

Analisis Sentimen Pada Ulasan Aplikasi WeTV di Google Play Store Menggunakan Algoritma NBC dan SVM

Sentiment Analysis on WeTV App Reviews on Google Play Store Using NBC and SVM Algorithms

¹Petronilia Palinggik Allorerung*, ²Rismayani

¹Teknik Informatika, Universitas Dipa Makassar,

²Rekayasa Perangkat Lunak, Universitas Dipa Makassar

Jalan Perintis Kemerdekaan KM. 9, Kec. Tamalanrea, Kota Makassar, Sulawesi Selatan, Indonesia
90245

*e-mail: petroniliaallorerung@gmail.com

(received: 22 Desember 2022, revised: 19 Maret 2023, accepted: 24 Maret 2023)

Abstrak

Sejak wabah *Covid-19* melanda Indonesia, semua kegiatan masyarakat menjadi sangat terbatas. Keputusan pemerintah tentang PPKM demi menurunkan tingkat kasus *Covid-19*, membuat masyarakat harus menurunkan tingkat kegiatan di luar rumah termasuk bekerja. Kegiatan yang akhir-akhir ini digemari oleh masyarakat salah satunya adalah menonton film melalui layanan *streaming online* yang tersedia di *Google Play Store*. Aplikasi dengan total unduhan dan rating yang tinggi menunjukkan minat masyarakat pada aplikasi tersebut. Layanan *streaming online* WeTV merupakan salah satu aplikasi yang memiliki unduhan dan rating yang tinggi pada *Google Play Store*. Layanan ini menyajikan beragam jenis konten dari berbagai negara. Kendati demikian, aplikasi WeTV juga memiliki kekurangan yang terlihat pada ulasan-ulasan dari para pengguna. Berdasarkan hal tersebut dilakukan penelitian tentang klasifikasi sentimen positif dan negatif dari pengguna aplikasi WeTV. Algoritma klasifikasi yang diimplementasikan ada dua yakni Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM). Selain mengklasifikasikan ulasan positif dan negatif, analisis sentimen ini juga bertujuan untuk melakukan komparasi performa dari kedua algoritma tersebut. Total data yang digunakan adalah 100 data. Setelah melakukan analisis sentimen, diperoleh kesimpulan bahwa metode klasifikasi SVM merupakan metode klasifikasi terbaik pada penelitian ini dengan *accuracy* 80.00%, *precision* 80.00%, dan *recall* 80.00%.

Kata kunci: *Google Play Store, NBC, Sentimen, SVM, WeTV*

Abstract

Since the *Covid-19* outbreak hit Indonesia, all community activities have become very limited. The government's decision regarding PPKM to reduce the level of *Covid-19* cases forced the community to reduce the level of activities outside the home including work. One activity that has recently been popular with the public is watching movies through the online streaming service available on the *Google Play Store*. Applications with high total downloads and ratings show people's interest in the application. The WeTV online streaming service is an application that has high downloads and ratings on the *Google Play Store*. This service provides various types of content from various countries. However, the WeTV application also has drawbacks that can be seen in the reviews from users. Based on this, research was conducted on the classification of positive and negative sentiments from WeTV application users. There are two classification algorithms implemented, namely Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM). Apart from classifying positive and negative reviews, this sentiment analysis also aims to compare the performance of the two algorithms. The total data used is 100 data. After conducting sentiment analysis, it was concluded that the SVM classification method was the best classification method in this study with 80.00% accuracy, 80.00% precision, and 80.00% recall.

Keywords: *Google Play Store, NBC, Sentiment, SVM, WeTV*

1 Pendahuluan

Kasus *Covid-19* pertama kali muncul di Wuhan, China pada 31 Desember 2019 yang kemudian menyebar ke wilayah lain di China. Pada awalnya, kasus ini diduga sebagai penyakit pneumonia dengan gejala seperti influenza pada umumnya namun dapat berkembang dengan cepat bahkan dapat mengakibatkan infeksi parah dan gagal organ [1]. Kasus serupa juga segera terdeteksi di beberapa negara di setiap benua seperti Asia, Eropa, Australia, Afrika, dan Amerika [2]. Di Indonesia, kasus terinfeksi *Covid-19* pertama kali dilaporkan pada awal tahun 2020 dan menurut laporan Kementerian Kesehatan Republik Indonesia, sepanjang tahun 2020 terdapat 743.198 kasus [3]. Karena peningkatan kasus positif di Indonesia masih berlangsung pada tahun berikutnya, maka pemerintah segera menerapkan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) untuk mengurangi kegiatan diluar rumah terutama yang menyebabkan keramaian.

