



ISSN : 2339 - 1871

JURNAL ILMIAH BETRIK

Besemah Teknologi Informasi dan Komputer

Editor Office : LPPM Sekolah Tinggi Teknologi Pagar Alam, Jln. Masik Siagim No. 75
Simpang Mbacang, Pagar Alam, SUM-SEL, Indonesia
Phone : +62 852-7901-1390.
Email : betrik@sttpagaralam.ac.id | admin.jurnal@sttpagaralam.ac.id
Website : <https://ejournal.sttpagaralam.ac.id/index.php/betrik/index>

PENERAPAN NATURAL LANGUAGE PROCESSING UNTUK ANALISIS SENTIMEN TERHADAP APLIKASI STREAMING

Paskariawan br Sihotang¹, Fanny Dameka br Sitanggung², Nadila Azriansyah³, Evta Indra^{4*}

Program Studi Sistem Informasi Universitas Prima Indonesia

Jl. Sampul No.4, Sei Putih Bar., Kec. Medan Petisah, Kota Medan, Sumatera Utara

Sur-el : evtaindra@unprimdn.ac.id

Abstrak: Aplikasi video streaming seperti Viu, Viki, dan WeTV sangat populer karena memberikan akses yang mudah dan nyaman bagi pengguna untuk menonton konten multimedia mereka. Menurut data dari *Google Play Store* per akhir tahun 2022, Viu, Viki, dan WeTV adalah layanan streaming teratas. VIU telah menerima lebih dari 100 juta unduhan dan 993.000 ulasan, Viki telah menerima lebih dari 50 juta unduhan dan 899.000 ulasan, dan WeTV telah menerima lebih dari 50 juta unduhan dan 545.000 ulasan meskipun dinilai terbaik dan menerima peringkat 5. Sangat penting untuk terus menilai seberapa baik layanan disampaikan kepada pengguna karena peringkat dan jumlah ulasan tidak selalu berarti bahwa tuntutan atau masalah pengguna telah terpenuhi. diperlukan sistem untuk memproses semua input pengguna, termasuk peringkat dan komentar. Untuk memastikan pengalaman yang baik bagi pengguna, penting bagi pengembang aplikasi untuk memahami perasaan dan harapan pengguna. karena itu, dilakukan penelitian untuk Analisis Sentimen terhadap Aplikasi Streaming Video menggunakan Pendekatan *Support Vector Machine* (SVM). Data diambil melalui proses scraping yang didasarkan pada informasi dari *Google Play Store* , pada aplikasi Viu, Viki, dan WeTV dengan jumlah 2400 data. Berdasarkan data yang ditemukan ulasan terbanyak terdapat pada ulasan tidak konsisten sebesar 89.5% dengan jumlah 2399 data. kemudian, pada ulasan dengan menggunakan diagram pie chart berdasarkan jam, data tersebut menggambarkan jumlah ulasan pada setiap jam terlihat bahwa jam 14.00 memiliki jumlah ulasan terbanyak dengan 7.4%. Pada penelitian ini diperoleh hasil berupa akurasi untuk pengujian model Regresi Logistik adalah 0,90 untuk data pelatihan dan 0,89 untuk data pengujian, selanjutnya pada pengujian model Decision Treedidapatkan hasil akurasi sempurna pada data latih sebesar 1.0, dan pada data uji sebesar 1.0.dan Hasil untuk pengujian akurasi menggunakan model SVM adalah 0,968 untuk data pelatihan dan 0,910 untuk data pengujian.

Kata Kunci: Penerapan Neuro Linguistic Program, Analisis Sentimen, Aplikasi Streaming.

Abstract: Video streaming applications such as Viu, Viki, and WeTV have gained immense popularity due to their easy and convenient access to multimedia content for users. According to data from the *Google Play Store* as of the end of 2022, Viu, Viki, and WeTV are among the top streaming services. Viu has received over 100 million downloads and 993,000 reviews, Viki has received over 50 million downloads and 899,000 reviews, and WeTV has received over 50 million downloads and 545,000 reviews, even though they are rated highly and have a 5-star rating. It is crucial to continually assess the quality of service delivered to users because ratings

and review counts do not always indicate whether user demands or issues have been met. Therefore, a system is needed to process all user inputs, including ratings and comments. To ensure a good user experience, it is important for application developers to understand user sentiments and expectations. Hence, research is conducted for Sentiment Analysis on Video Streaming Applications using the Support Vector Machine (SVM) approach. Data is collected through a scraping process based on information from the Google Play Store for the Viu, Viki, and WeTV applications, with a total of 2400 data points. Based on the data found, the majority of reviews are inconsistent, accounting for 89.5% with 2399 data points. Furthermore, using a pie chart based on the time of day, the data illustrates the distribution of reviews throughout each hour. It is observed that 14:00 has the highest number of reviews with 7.4%. This research yielded results in terms of accuracy for the Logistic Regression model, which achieved 0.90 for the training data and 0.89 for the testing data. Similarly, the Decision Tree model achieved perfect accuracy with a score of 1.0 for both the training and testing data. Lastly, the SVM model achieved an accuracy of 0.968 for the training data and 0.910 for the testing data

