

Implementasi *Support Vector Machine* dalam Analisis Sentimen Ulasan Aplikasi IndiHome TV di *Google Play Store*

*Agus Fajar Riany¹, Fenny Purwani², Irfan Dwi Jaya³

^{1,2,3}Sistem Informasi, UIN Raden Fatah Palembang

Jl. Panca Usaha No.2085, 5 Ulu, Kecamatan Seberang Ulu I, Kota Palembang

Email: ¹1930803043@radenfatah.ac.id, ²fennypurwani_uin@radenfatah.ac.id, ³irfan_dj@radenfatah.ac.id

ABSTRACT

Sentiment analysis is used to determine the responses or opinions of a group or individual regarding a topic of discussion in the context of the entire document. The Indihome TV application is currently widely used by the public, so that reviews of the Indihome TV application on the Google Play Store are very numerous. The exact number of reviews given by users is not yet known based on their sentiment class. Therefore, a method is needed to facilitate the analysis of these user reviews. The purpose of this study is to determine the polarity of sentiment towards the Indihome TV application and to determine the performance and accuracy resulting from the application of the Support Vector Machine algorithm. The method used to convert unstructured reviews into structured reviews uses the Text Mining method. The results of this study indicate that using the SVM algorithm in sentiment analysis of the Indihome TV application data produces the highest accuracy value at a ratio of 90:10 at 94%. Furthermore, from the results of data visualization, the most frequently appearing words are applications, watch, channel, open, please, good, login, indihome, complete and so on.

Keywords: *sentiment analysis; text mining; support vector machine; indihome tv*

ABSTRAK

Analisis sentimen digunakan untuk mengetahui tanggapan atau pendapat dari suatu kelompok atau individu terhadap suatu topik bahasan kontekstual keseluruhan dokumen. Aplikasi Indihome TV saat ini banyak digunakan oleh masyarakat, dengan begitu ulasan terhadap aplikasi indihome TV di *Google Play Store* menjadi sangat banyak. Ulasan yang diberikan pengguna belum diketahui secara pasti jumlah ulasan berdasarkan kelas sentimennya. Maka dari itu, diperlukan sebuah cara untuk mempermudah dalam menganalisis ulasan-ulasan pengguna tersebut. Tujuan dari penelitian ini yaitu untuk mengetahui polaritas sentimen terhadap aplikasi Indihome TV serta untuk mengetahui kinerja dan akurasi yang dihasilkan dari penerapan algoritma *Support Vector Machine*. Metode yang digunakan untuk mengubah ulasan yang tidak terstruktur menjadi ulasan yang terstruktur menggunakan metode *Text Mining*. Hasil dari penelitian ini menunjukkan bahwa dengan menggunakan algoritma SVM dalam analisis sentimen data aplikasi Indihome TV menghasilkan nilai akurasi tertinggi pada perbandingan 90:10 sebesar 94%. Selanjutnya, dari hasil visualisasi data, kata-kata yang paling sering muncul seperti aplikasi, tonton, channel, buka, tolong, bagus, login, indihome, lengkap dan sebagainya.

Kata kunci: *analisis sentimen; text mining; support vector machine; indihome tv*

1. PENDAHULUAN

Pada saat ini perkembangan teknologi dan internet telah menjadi pamor, tolak ukur kemajuan suatu negara (Ngafifi, 2014). Inilah sebabnya banyak negara, termasuk Indonesia, sangat tertarik dengan perkembangan teknologi dan internet. Di Indonesia perkembangan teknologi dan internet mengalami kemajuan yang cukup pesat, terbukti dengan berkembangnya teknologi ponsel pintar, berkembangnya jaringan 5G, serta penggunaan internet semakin mudah dan meluas. Kemudahan penggunaan internet telah mengubah kebiasaan masyarakat pada umumnya, dari mencari informasi menggunakan media cetak sekarang telah menggunakan internet.

Jumlah pengguna internet yang tinggi di Indonesia menghadirkan potensi yang sangat baik bagi Penyedia Jasa Internet (ISP). Berbagai jenis ISP berusaha memberikan layanan terbaik kepada semua orang dengan menghasilkan layanan internet berkualitas tinggi sesuai dengan kebutuhan semua orang. Selain swasta, negara juga ikut bersaing di sektor telekomunikasi melalui PT Telkom Indonesia (Telkom, 2021).

Indihome merupakan layanan serat optik digital yang menawarkan tiga layanan; televisi interaktif, telepon seluler, dan internet berkecepatan tinggi (Telkom, 2021). Aplikasi Indihome TV merupakan layanan televisi interaktif yang memberikan pelayanan terbaik kepada para pengguna untuk dapat menonton televisi secara mudah kapanpun dan dimana saja (Telkom, 2021).

Berdasarkan hasil observasi yang dilakukan Ulasan terkait aplikasi Indihome TV yang disampaikan pengguna, belum terklasifikasi secara pasti apakah ulasan tersebut benar termasuk ulasan positif atau ulasan negatif dan juga belum diketahui jumlah ulasan yang termasuk kelas positif dan kelas negatif. Oleh karena itu, ulasan pengguna terhadap aplikasi Indihome TV perlu dilakukan analisis.

Analisis sentimen mengidentifikasi penilaian berupa keluhan atau persepsi pengguna terhadap sebuah produk atau layanan. Perusahaan dapat menggunakan analisis sentimen untuk memahami tanggapan dan pendapat kelompok atau orang terhadap isu-isu yang disajikan dalam berbagai teks (Alwasi'a, 2020). Informasi yang diperoleh dari

pendekatan ini dapat digunakan oleh dunia usaha untuk berinovasi dan membuat perubahan.