Penyebaran *Covid-19* yang sangat cepat dan belum ditemukannya vaksin, menimbulkan kekacauan ditengah masyarakat dan mempengaruhi banyak sektor kehidupan seperti ekonomi, politik, industri, pendidikan, medis, dan sosial [4]. Berbagai aktivitas yang sebelumnya dilakukan secara langsung seperti bekerja, belajar, berbelanja, dan menonton film di bioskop telah dilakukan secara *online* selama PPKM. Hal ini di dukung dari data Asosiasi Penyelenggara Jasa Internet Indonesia kuartal II tahun 2020 yang menyebutkan bahwa pengguna internet di Indonesia sudah mencapai 73,7% atau 196,71 juta pengguna [5].

Aktivitas yang belakangan ini digemari oleh masyarakat adalah menonton film melalui layanan *streaming online* yang bisa didapatkan pada *Google Play Store*. WeTV merupakan salah satu aplikasi *streaming online* populer milik perusahaan Tiongkok yaitu Tencent yang telah diunduh lebih dari 50 juta kali di *Google Play Store*. Aplikasi ini menayangkan berbagai jenis konten film, drama, acara olahraga, dan acara televisi dari berbagai negara, terutama Asia [6].

Banyaknya jumlah unduhan pada *Google Play Store*, menunjukkan minat masyarakat terhadap aplikasi ini. Meski begitu, setiap aplikasi tentu memiliki kelebihan dan kekurangan, termasuk aplikasi WeTV. Disamping tersedianya berbagai konten yang sedang populer yang menjadi salah satu keunggulan aplikasi ini, juga terdapat beberapa kekurangan yang terlihat pada ulasan-ulasan yang diberikan oleh pengguna mengenai kualitas dari aplikasi WeTV. Namun karena banyaknya ulasan tersebut, developer menjadi kesulitan untuk membaca semuanya. Proses ini tentu sangat tidak efektif karena memerlukan banyak waktu dan tenaga serta menghambat proses perbaikan aplikasi.

Untuk mengatasi masalah tersebut, dilakukan penelitian ini yang bertujuan untuk mengklasifikasikan ulasan pada aplikasi WeTV di *Google Play Store* dengan menerapkan NBC dan SVM serta mengkomparasi kedua algoritma tersebut untuk mengetahui algoritma dengan performa tertinggi didasarkan pada tingkat *accuracy*, *precision*, dan *recall*. Analisis sentimen tersebut dilakukan agar para developer dapat melakukan perbaikan untuk lebih meningkatkan kualitas dari aplikasi WeTV sehingga menarik lebih banyak minat masyarakat dengan memperhatikan ulasan yang diberikan oleh pengguna.

2 Tinjauan Literatur

Penelitian sebelumnya juga dilakukan menggunakan algoritma yang sama dipakai oleh penulis yaitu dengan judul “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine” yang dilakukan oleh Herlinawati et al. pada tahun 2020. Pada penelitian ini, Support Vector Machine menghasilkan kinerja lebih tinggi di mana akurasi sebesar 81,22% [7].

Penelitian lainnya yaitu “Sentimen Analisis Mengenai Aplikasi Streaming Film Menggunakan Algoritma Support Vector Machine di Play Store” yang dilakukan oleh Al-shufi dan Erfina pada tahun 2021 yang menunjukkan bahwa tingkat akurasi aplikasi *streaming film* tertinggi adalah Iflix 92,67%, Disney hotstar 69,33%, WeTV 64,67%, Netflix 81,33% dan Vidio 62,00% [8].

Penelitian selanjutnya yaitu “Analisis Sentimen pada Review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine” yang dilakukan oleh Wahyudi dan Kusumawardhana pada tahun 2021 yang menghasilkan akurasi sebesar 85,54% [9] dan “Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma SVM Dan NBC” yang dilakukan oleh Muhammadin dan Sobari pada tahun 2021 dan menjadikan SVM algoritma dengan performa terbaik dengan akurasi 83,3%

<http://sistemasi.ftik.unisi.ac.id>

[10]. Adapula penelitian lain yaitu “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier” yang dilakukan oleh Salsabila et al. pada tahun 2022 dan menghasilkan nilai akurasi sebesar 95,10% [11].

Penelitian berikutnya adalah “Analisis Sentimen Masyarakat Indonesia Mengenai Vaksin COVID-19 pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier dan Support Vector Machine” yang dilakukan oleh Permatasari dan Irhamah pada tahun 2022 dan menjadikan metode SVM dengan performa tertinggi dibandingkan NBC dan SVM kernel memperoleh performa klasifikasi tertinggi dibandingkan SVM kernel RBF [12]. Kemudian judul selanjutnya yaitu “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM” oleh Isnain et al. pada tahun 2021 dan menghasilkan tingkat akurasi sebesar 74% [13] dan “Klasifikasi Data Review IMDb Berdasarkan Analisis Sentimen Menggunakan Algoritma Support Vector Machine” yang dilakukan oleh Cahyani et al. pada tahun 2022 dan menghasilkan nilai akurasi sebesar 86,5% [14].