Keywords :Neuro linguistic programming, Sentiment Analysis, Applications streaming

1. PENDAHULUAN

Analisis sentimen adalah proses yang menggunakan Natural Language Processing (NLP) untuk secara otomatis menambang sikap, pendapat, pandangan, dan emosi dari teks, audio, kicauan, dan sumber data [1]. Mencari tahu apa yang dipikirkan populasi umum tentang sesuatu juga disebut dengan analisis sentimen [2]. Sentimen terhadap aplikasi streaming video merupakan kegiatan untuk memahami perasaan dan sudut pandang pengguna terhadap layanan dan produk yang dipakai. Aplikasi video streaming seperti Viu , Viki, dan WeTV sangat populer karena memberikan akses yang mudah dan nyaman bagi pengguna untuk menonton konten multimedia mereka. Menurut data dari *GooglePlay Store* per akhir tahun 2022, Viu, Viki, dan WeTV adalah layanan streaming teratas. VIU telah menerima lebih dari 100 juta unduhan dan 993.000 ulasan, Viki telah menerima lebih dari 50 juta unduhan dan 899.000 ulasan, dan WeTV telah menerima lebih dari 50 juta unduhan dan 545.000 ulasan meskipun dinilai terbaik dan menerima peringkat 5. Namun, ada beberapa masalah yang dapat muncul saat menggunakan aplikasi ini. Masalah utama termasuk kualitas gambar yang kurang baik, buffering yang berulang, dan gangguan koneksi. Ini dapat mempengaruhi pengalaman penonton dan mengurangi tingkat kepuasan mereka. Selain itu, masalah privasi dan keamanan juga

merupakan perhatian utama karena aplikasi seringkali meminta akses informasi pribadi pengguna. Masalah lain termasuk keterbatasan konten yang tersedia di masing-masing aplikasi dan biaya yang dibutuhkan untuk menonton konten tertentu, pengguna sering merasa kesulitan untuk menemukan konten yang mereka cari dan membandingkan harga antar aplikasi. Sangat penting untuk terus menilai seberapa baik layanan disampaikan kepada pengguna karena peringkat dan jumlah ulasan tidak selalu berarti bahwa tuntutan atau masalah pengguna telah terpenuhi [3]. Menurut penjelasan yang diberikan, sistem diperlukan untuk memproses semua input pengguna, termasuk peringkat dan komentar [4]. Untuk memastikan pengalaman yang baik bagi pengguna, penting bagi pengembang aplikasi untuk memahami perasaan dan harapan pengguna. Beberapa penelitian tentang analisis sentimen dalam ulasan aplikasi meliputi, antara lain: Perbandingan sentimen *Corona Virus Disease* (Covid19) 2019 di Twitter Saat mengkategorikan suasana hati di Twitter seputar Penyakit Virus Corona 2019 (Covid-19) Dalam situasi tweet yang membandingkan nilai akurasi metode *Support Vector Machine* (SVM) dan Logistic Regression [5], nilai akurasi teknik *Support Vector Machine* (SVM) adalah 91,15% lebih tinggi. Selain itu, penelitian dilakukan dengan judul Analisis Sentimen Zoom Cloud Meetings di *Play Store*

Menggunakan Naïve Bayes dan *Support Vector Machine* [6], angka akurasi yang diperoleh dengan menggunakan pendekatan *Support Vector Machine* adalah 81,22%. Teknik *Support Vector Machine* menghasilkan nilai akurasi sebesar 85,54%, menurut studi berjudul Analisis Sentimen pada Aplikasi Grab di *GooglePlay Store* Menggunakan *Support Vector Machine* [7]. Pendekatan *Support Vector Machine* merupakan salah satu pilihan metodologi yang akan dimanfaatkan peneliti dalam analisis sentimen, menurut banyak referensi studi yang telah diselesaikan [8]. Untuk menentukan apakah dokumen tertulis mengandung sikap positif, negatif, dan netral, analisis sentimen menggunakan NLP [9]. Penting untuk mengklasifikasikan balasan dalam analisis sentimen untuk menentukan mana yang positif, negatif, dan netral [10]. Hal tersebut membantu pengembang aplikasi untuk memahami masalah yang dialami oleh pengguna. Diharapkan analisis sentimen terhadap aplikasi streaming video memberikan manfaat yang signifikan bagi pengembang aplikasi dan pengguna. Akibatnya, metode SVM digunakan dalam proses klasifikasi penelitian ini. Informasi ulasan yang digunakan mungkin konstruktif dalam bentuk rekomendasi atau destruktif dalam bentuk kritik terhadap layanan streaming. Analisis emosi digunakan untuk menentukan apakah suatu pendapat atau emosi memiliki kecenderungan untuk menjadi positif atau negatif [11]. Analisis sentimen dapat digunakan dalam situasi ini untuk memeriksa ulasan aplikasi yang tersedia di *GooglePlay Store*. Berdasarkan hal tersebut, penelitian ini mencoba menggunakan informasi ulasan dari *Google Play Store* untuk mengatasi masalah yang sekarang dialami platform streaming online. Metodologi dan jenis sumber data yang digunakan dalam penelitian ini berbeda dari penelitian sebelumnya. Pendekatan *Support Vector Machine* (SVM) dengan linear digunakan dalam penelitian penulis, dan jenis sumber data yang digunakan adalah data review aplikasi bernama Viu, Viki, dan