Berdasarkan hal di atas, sehingga penting untuk dilakukan kajian *review* pengguna terhadap aplikasi Indihome TV. Pada penelitian ini ulasan dikelompokkan menjadi dua yaitu berupa ulasan sentimen positif dan sentimen negatif. Analisis sentimen dapat dikelompokkan menjadi dua kelas seperti penelitian (Mahardika & Zuliarso, 2018) yang mengelompokkan sentimen menjadi sentimen positif dan sentimen negatif. Evaluasi *review* diolah dengan analisis sentimen menggunakan metode klasifikasi.

Metode *Lexicon Based* merupakan metode dengan pendekatan kamus dan bertujuan untuk mendapatkan bobot kalimat pada data set sehingga diketahui label kelas sentimen pada data set (Herdhianto, 2020). Maka dari itu, dalam penelitian ini, *Lexicon Based* digunakan untuk mengkategorikan ulasan dokumen atau data yang berupa kata dalam sebuah kalimat. Setelah data diberi label awal selanjutnya digunakanlah algoritma SVM.

Pada penelitian ini algoritma SVM digunakan untuk mengetahui

performa atau kinerja dari data yang dihasilkan dari proses *Lexicon Based*. Algoritma SVM dipilih karena dapat menentukan *hyperline* pembatas yang optimal (Ubaidillah et al., 2019), mempunyai akurasi klasifikasi tinggi, serta mempunyai proses pembelajaran yang paling cepat dan efisien (Budianto et al., 2019). Hal ini dapat dilihat jelas dari hasil penelitian (Herlinawati et al., 2020) yang mengungkapkan bahwa algoritma SVM memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma *Naïve Bayes*, dengan algoritma SVM memiliki nilai akurasi sebesar 81,2% dan algoritma NB memiliki akurasi sebesar 74,37%. Dengan demikian, algoritma SVM mengungguli algoritma NB sebesar 6,85% dalam klasifikasi *review* pengguna aplikasi *Zoom Meeting* di *Play Store*. Selain itu, penggunaan algoritma SVM untuk menilai ulasan pengguna terhadap *Google Meet* mencapai tingkat akurasi sebesar 94% (Fitri & Putri, 2020).

Setelah dilakukan pengujian akurasi model yang menggunakan algoritma SVM, selanjutnya akan dilakukan evaluasi model dengan menggunakan *Confusion Matrix*. *Confusion Matrix* digunakan untuk

melihat berapa banyak data dari kelas-kelas yang kemungkinan akan diklasifikasi benar dan berapa banyak data yang akan diklasifikasi salah.

Data masing-masing ulasan positif dan negatif tersebut selanjutnya akan dilakukan penggalian dan penemuan menggunakan *Wordcloud* dan kombinasi kata untuk mendapatkan kata-kata yang umum ditemukan di *review* pengguna Indihome TV sehingga kata-kata positif ataupun negatif yang paling sering diungkapkan oleh pengguna dalam memberikan komentar atau ulasan pendapat pada aplikasi Indihome TV dapat terlihat. Hasil penelitian ini diharapkan menghasilkan pengklasifikasian yang baik dan akurat sehingga dapat membantu peningkatan layanan aplikasi Indihome TV.

Berdasarkan uraian latar belakang permasalahan di atas, penelitian ini bertujuan untuk mendapatkan hasil polaritas sentimen pada aplikasi indihome TV serta mendapatkan hasil persentase performa dan akurasi yang diperoleh dari penerapan algoritma SVM dalam penentuan polaritas sentimen pada aplikasi indihome TV

2. METODE

Metode penelitian adalah metode yang digunakan untuk memperoleh data untuk menghasilkan suatu tujuan. Metode penelitian kuantitatif adalah metode penelitian yang menggunakan filsafat positivisme, untuk meneliti populasi atau sampel tertentu dengan pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif dengan tujuan untuk menguji hipotesis yang telah ditetapkan (Sugiyono, 2017).

Tujuan dari metode penelitian kuantitatif adalah untuk mengidentifikasi masalah yang ditetapkan sebagai tujuan penelitian, membuat prediksi untuk memutuskan masalah tersebut, mendapatkan data untuk diprediksi, dan menginterpretasi atau menganalisis sebuah data untuk melihat apakah mendukung untuk prediksi (Santoso & Madiistriyatno, 2021).

Metode kuantitatif yang dilakukan akan dijelaskan dengan pendekatan deskriptif. Pendekatan deskriptif merupakan suatu cara untuk mengetahui eksistensi dari variabel bebas (Arikunto, 2014). Dalam penelitian ini menerapkan metode

kuantitatif yang memanfaatkan *text mining* dan perhitungan algoritma *Support Vector Machine*.

2.1. Metode Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder mengacu pada informasi yang dikumpulkan atau diperoleh peneliti dari berbagai sumber yang tersedia. Sumber data pada penelitian ini memanfaatkan data ulasan pada aplikasi Indihome TV yang tersedia di *website Google Play*. Pada tahapan ini dilakukan pengumpulan data komentar atau ulasan Bahasa Indonesia. Data yang dikumpulkan adalah data komentar pengguna dari Google Play yang merupakan pesan atau pendapat pada aplikasi Indihome TV. Data yang diambil adalah data ulasan dari 26 Desember 2013 hingga 02 November 2024. Terdapat sekitar kurang lebih 54 ribu ulasan pengguna yang memberikan penilaian pada aplikasi Indihome TV. Namun, data yang berhasil dikumpulkan berdasarkan teknik ini berjumlah 17.462 data.