Selanjutnya penelitian dengan judul “Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naïve Bayes” yang dilakukan oleh Setiawan et al. pada tahun 2021 dan dihasilkan algoritma SVM yang terbaik dengan nilai akurasi 85% [15]. Lalu ada juga judul penelitian yang lain yaitu “Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 dengan Metode Naïve Bayes Classifier dan Support Vector Machine” yang dilakukan oleh Sujadi et al. pada tahun 2022 dan dihasilkan metode SVM sebagai metode terbaik dengan nilai akurasi 81,6% dan jika menggunakan Cross Validation akan dihasilkan rata-rata nilai akurasi metode SVM sebesar 74,4% [16].

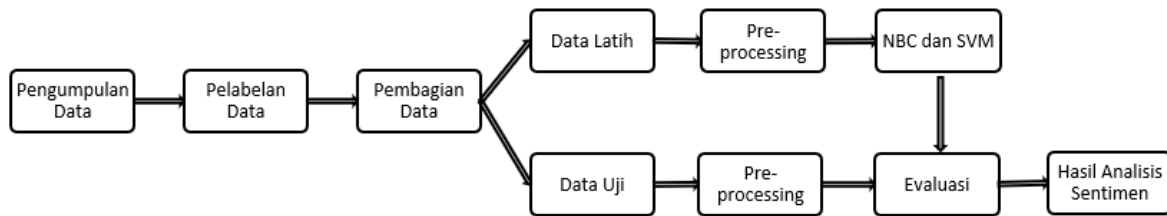
Ada juga penelitian dengan judul “Analisis Sentimen Review Aplikasi PeduliLindungi pada Google Play Store Menggunakan NBC” yang dilakukan oleh Locarso pada tahun 2022 dan dihasilkan nilai akurasi sebesar 83,3% [17]. Selanjutnya terdapat penelitian dengan judul “Penerapan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Terhadap Dampak Virus Corona di Twitter” yang dilakukan oleh Hasri dan Alita pada tahun 2022 dan dihasilkan metode Naïve Bayes sebagai metode dengan nilai akurasi tertinggi sebesar 81,07% [18].

Selain itu, terdapat penelitian berjudul “Perbandingan Metode Naïve Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca di Twitter” oleh Indriyani et al. pada tahun 2022 dan menjadikan Support Vector Machine sebagai metode dengan akurasi terbesar yaitu 87,27% [19] dan “Analisis Sentimen Aplikasi Transportasi Online KRL Access Menggunakan Metode Naïve Bayes” yang dilakukan oleh Nurwahyuni pada tahun 2019 dan dihasilkan nilai akurasi sebesar 84% [20]. Penelitian selanjutnya yaitu “Implementasi Algoritma Text Mining dan Cosine Similarity untuk Desain Sistem Aspirasi Publik Berbasis Mobile” yang dilakukan oleh Rismayani et al. pada tahun 2022 dan menghasilkan aplikasi yang menggunakan *method* text mining serta cosine similarity untuk pengukuran tingkat kemiripan fungsional setiap bidang pemerintahan dengan aspirasional yang dimasukkan masyarakat agar sesuai [21].

Berdasarkan penelitian di atas terutama pada penelitian ulasan pengguna pada suatu aplikasi, belum ada yang melakukan analisis sentimen yang berfokus pada satu layanan *streaming online*, yaitu WeTV kemudian melakukan komparasi terhadap dua algoritma. Maka dari itu, dilakukan pengelompokan sentimen pengguna terhadap aplikasi WeTV dengan NBC dan SVM lalu mengkomparasi keduanya untuk mendapatkan algoritma dengan performa tertinggi untuk *accuracy*, *precision*, dan *recall*.

3 Metode Penelitian

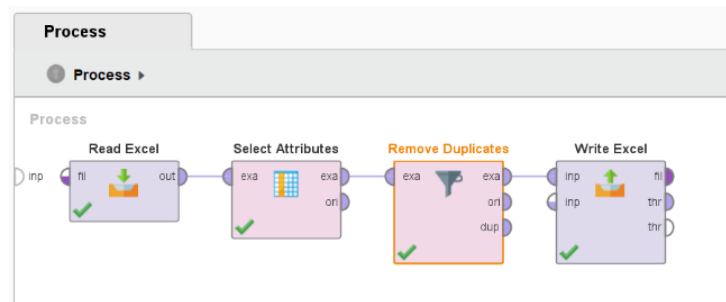
Metode yang dilakukan penulis pada proses analisis sentimen ini ditunjukkan pada Gambar 1.



Gambar 1. Tahap Analisis Sentimen

3.1 Pengumpulan Data

Tahap yang pertama adalah mengumpulkan data berupa ulasan pada WeTV di *Google Play Store* yang diperoleh melalui *tools Google Colab* menggunakan bahasa pemrograman Python. Data yang diperoleh dari hasil scrapping ini akan langsung di *export* ke dalam format excel. Dataset yang digunakan sebanyak 100 data dalam rentang waktu 25 Oktober 2022 sampai 2 Desember 2022, di mana terdiri dari 50 data positif dan 50 data negatif.