WeTV di *website Google Play Store*. Perumusan masalah dalam penelitian ini didasarkan pada deskripsi yang telah ditemukan, dan itu adalah Bagaimana NLP dapat digunakan untuk menganalisis ulasan pengguna dan komentar untuk sentimen. Tujuan dari penelitian ini; dataset yang diambil berupa hasil ulasan dari platform aplikasi streaming VIU,VIKI, dan WeTV dan dan Metode yang digunakan berupa *Support Vector Machine*. Batasan masalah pada penelitian ini adalah; dataset yang diambil berupa hasil ulasan dari platform aplikasi streaming VIU,VIKI, dan WeTV dan metode yang digunakan berupa *Support Vector Machine*.

2. METODE PENELITIAN

Pada penelitian ini diawali dengan studi literatur, data acquisition, data cleaning dan menggunakan metode Support Vector Machine (SVM).



Gambar 1. Metode Penelitian

2.1 Studi Literatur

Kegiatan penelitian yang dikenal sebagai "studi literatur" memanfaatkan metode pengumpulan informasi dan data serta berbagai sumber daya tambahan yang ditemukan di perpustakaan, seperti ensiklopedi, laporan dari studi serupa, artikel jurnal, catatan, dan bahan lain yang berkaitan dengan masalah yang dihadapi [12]. Jurnal penelitian dari penelitian sebelumnya, tesis, dan buku tentang penelitian semuanya dapat digunakan untuk menemukan literatur [13].

2.2 Data Acquisition

Tahap awal dalam pelabelan data melibatkan akuisisi data [14]. Suatu sistem yang bertindak untuk memperoleh, mengumpulkan, dan menyiapkan data sebelum mengolahnya untuk menghasilkan data yang dibutuhkan dapat disebut sebagai sistem akuisisi data [15]. Pada penelitian ini dataset yang digunakan berdasarkan data google play store di akhir tahun 2022 aplikasi Viu, Viki, dan WeTV data ini memiliki 4 kolom dan 1200 baris pengguna, jika ditotal dari ke-2 aplikasi data ini memiliki 2.400 baris atau data. Yang dimana data ini berisikan review atau ulasan pengguna aplikasi Viu, Viki, dan WeTV. Gambar di bawah ini menunjukkan data secara rinci.

reviewId	username	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	replyAt
3db761a-8768-4b22-34b3-4674616a3	Siti wahidul hasanah	https://play.googleusercontent.com/ACB-R...	Saya sudah berlangganan premium tetapi tidak b...	1	124	2020-01-01 00:00:00	03-36 09:12:23	Hi Vius, mohon maaf atas kelaknyamanannya...	2023-09-04
345074e3-8144-42c3-95a2-3414a27514	Abigail Maryam15	https://play.googleusercontent.com/AGNmy...	Viu meresihkan. Dlu bs nonton tanpa data kalo ...	1	19	2020-01-01 00:00:00	03-09 10:24:10	Hi Abigail Maryam, mohon maaf sudah bikin sed...	2023-10-30
86419296-464b-4b7e-3a61-3442288969	Snake Pit	https://play.googleusercontent.com/ACB-R...	Gak bisa sinkron antara HuaweiTv, di tombol me...	1	337	2020-01-01 00:00:00	02-17 07:15:20	Hi Snake Pit, mohon maaf pastikan kamu sudah m...	2023-08-01
5594a696-6982-464b-9232-39e1950a5d	Wijanarko Rofiq Nugroho	https://play.googleusercontent.com/ACB-R...	Langganan premium melalui SIM card sangat	1	24	2020-01-01 00:00:00	03-10 16:34:11	Hi Vius, mohon maaf atas kelaknyamanannya...	2023-03-11

Gambar 2. Dataset

2.3 Data Cleaning

Himpunan data ini menjalani pembersihan data untuk membuatnya siap untuk analisis. Pembersihan data dilakukan dengan menggunakan teknik Natural Language Processing (NLP). Natural Language Processing (NLP) adalah subbidang kecerdasan buatan yang berfokus pada pemrosesan bahasa alami dan dapat digunakan untuk mengatasi masalah dengan sistem komputer yang mampu mengenali bahasa lisan umum [16]. Sebelum data diproses oleh model, selama fase pembersihan atau penghapusan data, upaya dilakukan untuk membersihkan data agar menghasilkan hasil output terbaik [17].

2.4 Support Vector Machine

Support Vector Machine (SVM) adalah nama metode pembelajaran mesin yang digunakan

untuk analisis data dan pengenalan pola [18]. Jika dibandingkan dengan algoritma lain, SVM adalah metode yang sangat dicari karena tingkat presisi / akurasi dan kualitasnya yang tinggi [19]. Python juga digunakan untuk melakukan analisis sentimen menggunakan teknik Support Vector Machine, dan hasilnya meliputi laporan klasifikasi, kapasitas untuk memulihkan data (recall), hasil presisi (presisi), dan hasil akurasi (akurasi) [20].

