelitian yang dilakukan untuk mengetahui sentimen analisis aplikasi di *Google Play Store* [19]. Dalam penelitian [20], diperoleh hasil akurasi optimal 79,99% untuk klasifikasi sentimen *tweet* berbahasa Indonesia dengan menggunakan data sebanyak 2000 *tweet*. Maka dari itu pengambilan data ulasan aplikasi *Threads* diambil sebanyak 2000 data dan diambil dari 5 Juli sampai 16 Juli 2023 dengan menggunakan *web scraping* yaitu *Google Colaboratory* dengan aplikasi pemrograman *Python*.

B. Preprocessing

Tahap berikutnya dalam analisis sentimen adalah *preprocessing*. Tahap ini merupakan proses untuk menghilangkan kesalahan dan ketidakteraturan pada data mentah, sehingga data menjadi lebih bersih dan siap untuk dianalisis [21]. Berikut beberapa tahapan yang akan dilakukan:

- 1) *Case folding* adalah proses untuk mengubah semua huruf dalam teks agar menjadi huruf kecil [22]. Langkah ini membantu proses pencarian dan pengolahan teks, karena tidak semua teks memiliki konsistensi dalam penggunaan huruf kapital [23].
- 2) *Cleaning* adalah proses untuk menyederhanakan data teks dengan menghilangkan data yang tidak penting atau tidak bermakna. Proses ini dilakukan untuk memudahkan proses pengolahan data teks. Proses *cleaning* dilakukan untuk menilai kualitas data yang memiliki format data atau *file* yang tidak sesuai, dengan tujuan memperoleh data yang memiliki kualitas tinggi [17].
- 3) *Tokenization* merupakan proses membagi *string* input menjadi kata-kata yang terpisah. *Tokenization* secara umum membagi kumpulan karakter dalam teks menjadi satuan kata, dengan cara membagi kata-kata tertentu yang bisa dianggap sebagai pemisah kata atau tidak [24].
- 4) *Normalization* adalah proses menyesuaikan kata-kata dengan kaidah bahasa Indonesia yang baku. Proses ini digunakan untuk memastikan bahwa kata-kata yang di olah memiliki makna yang benar sesuai dengan KBBI [25].
- 5) *Stopword Removal* merupakan proses untuk membersihkan teks dari kata-kata umum yang tidak penting. Kata-kata itu berfungsi sebagai penghubung atau pengganti kata lain, tetapi tidak memiliki makna tersendiri [16].
- 6) *Stemming* adalah proses menyederhanakan kata-kata dengan menghilangkan imbuhan dan menjadi kata dasar. Proses ini menggunakan *library sastrawi* yang telah disesuaikan dengan kaidah bahasa Indonesia, sehingga hasil *stemming* juga akan mengikuti ketentuan yang terdapat dalam *library sastrawi* [26].

C. Pelabelan Dataset

Proses selanjutnya setelah *preprocessing* adalah memberikan label sentimen pada setiap dataset. Label sentimen ini berfungsi untuk menentukan akurasi pada proses berikutnya, karena label sentimen ini menentukan akurasi pada setiap dataset.

D. Pembagian Data

Tahap selanjutnya setelah preprocessing adalah pembagian data. Untuk mendapatkan hasil akurasi yang lebih tinggi, data dibagi menjadi dua bagian dengan rasio 80:20 [27], yaitu data *training* dengan perbandingan 80% dan data *testing* dengan perbandingan 20%. Data *training* berfungsi sebagai dasar dalam pengembangan model klasifikasi untuk memprediksi data *class* yang baru sedangkan fungsi dari data testing untuk untuk menilai kinerja algoritma *Support Vector Machine* dalam memprediksi data baru [28].