2.2. Text Preprocessing

Text preprocessing merupakan langkah pertama yang perlu dilakukan

pada sebuah dokumen, membantu menyiapkan teks yang belum tersusun sehingga menjadi data teks tersusun untuk diproses lebih lanjut. Berikut adalah langkah-langkah untuk *preprocessing text* yaitu (Sitanayah Que et al., 2020):

a. *Cleansing*

Proses membersihkan data menghilangkan tanda baca seperti; titik (.), koma (,), tanda tanya (?), dan tanda seru(!), lalu simbol atau karakter, *mention* atau nama pengguna, *emoticon*, *hashtag* (#), dan *link* atau URL.

b. *Spelling Normalization*

Koreksi ejaan merupakan langkah untuk memperbaiki kata-kata yang memiliki singkatan dan kesalahan ejaan. Jika koreksi kata tidak dilakukan, jumlah hitungan kata meningkat.

c. *Case Folding*

Case folding adalah tahapan yang dilakukan untuk mensejajarkan semua huruf, semua huruf diubah menjadi huruf kecil, dan semua tanda baca dan angka dihilangkan..

d. *Tokenizing*

Tokenizing adalah tahapan yang dilakukan dengan memisahkan teks yang awalnya berupa kalimat menjadi kata-kata.

e. *Filtering*

Filtering merupakan langkah mengekstraksi kata-kata penting dari hasil proses tokenisasi dan menghilangkan *stopwords*.

f. *Stemming*

Stemming merupakan langkah yang dilakukan untuk membuat perubahan pada kata berimbuhan menjadi bentuk kata dasar.

2.3. Algoritma *Support Vector Machine*

SVM adalah algoritma yang mengimplementasikan pemetaan non-linier untuk memperbesar ukuran data pelatihan asli (Handayanto & Herlawati, 2020). Algoritma SVM digunakan untuk menghitung masalah linier dengan mengimplementasikan transformasi matematis ke ruang pelatihan menggunakan fungsi kernel (Al-amrani et al., 2018). Tujuan dari algoritma SVM adalah untuk mengidentifikasi pemisah batas (*hyperlane*) dengan margin terbesar yang mungkin sehingga dua set data dapat dianalisis secara optimal (Fridayanti, 2023).

SVM adalah algoritma yang mengimplementasikan pemetaan non-linier untuk memperbesar ukuran data

pelatihan asli (Handayanto & Herlawati, 2020). Algoritma SVM digunakan untuk menghitung masalah linier dengan mengimplementasikan transformasi matematis ke ruang pelatihan menggunakan fungsi kernel (Al-amrani et al., 2018). Tujuan dari algoritma SVM adalah untuk mengidentifikasi pemisah batas (*hyperlane*) dengan margin terbesar yang mungkin sehingga dua set data dapat dianalisis secara optimal (Fridayanti, 2023).

SVM mengelompokkan data dengan membaginya menjadi dua kelas menggunakan *hyperlane*. *Hyperlane* terletak di antara dua ruang kelas, dengan jarak d pada setiap meja. Jarak d disebut sebagai margin, dan titik yang terletak pada margin disebut sebagai vektor pendukung (Ipmawati et al., 2024). Tujuan SVM adalah untuk mengidentifikasi *hyperlane* terbaik yang dapat memberikan informasi waktu nyata dari suatu titik tertentu.

Jika data dapat diproses secara akurat oleh *hyperlane* linier, SVM disebut sebagai SVM linier. Namun, jika data tidak dapat diproses secara linier, SVM menggunakan transformasi kernel untuk memindahkan data ke

dimensi yang lebih besar, di mana *hyperlane* linier dapat dibangun.

SVM dapat menangani data non-linier secara efisien dengan menggunakan beberapa jenis kernel, seperti kernel linier, polinomial, dan Gaussian (Ipmawati et al., 2024). Fungsi kernel adalah fungsi yang mengkonversi data ke dimensi yang lebih besar dengan tujuan memperbaiki struktur data dan mempermudah proses pemrosesan. Dalam hal ini, rumus umum untuk SVM linier dapat ditulis sebagai berikut (Ipmawati et al., 2024):

$$F(x) = \text{sign}(w \cdot x + b) \quad (1)$$

Dimana:

$F(x)$ = fungsi prediksi

w = vektor normal *hyperlane*

x = vektor fitur input

b = bias atau *intercept*

2.4. Confusion Matrix

Secara umum, mengukur keefektifan teknik klasifikasi dapat dilakukan dengan matriks konfusi (Utomo, 2020). Matriks Kesalahan adalah tabel dimana digunakan dalam evaluasi klasifikasi model untuk memvisualisasikan kinerja model (Bambang Seran & Supatman, 2024).

Confusion matrix diatas dapat digunakan untuk melihat berapa banyak data dari kelas-kelas yang kemungkinan akan diklasifikasi benar dan berapa banyak data yang akan diklasifikasi salah. Matriks konfusi memiliki dua nilai antara lain keakuratan dan tingkat kesalahan. Hasil klasifikasi dapat mengungkapkan seberapa banyak data yang telah dikategorikan secara akurat dan mengetahui kebenarannya. Sedangkan, untuk melihat tingkat kesalahan yang dilakukan dapat diketahui dengan melihat jumlah data yang terklasifikasi salah (Utomo, 2020).