3. HASIL DAN PEMBAHASAN

3.1 Data Acquisition

Dalam penelitian ini pengumpulan data dari aplikasi streaming Wetv, Viu dan Viki yang dikumpulkan menggunakan web scraping, terdapat 2.400 rows × 12 columns. Hasil dari pengumpulan data tersebut dapat dilihat pada gambar 3.1

reviewId	username	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	replyAt	sortOrder
86a3364-484-4 Acun Adani	https://play.googleusercontent.com/ACB-R...	SO BADI! Aplikasi	1	3	22.8.2000	2022-08-18 15:58	We're truly sorry	2022-08-18 19:01	newest	c
241cc14-6904-4 Anu Anas	https://play.googleusercontent.com/ACB-R...	Kali ngi ada bath	1	0	22.8.2000	2022-08-11 13:4			newest	c
af5c325f-cdf-4 Neneng Fauziah	https://play.googleusercontent.com/ACB-R...	Nga bagas keb	1	7	22.7.2000	2022-08-06 4:09			newest	c
4f4b38e-171e-4 Azriel Prayoga	https://play.googleusercontent.com/ACB-R...	Gk bisa cari serk	1	0	22.5.2000	2022-08-04 3:34			newest	c
308743-849-4 Sindy Sindy olo	https://play.googleusercontent.com/ACB-R...	loading nya lama	1	1	22.8.2000	2022-08-03 13:1			newest	c
44cc04e-634-4 Citra Kurni	https://play.googleusercontent.com/ACB-R...	Ko ga ada sub tr	1	0	22.8.2000	2022-08-02 12:0			newest	c
9a6a3217-571d-4 Meki Aniska	https://play.googleusercontent.com/ACB-R...	Tidak ada yang l	1	0		2022-08-02 5:58			newest	c
c086b1b-4988-4 Agus wijaya	https://play.googleusercontent.com/ACB-R...	Mau berjalan bal	1	10	22.7.2000	2022-08-13 5:25			newest	c
faf6eab-2b79-4 Wildan 2048	https://play.googleusercontent.com/ACB-R...	ga ada sub indo	1	0		2022-08-08 13:0			newest	c
6e1a54a-2d55-4 nana ssi	https://play.googleusercontent.com/ACB-R...	GA BISA DITON	1	0		2022-07-30 4:39			newest	c
5989137-7921-4 Baby Unna	https://play.googleusercontent.com/ACB-R...	Kebanyakan kila	1	0	22.7.2000	2022-07-27 16:5			newest	c
af01151-196b-4 Anan Niad Ulen	https://play.googleusercontent.com/ACB-R...	Film atau drama	1	1	22.7.2000	2022-07-26 11:12	WV doesn't have	2022-05-18 11:3	newest	c
027a685-6d2b-4 Sri Helipah	https://play.googleusercontent.com/ACB-R...	Kik pas lagi non	1	0	22.7.2000	2022-07-22 12:5			newest	c
1178a6b-6a0e-4 Dini Dina	https://play.googleusercontent.com/ACB-R...	Diel	1	0	22.8.2000	2022-07-18 15:5			newest	c
346f793b-4a4a-4 Lailatul Zayyah	https://play.googleusercontent.com/ACB-R...	menarik aku sll	1	0		2022-07-15 11:6			newest	c
1115041-c13-4 Dan	https://play.googleusercontent.com/ACB-R...	Baru download l	1	0	22.6.2000	2022-07-15 6:51			newest	c
15d852b-eb7e-4 Wardi Bawal	https://play.googleusercontent.com/ACB-R...	Tela bahasa ind	1	0	22.6.2000	2022-07-10 1:14			newest	c
89885-58a6a-4181921c1 214	https://play.googleusercontent.com/ACB-R...	the animation is	1	0		2022-07-08 14:5			newest	c

Gambar 3. Hasil Pengumpulan Data

3.2 Pre-processing

Prosedur ini diperlukan untuk memperbaiki kesalahan dalam data mentah, yang seringkali tidak lengkap dan memiliki format yang salah. Preprocessing banyak digunakan untuk membersihkan, menyusun, dan menyiapkan data untuk memenuhi persyaratan analisis atau pemodelan yang akan dilakukan. Beberapa langkah dalam preprocessing data meliputi:

3.2.1 Data manipulation

Dalam kode program ini, menggunakan metode rename() pada objek dataframe untuk mengubah nama kolom. Setelah mengganti nama kolom, menggunakan df.head() untuk menampilkan beberapa baris pertama dari dataframe yang sudah

tidak baku agar dapat diproses menggunakan algoritma xgboost. Hal ini terlihat dalam Gambar 3.7

```
[ ] from Sastravi.Stemmer.StemmerFactory import StemmerFactory

# proses stemming
def stemming(ulasan):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in ulasan.split():
        dt = stemmer.stem(w)
        do.append(dt)
    df = " ".join(do)
    print(df)
    return df

df['Ulasan'] = df['Ulasan'].apply(stemming)

df.to_csv('data_stemming.csv', index=False)
df = pd.read_csv('data_stemming.csv', encoding='latin1')
df.head()
```