E. Implementasi Algoritma Support Vector Machine

Algoritma SVM merupakan metode klasifikasi data teks yang dapat memisahkan data-data dari dua kelas yang berbeda dengan jelas [29]. *Support Vector Machine* adalah teknik yang menggunakan *supervised learning* untuk mengklasifikasikan data [30]. Pada penelitian ini, kernel *linear* dipilih karena memiliki beberapa keunggulan, yaitu akurasi yang tinggi, mencegah *overfitting*, waktu pelatihan yang lebih cepat, dan efisien untuk digunakan pada data yang besar [31]. Untuk menggunakan SVM, teks diubah menjadi data vektor. Proses klasifikasi sentimen menggunakan SVM dilakukan dengan memisahkan sentimen positif dan sentimen negatif menggunakan *hyperplane*. Secara umum, garis pemisah yang memiliki jarak terbesar ke titik data pelatihan terdekat dari setiap kelas akan memiliki akurasi yang lebih tinggi daripada garis pemisah lainnya. Margin adalah jarak antara titik data dalam kelas tertentu dengan *hyperplane* [32].

F. Evaluasi

Tahap evaluasi dilakukan untuk mengukur kinerja dari model *Support Vector Machine*. Kinerja ini diukur dengan menggunakan empat metrik, yaitu akurasi, presisi, *recall*, dan *f-measure*. Rumus untuk menghitung setiap metrik tersebut adalah sebagai berikut: [33]:

- 1) Akurasi adalah persentase dari semua data yang diprediksi dengan benar. Rumusnya sebagai berikut:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- 2) *Recall* merupakan ukuran seberapa banyak data kelas positif yang diprediksi dengan benar. Rumusnya sebagai berikut:

$$recall = \frac{TP}{TP + FN}$$

- 3) Presisi ukuran seberapa akurat model memprediksi data kelas positif. Rumusnya sebagai berikut:

$$precision = \frac{TP}{TP + FP}$$

- 4) *F-measure* adalah sebuah nilai yang dihasilkan dengan mengambil rata-rata dari *precision* dan *recall*. Rumusnya sebagai berikut:

$$f - measure = \frac{2x \text{ precision } x \text{ recall}}{\text{precision} + \text{recall}}$$

III. HASIL DAN PEMBAHASAN

A. Pembahasan

Penelitian ini tentang ulasan aplikasi *Threads* menggunakan algoritma *Support Vector Machine* dengan kernel *linear* dan pembagian data 80:20 dengan menghasilkan akurasi sebesar 88%. Akurasi ini diperoleh dengan mengklasifikasikan data ulasan menjadi dua kelas, yaitu ulasan positif dan ulasan negatif. Hasil analisis menunjukkan bahwa aplikasi *Threads* mendapat respons positif dari mayoritas pengguna di *Google Play Store*. Hal ini terlihat dari jumlah ulasan positif yang lebih banyak, yaitu sebanyak 1103, dibandingkan ulasan negatif yang hanya berjumlah 326. Penelitian tentang analisis sentimen menggunakan algoritma *Support Vector Machine* menunjukkan bahwa algoritma ini dipilih karena dapat memisahkan data input yang bersifat *non-linier* dan memiliki dimensi yang besar secara efektif, dengan memanfaatkan fungsi kernel [12]. Pada penelitian sebelumnya [34] membandingkan antara algoritma SVM dan *k-Nearest Neighbor*, algoritma SVM menunjukkan akurasi rata-rata sebesar 88,76% dan pada akurasi k-NN dengan iterasi K=2 sebesar 86,18%, dan pada iterasi K=6 sebesar 88,10%. Namun, kedua akurasi tersebut masih lebih rendah dari akurasi SVM sebesar 88,76%. Pada penelitian yang lain [35] menunjukkan

bahwa *accuracy* dari algoritma SVM sebesar 85% sedangkan NBC sebesar 82%. Berdasarkan hasil yang diperoleh, metode SVM menunjukkan akurasi yang lebih tinggi daripada metode algoritma lainnya, seperti NBC dan k-NN.