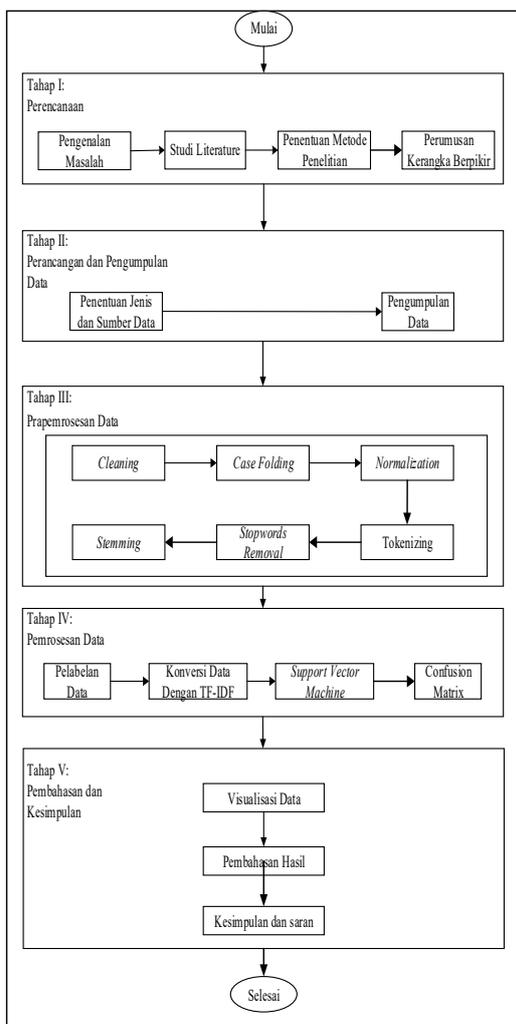
2.5. Visualisasi

Pada penelitian ini visualisasi data dilakukan untuk kata-kata yang sering muncul dalam data ditampilkan dalam bentuk *wordcloud* dan diagram batang. Tujuannya adalah untuk menentukan kata kunci mana yang paling sering muncul dalam data, serta seberapa sering kemunculannya.

2.6. Tahapan Penelitian

Tahapan penelitian adalah langkah-langkah yang akan dilakukan pada penelitian ini dan digambarkan dalam bentuk diagram alir. Diagram alir tersebut menggambarkan tentang

tahapan proses penelitian yang akan dilakukan sekaligus juga menggambarkan penelitian secara menyeluruh. Gambar 1 berikut ini merupakan tahapan penelitian.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Pengolahan data ulasan atau komentar pada aplikasi Indihome TV merupakan bagian utama sebagai hasil dalam penelitian ini yang selanjutnya

akan dibahas di pembahasan. Adapun pengolahan data yang dilakukan yaitu dengan pengumpulan data, *preprocessing text*, pengujian data menggunakan algoritma SVM sampai dengan proses visualisasi data.

a. Pengumpulan Data

Data pada penelitian ini diambil dari *play store* dengan *scraping data* pada aplikasi Indihome TV. Data komentar yang diperoleh adalah data bahasa Indonesia seluruh komentar pada aplikasi Indihome TV dan pada rentang waktu 26 Desember 2013 hingga 02 November 2024. Data yang akan diambil adalah *username*, tanggal dan waktu, *rating*, dan Komentar. Jumlah data yang diperoleh adalah sebanyak 17.462 data komentar.

Tabel 1. Contoh Data Hasil *Scraping* dari *Google Play*

Username	Rating	Tanggal/Waktu	Komentar
Pengguna Google	1	17/10/2024 07:52:12	👎 pas lagi nonton tiba2 loading padahal sinyal bagus 🤦
Pengguna Google	5	13/10/2024 13:06:51	Bagus, tingkatkan mutu siaran jangan ada siaran ngeleg
Pengguna Google	1	13/10/2024 10:47:59	ap nya g banget ngelag kl nonton, kebanyakan ngelag nya 🤦 🤦
Pengguna	1	09/10/20	apakah sudah

na	24	tidak support
Google	20:17:56	android go?

apl nya g banget ngelag kl nonton, kebanyakan ngelag nya 🙄 🙄	apl nya g banget ngelag kl nonton kebanyakan ngelag nya
---	--

b. Text Preprocessing

Data *raw* atau data mentah yang diperoleh dari proses *scraping* data yang berjumlah 17.462 data perlu dibersihkan terlebih dahulu sebelum diolah mesin. Tahapan ini disebut dengan *text preprocessing*. Langkah-langkah *preprocessing* yang dilakukan sebagai berikut:

1. Cleaning

Tahapan *cleaning* merupakan tahapan membersihkan data komentar yang telah diperoleh. Menurut (Ferdiana et al., 2019) terdapat beberapa elemen pengganggu yang dibersihkan, antara lain: Simbol –simbol, seperti *emoticon*, *hashtag*, *mention*, tanda baca, seperti titik, koma, tanda kurung, tanda tanya dan tanda seru, serta angka, data *duplicate* dan data spam.

Tabel 2. Contoh Data Komentar pada Proses *Cleaning*

Komentar	Hasil <i>Cleaning</i>
🙄 pas lagi nonton tiba2 loading padahal sinyal bagus 🙄	pas lagi nonton tiba loading padahal sinyal bagus
Bagus, tingkatkan mutu siaran jangan ada siaran ngeleg	Bagus tingkatkan mutu siaran jangan ada siaran ngeleg

2. Case Folding

Dalam tahap *preprocessing*, ada bagian *case folding* yang digunakan untuk mengubah bentuk huruf besar menjadi bentuk huruf kecil atau menyamakan semua bentuk huruf agar data teks yang diproses tetap sama (Albab et al., 2023).