Gambar 8. Proses Stemming

C. Tokenizing

Untuk mempermudah pengolahan teks dengan menghilangkan tanda baca, angka, dan simbol yang tidak penting. Pada kode program ini, teks dalam kolom 'content' dipisah menggunakan tanda koma sebagai pembatas. Potongan-potongan teks kemudian disimpan dalam list bernama 'nstr', dan elemen-elemen kosong dalam list tersebut dihapus menggunakan perulangan. Hasil tokenisasi kemudian disimpan kembali ke dalam kolom 'content' pada dataframe df. seperti yang digambarkan pada Gambar 3.8.

```
def token(ulasan):
    if isinstance(ulasan, float):
        Ulasan = str(ulasan)
    nstr = Ulasan.split(',')
    data = []
    a = -1
    for hu in nstr: #mengiterasi melakukan perulangan setiap elemen dalam nstr
        a = a + 1
        if hu == '':
            data.append(a)
        p = 0
        b = 0
        for q in data:
            b = q - p
            del nstr[b]
            p = p + 1
    return nstr
df['Ulasan'] = df['Ulasan'].apply(lambda x: token(x) if isinstance(x, str) else x)
df.head(10)
```

Gambar 9. Hasil Tokenizing

3.2.6 Analisis Sentimen

Dalam analisis ini, terdapat tiga analisis sentimen yang dilakukan. Pertama, analisis sentimen berdasarkan ulasan sentimen yang mencakup ulasan positif, ulasan negatif, dan ulasan yang tidak konsisten. Kedua, dilakukan pencarian 10 kalimat dengan jumlah like terbanyak disetiap sentimen, yaitu 10 kalimat dengan jumlah like terbanyak pada ulasan positif dan ulasan negatif. Ketiga, dilakukan analisis persentase

ulasan berdasarkan jam, yang menunjukkan persentase ulasan yang diterima pada setiap jam dalam sehari.

Analisis sentimen terhadap ulasan Dapat dilakukan dengan menggunakan metode atau algoritma analisis sentimen yang telah diimplementasikan sebelumnya, Kode yang diberikan merupakan implementasi analisis pada kolom "Ulasan" dalam DataFrame. Setiap ulasan akan diklasifikasikan ke dalam tiga kategori sentimen: positif (nilai 2), negatif (nilai 0), atau tidak konsisten (nilai 1). seperti pada tabel berikut:

Tabel 1. Hasil Sentimen Terhadap Ulasan

1.(Tidak Konsisten)	2143
2.(Positif)	156
3.(Negatif)	95

Pencarian 10 kalimat dengan jumlah like terbanyak di setiap sentimen Setelah mendapatkan hasil analisis sentimen, maka langkah selanjutnya mencari 10 kalimat berdasarkan jumlah like terbanyak di setiap sentimen, seperti dibawah ini

Kalimat terpositif:	Id Aplikasi	Ulasan	sentiment	Jumlah_like
1633	com.tencent.qqvideobn	[aku benar suka tonton drama atau film di apli...	-0.116667	1152
2063	com.tencent.qqvideobn	[sebenarnya aku nyaman pake viu udah dari lama...	-0.500000	1113
2071	com.tencent.qqvideobn	[app viki bagus buat nonton drakor sy kasih b...	-0.500000	592
2245	com.tencent.qqvideobn	[kalo turut aku aplikasi cukup bagus cuma minu...	-0.100000	541
1631	com.tencent.qqvideobn	[kalo turut aku aplikasi cukup bagus cuma minu...	-0.100000	470
1888	com.tencent.qqvideobn	[ini aplikasi kenapa awal bagus dan saya suka ...	-0.500000	324
1402	com.viki.android	[ini aplikasi kenapa awal bagus dan saya suka ...	-0.500000	314
2274	com.tencent.qqvideobn	[knp ga bisa login lagi sih padahal internet l...	-0.500000	309
2248	com.tencent.qqvideobn	[saya beri bintang 4 saja karena entah mengapa...	-0.800000	283
239	com.vuclip.viu	[so bad aplikasi untuk asi dramas movies tapi ...	-0.700000	261

Kalimat negatif:	Id Aplikasi	Ulasan	sentiment	Jumlah_like
2275	com.tencent.qqvideobn	[aku benar suka tonton drama atau film di apli...	-0.116667	710
401	com.vuclip.viu	[sebenarnya aku nyaman pake viu udah dari lama...	-0.500000	589
1421	com.viki.android	[app viki bagus buat nonton drakor sy kasih b...	-0.500000	192
1325	com.viki.android	[kalo turut aku aplikasi cukup bagus cuma minu...	-0.100000	153
1201	com.viki.android	[kalo turut aku aplikasi cukup bagus cuma minu...	-0.100000	153
1326	com.viki.android	[ini aplikasi kenapa awal bagus dan saya suka ...	-0.500000	139
1020	com.viki.android	[ini aplikasi kenapa awal bagus dan saya suka ...	-0.500000	139
1150	com.viki.android	[knp ga bisa login lagi sih padahal internet l...	-0.500000	126
1004	com.viki.android	[saya beri bintang 4 saja karena entah mengapa...	-0.800000	118
648	com.vuclip.viu	[so bad aplikasi untuk asi dramas movies tapi ...	-0.700000	34
850	com.viki.android			