B. Pengumpulan Data

Untuk dapat mengambil data, peneliti terlebih dahulu memasukkan library yang dibutuhkan. Setelah data berhasil diambil, maka data tersebut akan dijadikan dataset. Data yang diperoleh merupakan data ulasan tentang aplikasi *Threads* di *Google Play Store*. Proses mengumpulkan data menggunakan *Google Colaboratory* dan pemrograman *Python*. Library *google-play-scraper* digunakan untuk mengambil data dari *Google Play Store*. Hasil pengumpulan data yang di ambil dari bulan Juli sebanyak 2000 data dari tanggal 5 Juli sampai 16 Juli 2023 dan data disimpan dalam bentuk CSV agar memudahkan dalam memakai cara analisis sentimen.

TABEL I
 HASIL PENGUMPULAN DATA

Ulasan
Tolong diperketat untuk konten sensitif!! Tolong blokir saja akun yang menampilkan unsur pornografi!!
Orang pertama yang memberi ulasan Threads di Android. Terima kasih Instagram dan Meta sudah memberikan alternatif pilihan lain yang lebih bagus daripada Twitter yang limit limit itu

C. Preprocessing

1) *Cleaning* dan *Case Folding*

Tahap *cleaning* merupakan langkah untuk menghilangkan tanda baca, karakter khusus, *hashtag*, menghapus data duplikat, dan tanda baca lainnya. Pada langkah *case folding*, semua huruf dalam teks akan diubah menjadi huruf kecil.

TABEL II
 CLEANING DAN CASE FOLDING

Sebelum	Sesudah
Tolong diperketat untuk konten sensitif!! Tolong blokir saja akun yang menampilkan unsur pornografi!!	tolong diperketat untuk konten sensitif tolong blokir saja akun yang menampilkan unsur pornografi
Orang pertama yang memberi ulasan Threads di Android. Terima kasih Instagram dan Meta sudah memberikan alternatif pilihan lain yang lebih bagus daripada Twitter yang limit limit itu	orang pertama yang memberi ulasan threads android terima kasih instagram dan meta sudah memberikan alternatif pilihan lain yang lebih bagus daripada twitter yang limit limit itu

2) *Tokenization*

Pada tahap ini yaitu *tokenization*, kalimat akan dipecah menjadi beberapa bagian dengan menghapus teks yang bukan merupakan huruf sehingga mempermudah proses *stemming* di tahap selanjutnya. Tabel 3 menunjukkan Hasil dari proses *tokenization*.

TABEL III
 TOKENIZATION

Sebelum	Sesudah
tolong diperketat untuk konten sensitif tolong blokir saja akun yang menampilkan unsur pornografi	['tolong', 'diperketat', 'untuk', 'konten', 'sensitif', 'tolong', 'blokir', 'saja', 'akun', 'yang', 'menampilkan', 'unsur', 'pornografi']
orang pertama yang memberi ulasan threads android terima kasih instagram dan meta sudah memberikan alternatif pilihan lain yang lebih bagus daripada twitter yang limit limit itu	['orang', 'pertama', 'yang', 'memberi', 'ulasan', 'threads', 'android', 'terima', 'kasih', 'instagram', 'dan', 'meta', 'sudah', 'memberikan', 'alternatif', 'pilihan', 'lain', 'yang', 'lebih', 'bagus', 'daripada', 'twitter', 'yang', 'limit', 'limit', 'itu']

3) Normalization

Proses *normalization* bertujuan untuk memastikan bahwa kata-kata yang telah diolah memiliki makna yang sesuai dengan KBBI.

TABEL IV
 NORMALIZATION

Sebelum	Sesudah
['tolong', 'diperketat', 'untuk', 'konten', 'sensitif', 'tolong', 'blokir', 'saja', 'akun', 'yang', 'menampilkan', 'unsur', 'pornografi']	['tolong', 'diperketat', 'untuk', 'konten', 'sensitif', 'tolong', 'blokir', 'saja', 'akun', 'yang', 'menampilkan', 'unsur', 'pornografi']
['orang', 'pertama', 'yang', 'memberi', 'ulasan', 'threads', 'android', 'terima', 'kasih', 'instagram', 'dan', 'meta', 'sudah', 'memberikan', 'alternatif', 'pilihan', 'lain', 'yang', 'lebih', 'bagus', 'daripada', 'twitter', 'yang', 'limit', 'limit', 'itu']	['orang', 'pertama', 'yang', 'memberi', 'ulasan', 'threads', 'android', 'terima', 'kasih', 'instagram', 'dan', 'meta', 'sudah', 'memberikan', 'alternatif', 'pilihan', 'lain', 'yang', 'lebih', 'bagus', 'daripada', 'twitter', 'yang', 'limit', 'limit', 'itu']