Tabel 3. Contoh Data Komentar pada Proses *Case Folding*

Hasil <i>Cleaning</i>	Hasil <i>Case Folding</i>
pas lagi nonton tiba loading padahal sinyal bagus Bagus tingkatkan mutu siaran jangan ada siaran ngeleg apl nya g banget ngelag kl nonton kebanyakan ngelag nya	pas lagi nonton tiba loading padahal sinyal bagus bagus tingkatkan mutu siaran jangan ada siaran ngeleg apl nya g banget ngelag kl nonton kebanyakan ngelag nya

3. Normalization

Pada tahap ini dilakukan normalisasi ejaan yang bertujuan untuk memperbaiki kata-kata yang tidak baku, kata singkatan atau *slang word* menjadi kata baku sesuai dengan KKBI.

Tabel 4. Contoh Data Komentar pada Proses Normalisasi

Hasil <i>Case Folding</i>	Hasil Normalisasi
pas lagi nonton tiba loading padahal sinyal bagus bagus tingkatkan mutu siaran jangan ada siaran ngeleg	pas lagi menonton tiba loading padahal sinyal bagus bagus tingkatkan mutu siaran jangan ada siaran ngeleg

apl nya g banget	apl ya tidak banget
ngelag kl nonton	ngelag kalau
kebanyakan ngelag	menonton kebanyakan
nya	ngelag ya

4. Tokenizing

Tujuan tahapan *tokenizing* adalah untuk memecah kalimat menjadi bagian-bagian kecil yang disebut token.

Tabel 5. Contoh Data Komentar pada Proses *Tokenizing*

Hasil Normalisasi	Hasil <i>Tokenizing</i>
pas lagi menonton tiba loading padahal sinyal bagus	['pas', 'lagi', 'menonton', 'tiba', 'loading', 'padahal', 'sinyal', 'bagus']
bagus tingkatkan mutu siaran jangan ada siaran ngeleg	['bagus', 'tingkatkan', 'mutu', 'siaran', 'jangan', 'ada', 'siaran', 'ngeleg']
apl ya tidak banget ngelag kalau menonton	['apl', 'ya', 'tidak', 'banget', 'ngelag', 'kalau', 'menonton',
kebanyakan ngelag ya	'kebanyakan', 'ngelag', 'ya']

5. Stopword Removal

Penghapusan *stopword* digunakan untuk menghilangkan kata-kata yang tidak relevan dengan kalimat atau tidak memiliki makna untuk komentar.

Tabel 6. Contoh Data Komentar pada Proses *Tokenizing*

Hasil <i>Tokenizing</i>	Hasil <i>Stopword Removal</i>
['pas', 'lagi', 'menonton', 'tiba', 'loading', 'padahal', 'sinyal', 'bagus']	['pas', 'menonton', 'loading', 'sinyal', 'bagus']
['bagus', 'tingkatkan', 'mutu', 'siaran', 'jangan', 'ada', 'siaran', 'ngeleg']	['bagus', 'tingkatkan', 'mutu', 'siaran', 'siaran', 'ngeleg']
['apl', 'ya', 'tidak', 'banget', 'ngelag', 'kalau', 'menonton', 'kebanyakan', 'ngelag', 'ya']	['apl', 'ya', 'banget', 'ngelag', 'menonton', 'kebanyakan', 'ngelag', 'ya']

6. Stemming

Pada tahap stemming, kata diubah menjadi bentuk dasarnya dengan menghilangkan imbuhan, yaitu awalan, sisipan, akhiran, serta kombinasi awalan dan akhiran.

Tabel 7. Contoh Data Komentar pada Proses *Stemming*

Hasil <i>Stopword Removal</i>	Hasil <i>Stemming</i>
['pas', 'menonton', 'loading', 'sinyal', 'bagus']	pas tonton loading sinyal bagus
['bagus', 'tingkatkan', 'mutu', 'siaran', 'siaran', 'ngeleg']	bagus tingkat mutu siar siar ngeleg
['apl', 'ya', 'banget', 'ngelag', 'menonton', 'kebanyakan', 'ngelag', 'ya']	apl ya banget ngelag tonton banyak ngelag ya

c. Data Set

Data set dari hasil proses *text preprocessing* yang akan digunakan untuk tahap klasifikasi berjumlah 9965 data, berdasarkan data mentah yang berjumlah 17.462 data ulasan. Data ulasan yang berjumlah 9965 data inilah yang akan dilakukan klasifikasi dan pengujian sehingga menghasilkan suatu informasi yang dapat bermanfaat untuk pihak-pihak terkait.

d. Pelabelan dan Pengujian Data

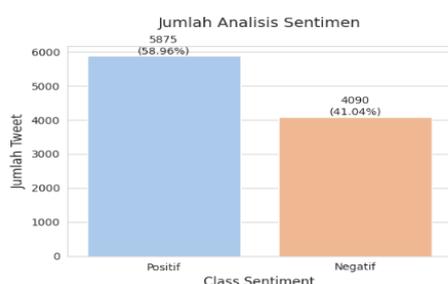
Pada tahap ini, data dilabelkan menggunakan metode *Lexicone Based*, yang merupakan kamus *lexicone* yang mengandung sentimen positif dan negatif. Setelah sentimen diberi label,

Kemudian dilakukan pengujian hasil klasifikasi menggunakan metode *Support Vector Machine*.