Gambar 2. Proses Penyimpanan Dataset

Gambar 2 menunjukkan bahwa data yang telah diperoleh dengan menggunakan bahasa pemrograman Python, akan diproses pada Rapidminer dengan hanya memilih atribut yang berisi ulasan atau komentar menggunakan operator “Select Atributes”, menghilangkan data yang duplikat dengan operator “Remove Duplicates”, kemudian dataset tersebut akan di simpan ke dalam format excel menggunakan operator “Write Excel”. Proses pengolahan dataset ini akan tetap menghasilkan 100 data karena tidak adanya duplikat pada data.

3.2 Pelabelan Data

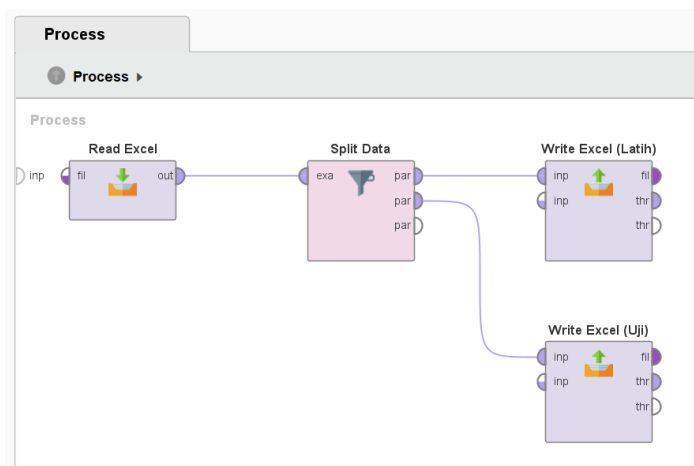
Tahap yang kedua yaitu melakukan pelabelan data untuk mengelompokkan sentimen sehingga membentuk *class* positif dan negatif. Pengelompokkan *class* ini didasarkan pada ulasan pengguna terhadap aplikasi WeTV. Jika ulasan tersebut berisikan kata positif atau netral dan merasa senang dalam menggunakan aplikasi WeTV, maka akan dikategorikan ke dalam *class* positif. Namun, jika ulasan tersebut berisikan kata negatif dan merasa kecewa dalam menggunakan aplikasi WeTV, maka akan dikategorikan ke dalam *class* negatif. Gambar 3 adalah contoh dari dataset yang berlabel.

	A	B
1	content	sentimen
2	Sbenarnya suka bnget ama wetv cm kok hars VIP jd hars pakai bayar klu bayar nya lewat liat ikl	positif
3	Sipp aplikasi hiburan terkeren bisa mengusir suntuk yang melanda We TV memang is the best	positif
4	Saya mau update ke Vlp gak bisa , gak ngerti dibilang untuk pembayaran error, ya sudah jadi cu	negatif
5	weTV makin di update bukannya tambah bagus trus nggak loading tp malah jd sering keluar se	negatif
6	Kenapa gak bisa daftar VIP ,, pembayaran selalu gagal... Tolong di perbaiki lagi min...	negatif
7	Sekarang kok kualitas we tv menurun, internet saya aktif, buat buka ig dll lancar saja, tp mau lia	negatif
8	Setelah diupdate ga makin bagus malah suka keluar2 sendiri. Enak2 nonton ehh lngsung tiba2	negatif
9	Berharap nanti unduhannya bisa disimpan di memori eksternal (ada pilihan tempat penyimpan	positif
10	Sudah langganan VIP tp masih ada iklan payahhhhhh...apalagi waktu masuk	negatif
11	Kenapa sih lagi enak2 nonton selalu iklan VIP tolong dong di perbaiki	negatif
12	Gak membantu . setiap mau nonton ada VIP harus membayar. Samakan aja sama YouTube ke	negatif

Gambar 3. Contoh Dataset Dengan Label

3.3 Pembagian Data

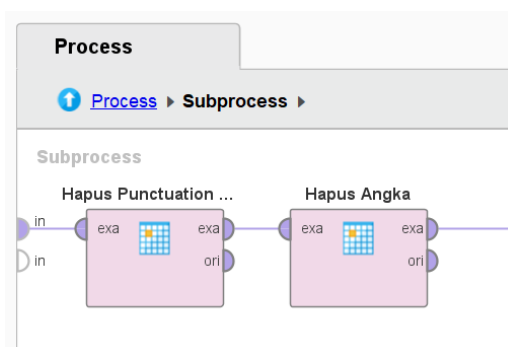
Tahap yang ketiga yaitu pembagian dataset jadi dua, yakni data latih 80% atau sebanyak 80 data sedangkan untuk data uji adalah 20% atau sebanyak 20 data. Keseluruhan dataset yang telah dibagi dapat dilihat pada https://drive.google.com/drive/folders/1SkXhsMv5nNgOIUfLNftJx7_ccEeK8ubo. Gambar 4 menunjukkan pembagian dataset menggunakan operator “Split Data”.