4) Stopword Removal

Tahap selanjutnya adalah melakukan proses *stopwords removal*, di mana kata-kata yang dianggap tidak memiliki makna yang jelas akan dihapus. Hal ini dilakukan agar data yang dihasilkan hanya berupa kalimat inti, tidak termasuk kata-kata yang berfungsi untuk menghubungkan kalimat-kalimat.

TABEL V
 STOPWORD REMOVAL

Sebelum	Sesudah
['tolong', 'diperketat', 'untuk', 'konten', 'sensitif', 'tolong', 'blokir', 'saja', 'akun', 'yang', 'menampilkan', 'unsur', 'pornografi']	['tolong', 'diperketat', 'konten', 'sensitif', 'tolong', 'blokir', 'akun', 'menampilkan', 'unsur', 'pornografi']
['orang', 'pertama', 'yang', 'memberi', 'ulasan', 'threads', 'android', 'terima', 'kasih', 'instagram', 'dan', 'meta', 'sudah', 'memberikan', 'alternatif', 'pilihan', 'lain', 'yang', 'lebih', 'bagus', 'daripada', 'twitter', 'yang', 'limit', 'limit', 'itu']	['orang', 'ulasan', 'threads', 'android', 'terima', 'kasih', 'instagram', 'meta', 'alternatif', 'pilihan', 'bagus', 'twitter', 'limit', 'limit']

5) Stemming

Langkah terakhir dalam proses *preprocessing* adalah proses *stemming*, dimana kata-kata yang memiliki imbuhan akan diubah menjadi bentuk kata baku dengan menggunakan *library Python Sastrawi*, seperti yang diperlihatkan pada Tabel 6.

TABEL VI
STEMMING

Sebelum	Sesudah
['tolong', 'diperketat', 'konten', 'sensitif', 'tolong', 'blokir', 'akun', 'menampilkan', 'unsur', 'pornografi']	tolong ketat konten sensitif tolong blokir akun tampil unsur pornografi
['orang', 'ulasan', 'threads', 'android', 'terima', 'kasih', 'instagram', 'meta', 'alternatif', 'pilihan', 'bagus', 'twitter', 'limit', 'limit']	orang ulas threads android terima kasih instagram meta alternatif pilih bagus twitter limit limit

D. Pelabelan Dataset

1) Translate

Pada langkah ini, menggunakan *library googletrans* untuk menerjemahkan teks dari bahasa Indonesia ke bahasa Inggris. Hal ini dilakukan karena pada proses ini, pelabelan menggunakan modul *vader lexicon* yang hanya mendukung bahasa Inggris, sehingga teks dari bahasa Indonesia perlu diterjemahkan terlebih dahulu dan bisa dilihat seperti pada Tabel 7.

TABEL VII
PROSES TRANSLATE

Sebelum	Sesudah
tolong ketat konten sensitif tolong blokir akun tampil unsur pornografi	<i>please be strict about sensitive content. please block accounts that display pornographic elements</i>
orang ulas threads android terima kasih instagram meta alternatif pilih bagus twitter limit limit	<i>people review threads android thank you instagram meta alternative choose good twitter limit limit</i>

2) Klasifikasi Sentimen

Setelah teks diterjemahkan ke bahasa Inggris menggunakan *library googletrans*, langkah selanjutnya adalah melakukan klasifikasi sentimen pada teks dengan menggunakan modul *vader lexicon*. Setelah menggunakan modul tersebut, selanjutnya akan melakukan pembobotan setiap kata agar bisa mendapatkan *score* dari setiap kata berdasarkan *score* kata-kata dalam *lexicon*. Setelah mendapatkan *score* dari tiap kata, akan dilakukan penjumlahan *score* untuk menentukan sentimen keluruhan teks. Jika jumlah *score* lebih besar dari 0, maka teks atau kalimat tersebut positif, jika jumlah *score* kurang dari 0, maka nilai tersebut negatif, hal ini dapat dilihat pada Tabel 8.