1. Pelabelan Data dengan *Lexicone Based*

Kelas sentimen dihasilkan berdasarkan penjumlahan hasil skor dari setiap kata secara keseluruhan dari sebuah kalimat atau data. Jumlah nilai sentimen di atas 0 (sentimen ≥ 0) dikatakan sebagai kelas sentimen positif, sedangkan jumlah sentimen di bawah 0 (sentimen < 0) dikatakan sebagai kelas sentimen negatif (Faradian et al., 2024).

Komentar kelas sentimen positif diperoleh sebanyak 5875 komentar. Dan untuk komentar kelas sentimen negatif diperoleh sebanyak 4090 komentar. Berikut ini grafik hasil analisis jumlah sentimen positif dan negatif dari aplikasi indihome TV.



Gambar 2. Hasil Jumlah Analisis Sentimen

2. Pembagian dan Konversi Data

Pembagian data uji dan latihan dilakukan untuk meningkatkan kualitas

data. Ini dilakukan dengan membagi data menjadi 80:20, 70:30, 60:40, 50:50, dan 40:60.

Setelah dilakukan pembagian data, selanjutnya sebelum data digunakan untuk proses pengujian data menggunakan Algoritma *Support Vector Machine*, terlebih dahulu dilakukan konversi data yaitu mengubah data teks menjadi numerik. Hal ini dilakukan karena proses pengolahan data menggunakan SVM menggunakan data numerik. Proses konversi ini menggunakan bantuan *library Tfidf vectorizer*.

```

➡ Shape of X_train_vec: (8968, 6805)
  Shape of X_test_vec: (997, 6805)

X_train_vec (as array):
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]

X_test_vec (as array):
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]

```

Gambar 3. Contoh Hasil Dari Konversi Data

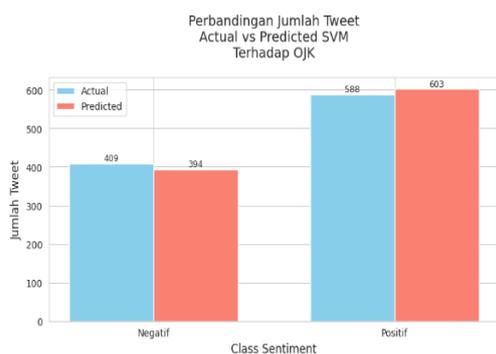
3. Pengujian Menggunakan Algoritma *Support Vector Machine*

Dalam proses pengujian model data dengan algoritma SVM ini, data yang digunakan sebanyak 9965 data ulasan telah diberi label dan diubah menjadi bentuk numerik.

Tabel 8. Akurasi SVM Berdasarkan Data Latih dan Data Uji

Perbandingan Data		Jumlah Data		Akurasi
Data Latih	Data Uji	Data Latih	Data Uji	
90%	10%	8.968	997	94%
80%	20%	7.972	1.993	93%
70%	30%	6.976	2.990	92%
60%	40%	5.979	3.986	91%
50%	50%	4.982	4.983	91%
40%	60%	3.986	5.979	91%

Berdasarkan hasil pengujian, nilai tertinggi diperoleh dari perbandingan 90:10 yaitu sebesar 94%.



Gambar 4. Grafik Hasil Pengujian Kelas Menggunakan SVM

Berdasarkan Hasil prediksi yang telah dilakukan dimana data testing berjumlah 997 data, yang merupakan data *testing* dari perbandingan 90:10. Dimana diperoleh kelas positif awal berjumlah 588, sedangkan berdasarkan hasil prediksi berjumlah 603. Kemudian kelas negatif data awal berjumlah 409, sedangkan berdasarkan hasil prediksi berjumlah 394. Berdasarkan hasil pengujian perbandingan data aktual dan

data hasil prediksi tidak terlalu mengalami perubahan yang signifikan.

e. Evaluasi (*Evaluation*)

Tahap evaluasi bertujuan untuk mengetahui performa akurasi dari hasil prediksi pengujian klasifikasi algoritma SVM yang telah dibangun. Dilakukan dengan menggunakan fungsi *confusion matrix* dimana perbandingan data yang digunakan pada pengujian ini yaitu 90:10, dimana 90% data *training* dan 10% data *testing*. Tabel 8 berikut merupakan hasil evaluasi yang dilakukan.

Classification Report:				
	precision	recall	f1-score	support
Negatif	0.94	0.91	0.93	409
Positif	0.94	0.96	0.95	588
accuracy			0.94	997
macro avg	0.94	0.94	0.94	997
weighted avg	0.94	0.94	0.94	997

Gambar 5. Hasil Evaluasi *Confusion Matrix*

Hasil evaluasi menunjukkan bahwa akurasi dari algoritma *Support Vector Machine* sebesar 94%. Artinya akurasi algoritma SVM dalam mengklasifikasi ulasan pengguna aplikasi Indihome TV pada kategori *Good Classification*.

f. Visualisasi

Visualisasi digunakan untuk mengumpulkan data berupa kata-kata yang paling sering disebutkan atau

muncul oleh pengguna aplikasi Indihome TV. Visualisasi dapat berguna untuk memperoleh informasi penting dari banyak teks komentar yang ada. Dalam penelitian ini, analisis sentimen akan divisualisasikan menggunakan *Wordcloud*. Berikut ini hasil visualisasi data pada ulasan secara keseluruhan, ulasan positif dan negatif.



Gambar 6. *Wordcloud* Keseluruhan Data Ulasan



Gambar 7. *Wordcloud* Kelas Positif



Gambar 8. *Wordcloud* Kelas Negatif

3.2. Pembahasan

Berdasarkan hasil yang telah di dapatkan, terdapat temuan dalam penelitian ini yang akan dibahas pada bagian pembahasan. Pembahasan dari hasil analisis yang telah dilakukan mulai dari pengumpulan data dengan teknik *scrapping* data *Google Play* pada aplikasi Indihome TV dengan rentang waktu 26 Desember 2013 – 02 November 2024. Diperoleh data sebanyak 17.462 data ulasan.