Gambar 10. Hasil 10 Kalimat Dengan Jumlah Like Terbanyak Disetiap Sentimen

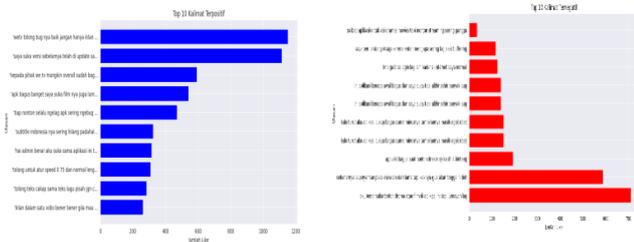
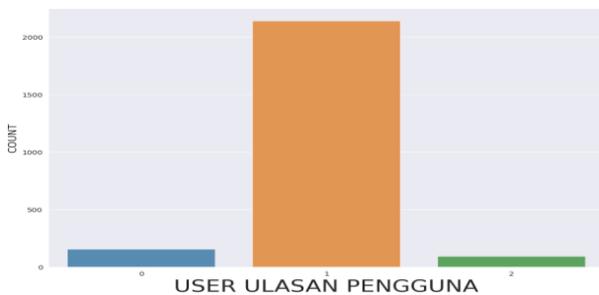
Analisis persentase ulasan berdasarkan jam Analisis persentase ulasan berdasarkan jam menunjukkan pola penggunaan Aplikasi Viu, Viki, dan WeTv pada berbagai waktu. Berdasarkan data yang dianalisis, terdapat variasi dalam jumlah ulasan pada jam-jam tertentu sebagai berikut:

Jam	jumlah_ulasan
00	87
01	112
02	119
03	94
04	134
05	125
06	129
07	114
08	104
09	88
10	112
11	119
12	147
13	169
14	176
15	126
16	116
17	79
18	58
19	21
20	32
21	34
22	54
23	54

Gambar 11. Hasil Persentase Ulasan Berdasarkan Jam

Visualisasi

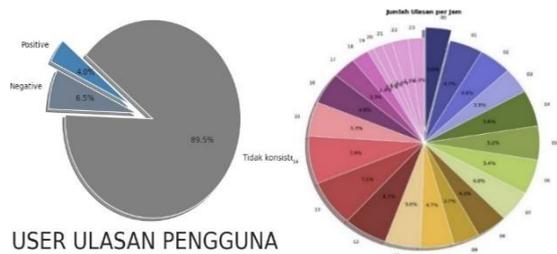
Dalam penelitian ini, melakukan dua tahap visualisasi untuk menganalisis ulasan pengguna. Tahap pertama, menggunakan bar plot untuk menampilkan jumlah ulasan pengguna berdasarkan sentimen. Hasil visualisasi ini memberikan gambaran yang jelas tentang distribusi ulasan pengguna dalam kategori sentimen positif, negatif, dan tidak konsisten. Selain itu, melalui visualisasi ini dapat melihat hasil terkait dengan kalimat 10 ternegatif berdasarkan jumlah dislike terhadap sentimen



Gambar 12. Visualisasi Bar Plot

Tahap kedua melibatkan penggunaan pie chart untuk memperlihatkan proporsi ulasan pengguna berdasarkan sentimen. *Pie chart* memberikan gambaran visual yang jelas mengenai persentase ulasan pengguna yang tergolong dalam kategori sentimen positif, negatif, dan tidak konsisten. Selain itu, tahap

ini juga mencakup visualisasi ulasan berdasarkan jam menggunakan diagram pie chart. Dengan menggunakan diagram pie chart berdasarkan jam, data tersebut menggambarkan jumlah ulasan pada setiap jam. Terlihat bahwa jam 14.00 memiliki jumlah ulasan terbanyak dengan 7.4% ulasan, diikuti oleh jam 13.00 dengan 7.1% ulasan. Jam-jam tersebut menunjukkan tingkat aktivitas pengguna yang tinggi dalam memberikan ulasan pada aplikasi tersebut.



Gambar 13. Hasil Visualisasi Pie Chart

3.3 TF IDF Vectorizer

Metode frekuensi istilah (TF) menentukan seberapa sering sebuah kata muncul dalam dokumen. Inverse Document Frequency (IDF) mengukur signifikansi kata di dalam korpus atau kumpulan dokumen. IDF digunakan untuk meningkatkan bobot kata-kata yang jarang muncul dan mengurangi bobot istilah yang biasa ditemukan dalam korpus.