TABEL VIII
PROSES KLASIFIKASI SENTIMEN

Sebelum	Score	Compound
<i>please be strict about sensitive content. please only block accounts that display pornographic elements</i>	{'neg': 0.166, 'neu': 0.571, 'pos': 0.263, 'compound': 0.1779}	0.1779
<i>people review threads android thank you instagram meta alternative choose good twitter limit limit</i>	{'neg': 0.0, 'neu': 0.69, 'pos': 0.31, 'compound': 0.6597}	0.6597

Tahap selanjutnya setelah mendapatkan nilai *compound* adalah klasifikasi sentimen. Klasifikasi sentimen ini dilakukan dengan menggunakan nilai *compound* tersebut untuk menentukan apakah sentimen dari teks tersebut positif atau negatif, Gambar 2 menunjukkan proses tersebut.

```
def condition(c):
    if c>=0.0000:
        return "positif"
    else:
        return 'negatif'

translate['sentimen'] = translate['compound'].apply(condition)

translate.head()
```

Gambar 2 Proses Pelabelan

Dapat dilihat pada tabel 9 merupakan beberapa contoh kalimat dengan sentimen positif dan negatif.

TABEL IX
 CONTOH KALIMAT SENTTIMEN POSITIF DAN SENTTIMEN NEGATIF

Positif	Negatif
tolong ketat konten sensitif tolong blokir akun tampil unsur pornografi	scrolling seret seret bug
bagus tingkat upload instagram langsung masuk threads upload kali instagram threads simpel praktis story instagram	spesifikasi rendah layak tampil hancur

E. Implementasi Algoritma Support Vector Machine

Dalam tahap implementasi, langkah awal yang diambil adalah membagi data dengan menggunakan perbandingan 80:20 dari keseluruhan dataset. Dalam perbandingan ini, akan dibagi menjadi dua bagian, yaitu 80% data training dan 20% data testing. Gambar 3 menunjukkan proses pembagian data.

```
[ ] from sklearn.model_selection import train_test_split
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=386)
```

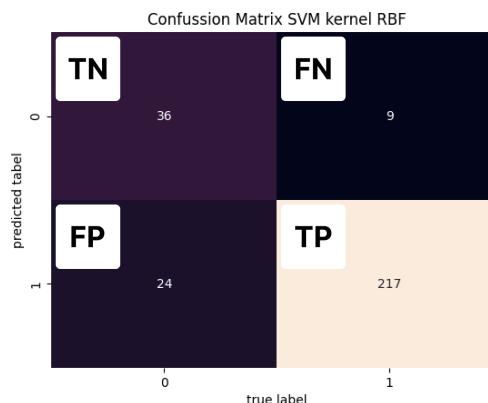
Gambar 3 Proses pembagian data

Setelah membagi data, proses selanjutnya adalah mengimplementasikan algoritma pada dataset. Dalam tahap implementasi, peneliti mengaplikasikan algoritma *Support Vector Machine* dengan kernel *linear*. Melalui pengujian dengan algoritma ini, harapannya adalah mampu menghasilkan output berupa tingkat akurasi, seperti yang ditunjukkan pada Gambar 4.

```
svm = SVC(kernel='linear')
svm.fit(X_train_tfidf, y_train)
y_pred = svm.predict(X_test_tfidf)
```

Gambar 4 Implementasi Algoritma

Setelah proses klasifikasi selesai dilakukan, tahap selanjutnya setelah proses klasifikasi adalah evaluasi performa algoritma. *Confusion matrix* adalah salah satu metode evaluasi performa algoritma klasifikasi. Pada gambar 5 merupakan hasil *confusion matrix*.