Dalam penelitian ini, tahapan penelitian dalam analisis ini yaitu *text preprocessing*, pelabelan data dengan *Lexicon*, klasifikasi data dengan *Support Vector Machine* (SVM), pengujian dengan *Confusion Matrix*, serta visualisasi.

Tahapan *text preprocessing* dilakukan untuk pembersihan data. dalam tahap ini terdapat beberapa proses, yaitu *cleaning*, *case folding*, *normalization*, *tokenization*, *stopword removal* dan *stemming*. Data bersih yang didapat dari proses ini adalah sebanyak 9965 ulasan.

Setelah data dibersihkan dan distrukturkan, selanjutnya proses pelabelan data. Dari proses ini didapatkan jumlah data ulasan

mengenai aplikasi Indihome TV yang termasuk kelas sentimen positif sebanyak 5875 ulasan, sedangkan jumlah data ulasan mengenai aplikasi Indihome TV yang termasuk kelas sentimen negatif sebanyak 4090 ulasan.

Sebelum dilakukan pengujian terhadap data yang telah diberi label menggunakan algoritma SVM, data harus dibagi menjadi data latih dan data uji. Pembagian data ini adalah 90:10, 80:20, 70:20, 60:40, 50:50, dan 40:60.

Berdasarkan hasil pengujian yang dilakukan menggunakan algoritma SVM yang dilakukan dengan perbandingan yang beragam, dihasilkan bahwa nilai akurasi tertinggi yaitu dihasilkan dari perbandingan 90:10 dengan nilai akurasi sebesar 94%. Dari hasil perbandingan ini menunjukkan bahwa SVM merupakan algoritma yang dapat digunakan dalam pengklasifikasian data ulasan pengguna aplikasi Indihome TV dilihat dari nilai akurasi yang sangat baik.

Berdasarkan hasil perbandingan 90:10 dengan data latih berjumlah 8.968 dan data uji 997. Dimana hasil data uji yang digunakan untuk memeriksa apakah data tersebut terklasifikasi sesuai dengan kelasnya. Dihasilkan bahwa dari data uji yang berjumlah 997

yaitu kelas positif awal berjumlah 588, sedangkan berdasarkan hasil prediksi berjumlah 603. Kemudian kelas negatif data awal berjumlah 409, sedangkan berdasarkan hasil prediksi berjumlah 394. Akhirnya, hasil pengujian perbandingan data aktual dan data hasil prediksi tidak terlalu mengalami perubahan yang signifikan.

Setelah dilakukan pengujian menggunakan algoritma SVM selanjutnya dilakukan evaluasi terhadap model yang telah dibuat menggunakan *Confusion Matrix*. *Confusion Matrix* digunakan untuk mengevaluasi hasil klasifikasi data. Berdasarkan hasil perbandingan 90:10, 90% data *training* dan 10% data *testing*. Data *testing* sebanyak 997 data diperoleh hasil yaitu terdapat 566 komentar positif pengguna Indihome TV yang benar masuk dalam komentar positif dan terdapat 37 komentar positif yang seharusnya masuk ke dalam kelas negatif. Selanjutnya, terdapat 372 data negatif yang benar masuk dalam komentar negatif serta 22 data komentar negatif yang seharusnya masuk ke dalam kelas positif. Akhirnya, dari hasil evaluasi model menggunakan *Confusion Matrix* diperoleh nilai akurasi sebesar 94%, *precision* negatif dan positif sebesar

94%, *recall* negatif sebesar 91% dan *recall* positif 96%, dan *F1-Score* negatif sebesar 93% dan *F1-Score* positif 95%.

Selanjutnya, *bar chart* dan *wordcloud* digunakan untuk divisualisasikan data hasil klasifikasi. *Wordcloud* mengumpulkan informasi tentang kata-kata populer seperti kata aplikasi, tonton, channel, buka, tolong, bagus, login, indihome, lengkap dan sebagainya merupakan kata yang paling banyak muncul pada ulasan positif maupun negatif. Penilaian positif dari pengguna diantaranya tentang aplikasi, channel, pakai, suka dan indihome yang dapat diartikan sebagai pengguna menyukai aplikasi aplikasi indihome TV serta terdapat juga kata bagus, mantap.

Sedangkan beberapa penilaian negatif dari pengguna terdapat pula diantaranya seperti kata kata muncul yang memiliki arti masuk terkadang pengguna kesulitan untuk masuk, kata aplikasi yang berarti aplikasi sering mengalami *error* atau gangguan, channel dapat diartikan saluran pada aplikasi kurang lengkap karena terdapat beberapa saluran yang tidak ada dan lain sebagainya. Dari beberapa penilaian negatif dapat dilakukan tindak lanjut dengan perbaikan aplikasi

Indihome TV serti memperbaiki kesalahan atau kegagalan saat pengguna akan login ke dalam aplikasi, kemudian saat pengguna akan melakukan pembaruan, dan sebagainya. Informasi tersebut dapat berguna bagi PT Telekomunikasi Indonesia selaku pengembang aplikasi Indihome TV maupun pihak lain yang membutuhkan.