3.3.1 Splitting Data

Proses memisahkan data menjadi dua atau lebih bagian untuk digunakan dalam tahap yang berbeda dalam analisis data. Biasanya, Data pelatihan dan data pengujian adalah dua kategori data. Data pengujian digunakan untuk menilai kinerja model setelah dilatih menggunakan data pelatihan. Tujuan dari *splitting data* untuk menghindari *overfitting* dan *underfitting* pada model, serta untuk mengevaluasi seberapa baik model dapat digeneralisasi pada data yang belum pernah dilihat sebelumnya. Setelah proses *splitting data* dilakukan, didapatkan hasil sebagai berikut: Jumlah data pada *training set* : 1915 Jumlah data pada *testing set* : 479

Tabel 2. Hasil Spliting data

Training set	1915
Testing set	479
Total	2394

3.3.2 Pengolahan Machine Learning

Dengan menyajikan data dengan algoritma pembelajaran, sekelompok teknik yang dikenal sebagai *Machine Learning* dapat membantu dalam menangani dan membuat prediksi tentang jumlah data yang sangat besar. penelitian ini melakukan pengolahan menggunakan tiga model machine learning, yaitu *Logistic Regression*, *Decision Tree*, dan *Support Vector Machine (SVM)*.

A Logistic Regression

Penelitian ini menggunakan model Logistic Regression dengan fitur TF-IDF untuk mengklasifikasikan sentimen pada data ulasan. Model ini memberikan akurasi 0.90 pada data latih dan 0.88 pada data uji. Dalam evaluasi model juga menggunakan *confusion matrix* untuk melihat seberapa baik model dapat mengklasifikasikan sentimen positif, negatif, dan tidak konsisten.

```
Confusion matrix on train is:
[[ 56 70  0]
 [ 0 1714  0]
 [  0  63 12]]
-----
Model accuracy on train is: 0.9305483028720627

Confusion matrix on test is:
[[ 5 25  0]
 [ 0 429  0]
 [ 0 20  0]]
-----
Model accuracy on test is: 0.906054279749478

Classification report on test is:
precision    recall  f1-score   support

0           1.00     0.17     0.29         30
1           0.91     1.00     0.95        429
2           0.00     0.00     0.00         20

accuracy          0.91         479
macro avg         0.64         0.39     0.41         479
weighted avg     0.87         0.91     0.87         479
```

Gambar 14. Hasil Logistic Regression

B Decision Tree

Dalam penelitian ini, peneliti menggunakan model *Decision Tree* untuk analisis sentimen. Model ini memiliki akurasi sempurna yaitu 1.0 pada data latih maupun

data uji. Selanjutnya, model dievaluasi menggunakan *confusion matrix* untuk melihat seberapa baik model mengklasifikasikan sentimen positif, negatif, dan tidak konsisten. Selain itu, peneliti juga menghitung *Kappa Score* sebagai matrix evaluasi tambahan. *Kappa Score* yang diperoleh adalah 1.0.

```
Model accuracy on train is: 1.0
Model accuracy on test is: 1.0
-----
confusion_matrix train is: [[ 126  0  0]
 [  0 1714  0]
 [  0  0 75]]
confusion_matrix test is: [[ 30  0  0]
 [  0 429  0]
 [  0  0 20]]
Wrong predictions out of total
-----
0 / 479
-----
KappaScore is: 1.0
```

Gambar 15. Hasil Decision Tree

C Support Vector Machine (SVM)

Dalam penelitian ini, peneliti menggunakan model *Support Vector Machine (SVM)* dengan kernel linear untuk klasifikasi. Model ini memiliki akurasi 0.968 pada data latih dan 0.910 pada data uji. *Confusion matrix* digunakan untuk melihat seberapa baik model mengklasifikasikan sentimen positif, negatif, dan tidak konsisten. Peneliti juga menyajikan classification report yang mencakup precision, recall, dan f1-score untuk setiap kelas sentimen.

Tabel 3. Hasil Pengujian Support Vector Machine (SVM)

Kelas	Precision	Recall	F1-Score	Support
Negatif(0)	1.00	0.40	0.57	30
Tidak konsisten(1)	0.92	1.00	0.96	429
Positif(2)	1.00	0.10	0.18	20

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil penelitian analisis sentimen data review pengguna dari Aplikasi *Viu*, *Viki*, dan *WeTv*. Ada banyak kesimpulan yang bisa diambil. Data diambil melalui proses *scraping* yang didasarkan pada informasi dari *Google Play Store*, pada

aplikasi *Viu*, *Viki*, dan *WeTv* dengan jumlah 2400 data. Berdasarkan data yang ditemukan ulasan terbanyak terdapat pada ulasan tidak konsisten sebesar 89.5% dengan jumlah 2399 data. Pada ulasan dengan menggunakan diagram pie chart berdasarkan jam, data tersebut menggambarkan jumlah ulasan pada setiap jam. Terlihat bahwa jam 14.00 memiliki jumlah ulasan terbanyak dengan 7.4%.