Gambar 5 Visualisasi Confussion Matrix

Keterangan:

TN : Hasil dari 286 data yang di *testing*, ditemukan 36 data yang menjadi *True Negative* (TN)

FN : Hasil dari 286 data yang di *testing*, ditemukan 9 data yang menjadi *False Negative* (FN)

TP : Hasil dari 286 data yang di *testing*, ditemukan 217 data yang menjadi *True Positive* (TP)

FP : Hasil dari 286 data yang di *testing*, ditemukan 24 data yang menjadi *False Positive* (FP)

F. Pengujian Evaluasi

Setelah melakukan proses validasi menggunakan algoritma *Support Vector Machine*, langkah selanjutnya merupakan proses terakhir dalam penelitian ini yaitu melakukan pengujian evaluasi. Kernel *linear* memiliki akurasi yang tinggi, sehingga digunakan dalam evaluasi ini. Hasil evaluasi menunjukkan akurasi sebesar 88%. Pada sentimen negatif menghasilkan *precision* mencapai 80%, *recall* sebesar 60%, dan untuk *f1-score* sebesar 69% dan untuk sentimen positif menghasilkan *precision* mencapai 90%, *recall* sebesar 96%, dan *f1-score* sebesar 93%. Pada Gambar 6 dan Gambar 7 menunjukkan hasil evaluasi ini.

```
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi:", accuracy)

Akurasi: 0.8846153846153846
```

Gambar 6 Hasil Akurasi Algoritma

	precision	recall	f1-score	support
-1	0.80	0.60	0.69	60
1	0.90	0.96	0.93	226
accuracy			0.88	286
macro avg	0.85	0.78	0.81	286
weighted avg	0.88	0.88	0.88	286

Gambar 7 Hasil Pengujian Evaluasi

Hasil dari semua tahap menunjukkan bahwa aplikasi *Threads* memiliki sentimen yang positif. Sentimen positif ini akan menjadi daya tarik bagi pengguna yang tertarik untuk mengunduh atau menggunakan aplikasi tersebut. Akurasi dalam pemilihan kernel ternyata bisa membuat akurasi menjadi tinggi. Pemilihan kernel yang tepat dapat membantu meningkatkan akurasi. Rekomendasi untuk penggunaan SVM dalam analisis sentimen dapat dijadikan pedoman untuk penelitian selanjutnya atau penerapan pada aplikasi serupa.

IV. KESIMPULAN

Berdasarkan pada hasil yang diperoleh terhadap ulasan sentimen aplikasi *Threads* sebanyak 2000 ulasan, dari 1429 ulasan yang telah melalui penghapusan ulasan *double*, sebanyak 1103 ulasan mendapatkan sentimen positif dan sebanyak 326 ulasan mendapatkan sentimen negatif. Hasil dari proses pengujian *review* yang diimplementasikan menggunakan metode algoritma *Support Vector Machine* menghasilkan hasil *accuracy* sebesar 88%. Untuk proses pengujian sentimen negatif, memperoleh *precision* sebesar 80%, untuk *recall* sebesar 60%, dan untuk *f1-score* sebesar 69%. Sedangkan sentimen positif memperoleh *precision* 90%, untuk *recall* 96%, dan untuk *f1-score* sebesar 93%. Dari hasil analisis sentimen terlihat bahwa rata-rata sentimen yang diberikan mempunyai arti positif sehingga membuat pengguna aplikasi *Threads* sangat puas dengan fitur yang diberikan. Selanjutnya saran untuk melakukan penelitian ini adalah membandingkan klasifikasi dengan kernel lain seperti *rbf*, *sigmoid* dan *polinomial* atau bisa juga menggunakan algoritma lain seperti *Naive Bayes* atau *k-Nearest Neighbour* dan menggunakan metode *k-fold cross* untuk membandingkan hasil akurasi setiap percobaan.