Gambar 4. Proses Pembagian Data

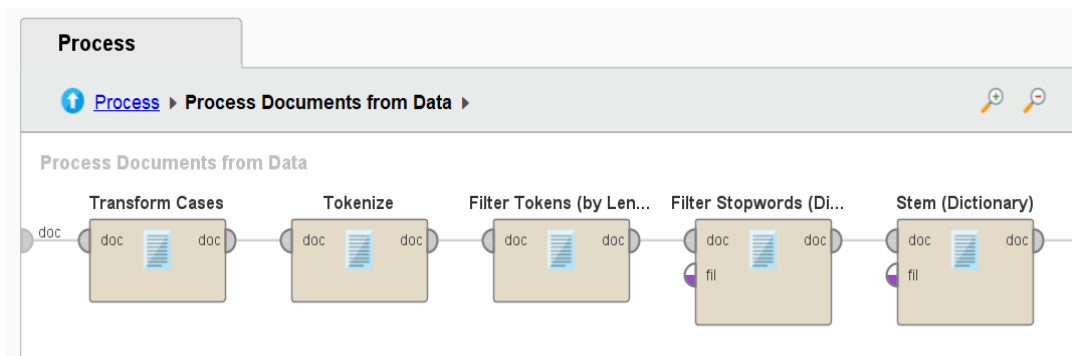
3.4 Pre-processing

Tahap berikutnya yaitu pre-processing terhadap dataset yang telah diberi label. Sebelum dilakukan klasifikasi, data latih dan data uji harus melakukan tahap pre-processing. Proses ini bertujuan untuk membuat data menjadi terstruktur sehingga dapat memudahkan dalam pemrosesan data. Tahap pre-processing ini berguna dalam mengurangi kata-kata yang tidak diperlukan. Tahap ini dibagi ke dalam dua operator yaitu, “Subprocess” dan “Process Documents from data”. Pada operator “Subprocess”, dilakukan tahap cleansing untuk menghapus simbol dan angka yang terdapat pada ulasan pengguna WeTV. Gambar 5 menunjukkan operator “Replace” yang digunakan pada tahap ini.



Gambar 5. Proses Pre-processing (Cleansing Data Dalam Subprocess)

Tahap kedua terjadi di operator “Process Documents from Data”. Adapun operator untuk melakukan pre-processing ini yaitu “Transform Cases”, “Tokenize”, “Filter Stopword (Dictionary)”, dan Stem (Dictionary). Gambar 6 menunjukkan model pre-processing yang terjadi dalam “Process Documents from Data”.



Gambar 6. Proses Pre-processing (Dalam Process Documents From Data)

Berikut penjelasan dari 7 tahapan pre-processing yang telah dilakukan.

1. Cleansing

Cleansing merupakan tahap untuk menghapus hal yang tidak dibutuhkan pada teks pada proses analisis sentimen seperti tag (#) username (@), angka, URL, RT, dan lain sebagainya. Tahap ini berfungsi untuk mengurangi noise pada dataset. Pada penelitian ini, dilakukan cleansing menggunakan parameter regular expression yang terdapat pada operator “Replace”. Adapun regular expression yang digunakan yaitu “[!\"#\$%&'()*+,-./:;<=>?@[\\|_`{|}~]” dan [0-9]. Tabel 1 merupakan contoh dari hasil cleansing.

Tabel 1. Contoh Hasil Cleansing

Sebelum	Sesudah
Kenapa suaranya suka hilang saat nonton, sebelumnya gak pernah gini..	Kenapa suaranya suka hilang saat nonton sebelumnya gak pernah gini

2. Case Folding

Transform case adalah operator yang digunakan pada tahap ini yang berfungsi untuk mengonversikan seluruh huruf menjadi besar atau kecil. Pada penelitian ini, akan dilakukan perubahan kata menjadi huruf kecil. Tabel 2 adalah contoh dari hasil tahap ini.

Tabel 2. Contoh Hasil Case Folding

Sebelum	Sesudah
Kenapa suaranya suka hilang saat nonton sebelumnya gak pernah gini	kenapa suaranya suka hilang saat nonton sebelumnya gak pernah gini

3. Tokenization

Tokenisasi merupakan tahap pemotongan *text* menjadi *term* agar semakin berarti. Tabel 3 merupakan contoh dari hasil tokenisasi.

Tabel 3. Contoh Hasil Tokenization

Sebelum	Sesudah
kenapa suaranya suka hilang saat nonton sebelumnya gak pernah gini	['kenapa', 'suaranya', 'suka', 'hilang', 'saat', 'nonton', 'sebelumnya', 'gak', 'pernah', 'gini']

4. Filter Token

Filter token merupakan tahap untuk pemilihan kata berdasarkan jumlah karakter yang telah ditentukan. Pada penelitian ini, kata dari proses tokenisasi yang memiliki 4 hingga 24 karakter akan diproses. Tabel 4 merupakan contoh dari hasil filter token.