Temuan akurasi untuk pengujian model *Regresi Logistik* adalah 0,90 untuk data pelatihan dan 0,89 untuk data pengujian. Pada pengujian model *Decision Tree* didapatkan hasil akurasi sempurna pada data latih sebesar 1.0, dan pada data uji sebesar 1.0. Hasil untuk pengujian akurasi menggunakan model SVM adalah 0,968 untuk data pelatihan dan 0,910 untuk data pengujian. Hal ini menunjukkan bahwa banyaknya pengguna memberikan ulasan terdapat pada jam 14.00. Dari pengujian dengan model *DecisionTree* dan SVM didapatkan hasil akurasi lebih baik menggunakan model *decision tree* dengan akurasi data train sebesar 0.1 yang menunjukkan bahwa model tersebut dapat memprediksi dengan sempurna.

4.2 Saran

Beberapa ide yang dapat dipertimbangkan untuk penyelidikan tambahan meliputi:

- 1 Pada ulasan yang tidak konsisten, melakukan analisis yang mendalam dengan mengidentifikasi adanya sentimen yang tidak konsisten dalam data ulasan pengguna, perlu dilakukan analisis lebih lanjut untuk memahami penyebabnya.
- 2 Pada sentiment rentang waktu tidak ditemukan pola waktu mana yang lebih baik untuk melihat ulasan, maka penelitian lebih lanjut perlu mengidentifikasi pola waktu, di waktu mana yang banyak berdampak positif, negative, dan tidak konsisten pada setiap data ulasan.
- 3 Diperlukannya data lebih luas terhadap cakupan pengguna dan variabilitas akan memberikan temuan yang lebih

kredibel dalam memahami sentiment terhadap aplikasi streaming.

DAFTAR PUSTAKA

- [1] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, vol. 1, no. 1, pp. 299–305, 2021.
- [2] A. Andreyestha and A. Subekti, "Analisa Sentiment Pada Ulasan Film Dengan Optimasi Ensemble Learning," *J. Inform.*, vol. 7, no. 1, pp. 15–23, 2020, doi: 10.31311/ji.v7i1.6171.
- [3] M. R. Nadhif, D. Wisnu Brata, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi TIX ID di Indonesia pada Google Play Store menggunakan Support Vector Machine," vol. 6, no. 8, pp. 3932–3937, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] M. T. Anjasmoros, I. Istiadi, and F. Marisa, "Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode SVM Dan NBC (Studi Kasus: Komentar Pada Play Store)," *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2020)*, no. Ciastech, pp. 489–498, 2020.
- [5] K. Kelvin, J. Banjarnahor, E. I. -, and M. NK Nababan, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 47–52, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365.
- [6] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [7] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.

- [8] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, no. 11, pp. 4704–4713, 2018.
- [9] J. Nurvania and K. M. Lhaksamana, "Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM)," vol. 8, no. 4, pp. 4124–4135, 2021.
- [10] F. Sodik and I. Kharisudin, "Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," vol. 4, pp. 628–634, 2021.
- [11] M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [12] V. Melinda and M. Zainil, "Penerapan Model Project Based Learning untuk Meningkatkan Kemampuan Komunikasi Matematis Siswa Sekolah Dasar (Studi Literatur)," *J. Pendidik. tambusai*, vol. 4, pp. 1526–1539, 2020, [Online]. Available: <https://jptam.org/index.php/jptam/article/download/618/545>
- [13] S. N. Lase, Y. Yenny, O. Owen, M. Turnip, and E. Indra, "Application of Data Mining To Predicate Stock Price Using Long Short Term Memory Method," *Infokum*, vol. 10, no. 2, pp. 1001–1005, 2022, [Online]. Available: <http://infor.seaninstitute.org/index.php/inforokum/article/view/454>
- [14] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews," *2020 Int. Conf. Contemp. Comput. Appl. IC3A 2020*, no. May, pp. 217–220, 2020, doi: 10.1109/IC3A48958.2020.233300.
- [15] D. T. Husni *et al.*, "Analisis Big Data Penjualan Video Games Menggunakan Eda," *J. Tek. Inf. dan Komput.*, vol. 5, no. 1, p. 43, 2022, doi: 10.37600/tekinkom.v5i1.517.
- [16] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [17] R. Sitanggung *et al.*, "Sentiment Analysis Compare Linear Regression and Decision Tree Regression Algorithm To Determine Film Rating Accuracy," *Infokum*, vol. 10, no. 2, pp. 880–890, 2022.
- [18] H. Christanto, J. Rahmad, S. H. Sinurat, and D. Ryan, "Analisis Perbandingan Decision Tree , Support Vector Machine , dan Xgboost dalam Mengklasifikasi Review Hotel Trip Advisor," vol. 9, no. 1, pp. 306–319, 2023.
- [19] S. I. Nurhafida and F. Sembiring, "Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM)," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 317–327, 2022.
- [20] A. Triayudi and W. O. Widyarto, "Comparison J48 and Naïve Bayes Methods in Educational Analysis," *J. Phys. Conf. Ser.*, vol. 1933, no. 1, pp. 15–20, 2021, doi: 10.1088/1742-6596/1933/1/012062.