DAFTAR PUSTAKA

- [1] Inez, "Apa Itu Threads? Aplikasi Baru yang Digadang-gadang Jadi Pengganti Twitter, Trending di Media Sosial," *tribunnews.com*. Accessed: Sep. 09, 2023. [Online]. Available: <https://jateng.tribunnews.com/2023/07/07/apa-itu-threads-aplikasi-baru-yang-digadang-gadang-jadi-pengganti-twitter-trending-di-media-sosial>
- [2] A. Z. Yonatan, "Indonesia Peringkat 4, Ini Dia 7 Negara Pengguna Internet Terbesar di Dunia," *data.goodstats.id*. Accessed: Nov. 14, 2023. [Online]. Available: <https://data.goodstats.id/statistic/agneszeffanyayonatan/indonesia-peringkat-4-ini-dia-7-negara-pengguna-internet-terbesar-di-dunia-FLw6V>
- [3] M. Syaharani, "10 Negara Dengan Pengguna Smartphone Terbanyak Di Dunia, Indonesia Masuk Daftar!," *goodstats.id*. Accessed: Nov. 14, 2023. [Online]. Available: <https://goodstats.id/article/10-negara-dengan-pengguna-smartphone-terbanyak-di-dunia-indonesia-masuk-daftar-fDv25>
- [4] R. B. Lubis, "Membandingkan Tingkat Kecepatan Internet Kota-Kota di Kawasan Asia Tenggara," *goodstats.id*. Accessed: Nov. 14, 2023. [Online]. Available: <https://goodstats.id/article/membandingkan-tingkat-kecepatan-internet-kota-kota-di-kawasan-asia-tenggara-tlIu7>
- [5] D. V. Putsanra, "Apa Itu Threads di Instagram dan Bagaimana Menggunakannya?," *tirto.id*. Accessed: Sep. 09, 2023. [Online]. Available: <https://tirto.id/apa-itu-threads-di-instagram-dan-bagaimana-menggunakannya-gMGs>
- [6] Samita Sadya, "APIII: Pengguna Internet Indonesia 215,63 Juta pada 2022-2023," *dataindonesia.id*. Accessed: Sep. 09, 2023. [Online]. Available: <https://dataindonesia.id/digital/detail/apiii-pengguna-internet-indonesia-21563-juta-pada-20222023>