Tabel 4. Contoh Hasil Filter Token

Sebelum	Sesudah
['kenapa', 'suaranya', 'suka', 'hilang', 'saat', 'nonton', 'sebelumnya', 'gak', 'pernah', 'gini']	['kenapa', 'suaranya', 'suka', 'hilang', 'saat', 'nonton', 'sebelumnya', 'pernah', 'gini']

5. Stopword Removal

Stopword removal yaitu teknik penghapusan token yang tidak menggambarkan sesuatu. Token yang dihapus didasarkan pada *dictionary* yang digunakan. Tabel 5 merupakan contoh dari hasil tahap ini.

Tabel 5. Contoh Hasil Penghapusan Token

Sebelum	Sesudah
['kenapa', 'suaranya', 'suka', 'hilang', 'saat', 'nonton', 'sebelumnya', 'pernah', 'gini']	['suaranya', 'suka', 'hilang', 'nonton', 'gini']

6. Stemming

Stemming merupakan tahap untuk mengonversikan *term* berimbuhan menjadi *term* dasar agar ulasan yang diperoleh menjadi lebih spesifik ketika dilakukan pengkategorian. Tabel 6 merupakan contoh dari hasil stemming.

Tabel 6. Contoh Hasil Stemming

Sebelum	Sesudah
['suaranya', 'suka', 'hilang', 'nonton', 'gini']	['suara', 'suka', 'hilang', 'nonton', 'gini']

7. Weighting

Weighting adalah tahap untuk melakukan pembobotan kata berdasarkan jumlah kemunculannya berdasarkan *Term Frequency-Inverse Document Frequency* (TF-IDF).

3.5 Klasifikasi Analisis Sentimen

Setelah semua tahap di atas selesai, kemudian dilakukan pengklasifikasikan ulasan atau data ke *class* tertentu. Proses tersebut bertujuan untuk melatih data dan menggunakan *method* klasifikasi. Penelitian ini menerapkan dua algoritma klasifikasi, sehingga digunakan operator “Naïve Bayes” dan “SVM” untuk dilakukan perbandingan hasil. Setelah model yang diusulkan dibuat, langkah selanjutnya yaitu melakukan pengujian model dengan dataset yang dipisah sebagai data latih dan data uji. Operator yang dioperasikan agar menyajikan nilai *accuracy*, *precision*, dan *recall* adalah “Performance”.

3.5.1 Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) adalah algoritma klasifikasi yang menerapkan teorema Bayes dalam perhitungan peluang. Berlandaskan teorema Bayes, maka dilakukan prediksi peluang di masa depan yang didasarkan pada pengetahuan pada masa lalu [11].

$$P(M|N) = \frac{P(N|M)P(M)}{P(N)} \tag{1}$$

Di mana:

P(M|N) = probabilitas M terjadi didasarkan peristiwa N

P(M) = probabilitas peristiwa M

P(N|M) = probabilitas N terjadi didasarkan peristiwa M

P(N) = probabilitas peristiwa N

3.5.2 Support Vector Machine

SVM yaitu *method* yang digunakan dalam analisis dan pengenalan model dalam pengklasifikasian. Penelitian ini, berfokus pada klasifikasi dua *class* yakni positif dan negatif. Teks merupakan data yang cocok mengimplementasikan algoritma SVM dalam pengklasifikasiannya karena cenderung berdimensi tinggi, yang memiliki fitur-fitur yang tidak relevan namun berkorelasi dengan yang lain dan disusun pada kelompok yang terbagi dengan linear [22].

3.6 Evaluasi Analisis Sentimen

Pada tahap ini, evaluasi dilakukan demi menguji hasil klasifikasi dengan mengukur tingkat keakuratan suatu metode dalam menganalisis komentar. Parameter pengujiannya adalah accuracy, precision, dan recall. Model confusion matrix dijelaskan pada Gambar 7 berikut.

	True Positive	True Negative
Prediction Positive	TP	FP
Prediction Negative	FN	TN

Gambar 7. Confusion Matrix

Accuracy merupakan total data yang terklasifikasi tepat, yaitu dimana hasil *prediction* sesuai dengan nilai label yang diatur sebelumnya [22]. Nilai accuracy dihasilkan melalui rumus:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

Precision adalah data yang diklasifikasikan menjadi True Positive dari semua data yang di *prediction* menjadi *class* positif [22]. Nilai precision dihasilkan melalui rumus:

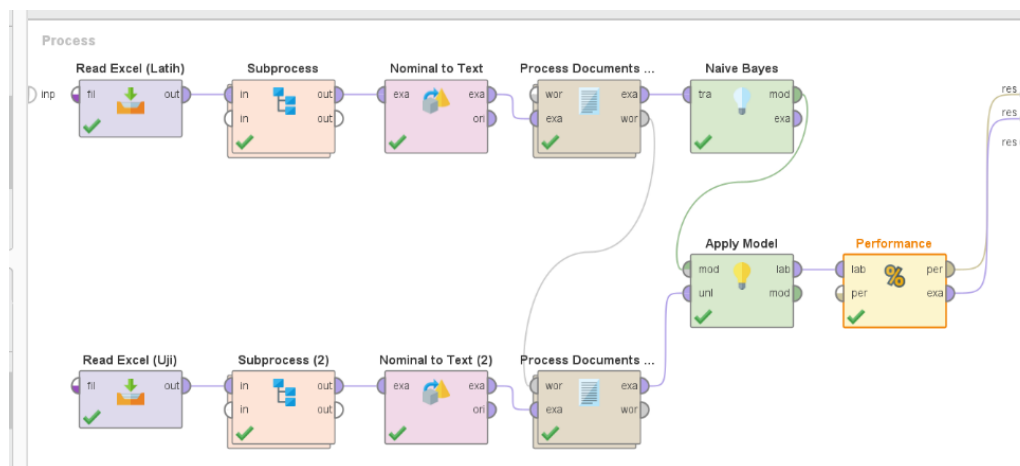
$$precision = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

Recall merupakan jumlah data terklasifikasi True Positive dari seluruh data yang sebenarnya positif [22]. Nilai recall dihasilkan melalui rumus:

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

4 Hasil dan Pembahasan

Penelitian ini bermaksud menganalisa sentimen yaitu ulasan pada WeTV di *Google Play Store*. Adapun algoritma yang diimplementasikan yakni NBC dan SVM lalu dilakukan komparasi algoritma tersebut agar mendapatkan algoritma dengan performa tertinggi. Dataset ulasan pada penelitian ini berjumlah 100 ulasan, dimana 80% menjadi data latih atau 80 ulasan dan 20% menjadi data uji atau 20 ulasan.



Gambar 8. Proses Utama

Gambar 8 menjelaskan proses utama dalam penelitian ini. Operator “Read Excel” berfungsi membaca dataset dengan format excel. Operator “Subprocess” berfungsi melakukan pre-processing tahap pertama yaitu cleansing data. Operator “Process Document from Data” berfungsi melakukan pre-processing tahap kedua dan pembobotan kata. Operator “Naïve Bayes” dan “SVM” merupakan operator dari algoritma klasifikasi yang digunakan. Operator “Apply Model” berfungsi untuk menerapkan model algoritma dari data latih agar melakukan pengklasifikasian pada data uji. Operator “Performance” digunakan untuk menampilkan nilai accuracy, precision, dan recall.

Row No.	sentimen	prediction(s...	confidence(...	confidence(...	text
1	negatif	positif	1	0	kualitas turun...
2	negatif	negatif	0	1	enak nonton i...
3	negatif	negatif	0	1	eror mulu no...
4	positif	negatif	0	1	daftar kode v...
5	positif	positif	1	0	bagus suka u...

Gambar 9. Hasil Prediksi Menggunakan RapidMiner

Gambar 9 menampilkan perbandingan antara label hasil prediksi dengan label sentimen yang telah ditentukan sebelumnya.

Tabel 7. Perbandingan Sentimen Dari Hasil Prediksi

Algoritma	Komentar Positif	Komentar Negatif
NBC	8	12
SVM	10	10

Tabel 7 menunjukkan perbandingan jumlah sentimen dari hasil prediksi ulasan. Dari analisis sentimen dengan algoritma Naïve Bayes Classifier, diperoleh sebanyak 8 data berisikan ulasan positif dan sebanyak 12 data berisikan ulasan negatif. Dari hasil tersebut, dapat diketahui bahwa ulasan pengguna aplikasi WeTV didominasi oleh komentar negatif. Sedangkan jika menggunakan algoritma Support Vector Machine, diperoleh sebanyak 10 data berisikan ulasan positif dan sebanyak 10 data berisikan ulasan negatif. Hasil tersebut menunjukkan bahwa jumlah komentar positif dan negatif terhadap aplikasi WeTV sama.

Tabel 8 di bawah ini, merupakan hasil confusion matrix dari *method* NBC dan SVM.

Tabel 8. Confusion Matrix Masing-Masing Algoritma

Algoritma	TP	FP	TN	FN
NBC	8	4	6	2
SVM	8	2	8	2

Berdasarkan Tabel 8, nilai *mean accuracy*, *precision*, dan *recall* ditunjukkan pada Tabel 9.