4. KESIMPULAN

Berdasarkan hasil dari serangkaian tahapan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa proses analisis sentimen dengan teknik *text mining* menggunakan algoritma *Support Vector Machine* dapat diterapkan dalam menentukan polaritas sentimen untuk mengetahui ulasan pengguna aplikasi sentimen apakah termasuk kelas positif atau negatif. Dimana berdasarkan proses pengujian yang telah dilakukan dengan menggunakan data *training* dan data testing dengan pengujian dengan hasil tertinggi pada perbandingan 90:10, hasilnya menunjukkan bahwa pada data testing, data yang terklasifikasi masuk kedalam kelas positif sebanyak 588 data dan yang data terklasifikasi masuk kedalam kelas negatif sebanyak 409

data. Dari pengujian ini menghasilkan akurasi sebesar 94%.

Adapun saran bagi penelitian selanjutnya yaitu agar menambahkan label sehingga hasilnya dapat lebih bervariasi, kemudian menggunakan algoritma lain seperti *Naïve Bayes*, *K-Nearest Neighbors* sehingga dapat diketahui perbandingan hasil metodenya dalam menganalisis sentimen dalam menentukan polaritas kelas pengguna aplikasi Indihome TV, serta menambahkan data ulasan dalam bahasa asing seperti ulasan berbahasa Inggris.

DAFTAR PUSTAKA

- Al-amrani, Y., Lazaar, M., & Eddine, K. (2018). Sentiment Analysis Using Hybrid Method of. *Journal of Theoretical and Applied Information Technology*, 96(7), 1886–1895. www.jatit.org
- Albab, M. U., Karuniawati P, Y., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic. *Jurnal TRANSFORMATIKA*, 20(2), 1–10. <https://journals.usm.ac.id/index.php/transformatika/page1>
- Alwasi'a, A. (2020). ANALISIS SENTIMEN PADA REVIEW APLIKASI BERITA ONLINE MENGGUNAKAN METODE MAXIMUM ENTROPY (Studi Kasus: Review Detikcom pada Google Play 2019. In *Skripsi*.
- Arikunto, S. (2014). *Prosedur Penelitian Suatu Pendekatan Praktik*. Rineka.
- Bambang Seran, Y., & Supatman, S. (2024). Analisis Sentimen Masyarakat Terhadap Kinerja Kerja Presiden Joko Widodo Menggunakan Algoritma Support Vector Machine. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(4), 7190–7195. <https://doi.org/10.36040/jati.v8i4.10171>
- Budianto, A., Ariyuana, R., & Maryono, D. (2019). Perbandingan K-Nearest Neighbor (Knn) Dan Support Vector Machine (Svm) Dalam Pengenalan Karakter Plat Kendaraan Bermotor. *Jurnal Ilmiah Pendidikan Teknik Dan Kejuruan*, 11(1), 27. <https://doi.org/10.20961/jiptek.v11i1.18018>
- Faradian, H., Rubhasy, A., & Wijaya, Y. F. (2024). Analisis Sentimen Terhadap Penutupan Tiktok Shop Menggunakan Algoritma Naive Bayes Classifier Pada Media Sosial X. *Jurnal Ilmiah Sains Dan Teknologi*.
- Ferdiana, R., Jatmiko, F., Purwanti, D. D., Ayu, A. S. T., & Dicka, W. F. (2019). Dataset Indonesia untuk Analisis Sentimen. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(4), 334. <https://doi.org/10.22146/jnteti.v8i4.533>
- Fitri, D. A., & Putri, A. (2022). Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine. *Jurnal Computer Science and Information Technology (CoSciTech)*, 3(3), 472–478.
- Fridayanti, R. (2023). *Analisis Sentimen Review Pengguna Aplikasi*

- Photomath dengan Metode Support Vector Machine (SVM).*
- Handayanto, R. T., & Herlawati. (2020). *Data Mining dan Machine Learning Menggunakan Matlab dan Python*. Informatika.
- Herdhianto, A. (2020). *Sentiment Analysis Menggunakan Naïve Bayes Classifier (NBC) pada Tweet Tentang Zakat*.
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293. <https://doi.org/10.24114/cess.v5i2.18186>
- Ipmawati, J., Saifulloh, S., & Kusnawi, K. (2024). Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 247–256. <https://doi.org/10.57152/malcom.v4i1.1066>
- Mahardika, Y. S., & Zuliarso, E. (2018). Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes. *Prosiding SINTAK 2018, 2015*, 409–413.
- Ngafifi, M. (2014). Kemajuan Teknologi Dan Pola Hidup Manusia Dalam Perspektif Sosial Budaya. *Jurnal Pembangunan Pendidikan: Fondasi Dan Aplikasi*, 2(1), 33–47. <https://doi.org/10.21831/jppfa.v2i1.2616>
- Santoso, I., & Madiistriyatno, H. (2021). *Metodologi Penelitian Kuantitatif*. Indigo Media. https://books.google.co.id/books?id=bRFTEAAAQBAJ&printsec=frontcover&hl=id&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Sitanayah Que, V. K., Iriani, A., & Dwi Purnomo, H. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization). *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 162–170. www.tripadvisor.com,
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Alfabeta.
- Telkom. (2021). Digitalization for a Better Future. *Digitalization for a Better Future*, 1–486. <https://www.telkom.co.id>
- Ubaidillah, M. J., Munadhif, I., & Rinanto, N. (2019). Klasifikasi Gelombang Otot Lengan Pada Robot Manipulator Menggunakan Support Vector Machine. *Rekayasa*, 12(2), 91–97. <https://doi.org/10.21107/rekayasa.v12i2.5406>
- Utomo, D. P. (2020). *Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung*. 4(April), 437–444. <https://doi.org/10.30865/mib.v4i2.2080>