- [7] A. R. Satria, S. Adinugroho, and S. Suprpto, "Analisis Sentimen Ulasan Aplikasi Mobile menggunakan Algoritma Gabungan Naive Bayes dan C4.5 berbasis Normalisasi Kata Levenshtein Distance," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 11, pp. 4154–4163, 2020, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/8270>
- [8] S. Fransiska and A. Irham Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Sci. J. Informatics*, vol. 7, no. 2, pp. 2407–7658, 2020, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>
- [9] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [10] G. Radiena and A. Nugroho, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine," *J. Pendidik. Teknol. Inf.*, vol. 6, no. 1, pp. 1–10, 2023, doi: 10.37792/jukanti.v6i1.836.
- [11] A. C. Najib, A. Irsyad, G. A. Qandi, and N. A. Rakhmawati, "Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter," *Fountain Informatics J.*, vol. 4, no. 2, p. 41, 2019, doi: 10.21111/fij.v4i2.3573.
- [12] R. A. Saputra, D. Puspitasari, and T. Baidawi, "Deteksi Kematangan Buah Melon Dengan Algoritma Support Vector Machine Berbasis Ekstraksi Fitur GLCM," *J. Infortech*, vol. 4, no. 2, pp. 200–206, 2022, [Online]. Available: <https://ejournal.bsi.ac.id/ejournal/index.php/infortech/article/view/14436>
- [13] N. Pratiwi and Y. Setyawan, "Analisis Akurasi Dari Perbedaan Fungsi Kernel Dan Cost Pada Support Vector Machine Studi Kasus Klasifikasi Curah Hujan Di Jakarta," *J. Fundam. Math. Appl.*, vol. 4, no. 2, pp. 203–212, 2021, doi: 10.14710/jfma.v4i2.11691.
- [14] M. Ichwan, I. A. Dewi, and Z. M. S., "Klasifikasi Support Vector Machine (SVM) Untuk Menentukan TingkatKemanisan Mangga Berdasarkan Fitur Warna," *MIND J.*, vol. 3, no. 2, pp. 16–23, 2019, doi: 10.26760/mindjournal.v3i2.16-23.
- [15] M. F. Naufal, T. Arifin, and H. Wirjawan, "Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naive Bayes untuk Klasifikasi Cyberbullying pada Media Sosial," *Jurasik (Jurnal Ris. Sist.)*, vol. 8, pp. 82–90, 2023, [Online]. Available: <http://tunasbangsa.ac.id/ejournal/index.php/jurasik/article/view/544%0Ahttp://tunasbangsa.ac.id/ejournal/index.php/jurasik/article/download/544/522>
- [16] S. Rita *et al.*, "PENGGUNAAN SUPPORT VECTOR MACHINE UNTUK ANALISIS USING SUPPORT VECTOR MACHINE FOR SENTIMENT ANALYSIS OF," vol. 20, no. 2, pp. 131–140, 2023.
- [17] D. Safrida Putri and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support Vector Machine," *J. Ilm. Inform.*, vol. 11, no. 01, pp. 32–40, 2023, doi: 10.33884/jif.v11i01.6611.
- [18] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [19] A. Rahman, E. Utami, and S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naive Bayes dan Algoritma Genetika," *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 60–71, 2021, doi: 10.31603/komtika.v5i1.5188.
- [20] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [21] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naive Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [22] C. Villavicencio, J. J. Macrohon, X. A. Imbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naive bayes," *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.
- [23] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naive bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/teknol.v10i2.419.
- [24] R. Ramadhan, M. Afdal, I. Permana, and M. Jazman, "Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 3, pp. 715–724, 2023, doi: 10.30865/jurikom.v10i3.6396.
- [25] T. Malik Iryana and P. Pandu Adikara, "Analisis Sentimen Masyarakat Terhadap Mass Rapid Transit Jakarta Menggunakan Metode Naive Bayes Dengan Normalisasi Kata," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2548–964, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [26] R. Maulana, A. Voutama, and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC," *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023, doi: 10.54914/jtt.v9i1.609.
- [27] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *J. Sains dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022, doi: 10.34128/jsi.v8i1.419.
- [28] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, "Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naive Bayes Classifier," *J. Asimetrik J. Ilm. Rekayasa Inov.*, vol. 4, pp. 239–248, 2022, doi: 10.35814/asiimetrik.v4i1.3577.
- [29] O. Somantri, S. Wiyono, and D. Dairoh, "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," *Sci. J. Informatics*, vol. 3, no. 1, pp. 34–45, 2016, doi: 10.15294/sji.v3i1.5845.
- [30] R. W. Pratiwi, S. F. H. D. Dairoh, D. I. Af'idah, Q. R. A., and A. G. F., "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 1, pp. 40–46, 2021, doi: 10.20895/inista.v4i1.387.
- [31] A. Z. Praghakusma and N. Charibaldi, "Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi)," *JSTIE (Jurnal Sarj. Tek. Inform.)*, vol. 9, no. 2, p. 88, 2021, doi: 10.12928/jstie.v9i2.20181.
- [32] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [33] D. Ardiansyah, A. Saepudin, R. Aryanti, E. Fitriani, and Royadi, "Analisis Sentimen Review Pada Aplikasi Media Sosial Tiktok Menggunakan Algoritma K-Nn Dan Svm Berbasis Pso," *J. Inform. Kaputama*, vol. 7, no. 2, pp. 233–241, 2023, doi: 10.59697/jik.v7i2.148.
- [34] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [35] P. R. Saputra Andri, Subing Mulia, "Perbandingan Metode Naive Bayes Classifier Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022," *Teknomatika*, vol. 13, no. 01, pp. 22–31, 2023.