Tabel 9. Accuracy, Precision, dan Recall

Algoritma	Accuracy	Precision	Recall
NBC	70.00%	70.83%	70.00%
SVM	80.00%	80.00%	80.00%

Dari Tabel 9, dapat diketahui bahwa nilai *accuracy* NBC sebesar 70,00%, sedangkan SVM sebesar 80.00%. Nilai *precision* NBC sebesar 70.83% dan untuk SVM sebesar 80.00%. Sementara nilai *recall* NBC sebesar 70.00% dan nilai *recall* SVM sebesar 80.00%. Berdasarkan ketiga parameter tersebut, dapat dilihat bahwa klasifikasi dengan algoritma SVM adalah yang terbaik dalam menganalisis sentimen pada penelitian ini.

5 Kesimpulan

Pengklasifikasian sentimen di penelitian ini menerapkan dua *method* yakni NBC dan SVM untuk menganalisis ulasan aplikasi WeTV pada *Google Play Store*. Total dataset adalah 100 data yang berisikan 50 data positif dan 50 data negatif di mana persentase data latih 80% dan data uji 20%. Pemilihan algoritma terbaik pada analisis sentimen ini didasarkan pada nilai *accuracy*, *precision*, dan *recall*. Pada penelitian ini, hasil *accuracy* NBC dan SVM sebesar 70.00% dan 80.00%. Hasil *precision* NBC dan SVM sebesar 70.83% dan 80.00%. Dan hasil *recall* NBC dan SVM sebesar 70.00% dan 80.00%. Dari hasil tersebut, diketahui bahwa SVM mendapatkan performa tertinggi sehingga menjadi algoritma terbaik dalam penelitian ini, di mana tidak terdapat komentar yang mendominasi. Berdasarkan hasil tersebut, dapat diketahui bahwa total pengguna aplikasi WeTV yang merasa puas dan tidak puas sama.

Referensi

- [1] J. Majid, "Perlakuan Social Distancing: Upaya Masyarakat dan Tenaga Medis Dalam Mengurangi Penyebaran Virus Corona COVID-19," *J. Multidisiplin Dehasen MUDE*, vol. 1, no. 2, Art. no. 2, Apr. 2022, doi: 10.37676/mude.v1i2.2090.
- [2] K. C. Media, "Apa Itu Virus Corona? Halaman all," *KOMPAS.com*, Mar. 22, 2020. <https://www.kompas.com/skola/read/2020/03/22/183000269/apa-itu-virus-corona> (accessed Mar. 21, 2023).
- [3] Kompasiana.com, "Dampak akibat Covid-19 di Indonesia," *KOMPASIANA*, Jul. 16, 2021. <https://www.kompasiana.com/dinsalsabilla/60f11ef815251016f8403d22/dampak-akibat-covid-19-di-indonesia> (accessed Mar. 21, 2023).
- [4] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 1, Art. no. 1, Feb. 2021, doi: 10.20895/dinda.v1i1.180.
- [5] "Pengguna Internet Indonesia Mendekati Angka 200 Juta." <https://tekno.kompas.com/read/2020/11/10/07160027/pengguna-internet-indonesia-mendekati-angka-200-juta> (accessed Mar. 21, 2023).
- [6] "Download Aplikasi WeTV, Layanan Streaming Gratis! - Tertekno." <https://tertekno.com/download-aplikasi-wetv-layanan-streaming-gratis/> (accessed Mar. 21, 2023).
- [7] 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 J. Comput. Eng. Syst. Sci.*, vol. 5, no. 2, Art. no. 2, Jul. 2020, doi: 10.24114/cess.v5i2.18186.

- [8] M. F. Al-shufi and A. Erfina, "Sentimen Analisis Mengenai Aplikasi Streaming Film Menggunakan Algoritma Support Vector Machine di Play Store," *Semin. Nas. Sist. Inf. Dan Manaj. Inform. Univ. Nusa Putra*, vol. 1, no. 01, Art. no. 01, Sep. 2021.
- [9] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, Art. no. 2, Sep. 2021, doi: 10.31294/ji.v8i2.9681.
- [10] A. Muhammadin and I. A. Sobari, "Analisis Sentimen pada Ulasan Aplikasi Kredivo dengan Algoritma SVM dan NBC," *Reputasi J. Rekayasa Perangkat Lunak*, vol. 2, no. 2, Art. no. 2, Dec. 2021, doi: 10.31294/reputasi.v2i2.785.
- [11] S. M. Salsabila, A. A. Murtopo, and N. Fadhilah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier," *J. Minfo Polgan*, vol. 11, no. 2, Art. no. 2, Aug. 2022, doi: 10.33395/jmp.v11i2.11640.
- [12] R. W. Permatasari and I. Irhamah, "Analisis Sentimen Masyarakat Indonesia Mengenai Vaksin COVID-19 pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier dan Support Vector Machine," *J. Sains Dan Seni ITS*, vol. 11, no. 2, Art. no. 2, Jul. 2022, doi: 10.12962/j23373520.v11i2.73995.
- [13] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Sentimen Analisis Publik terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM," *J. Data Min. Dan Sist. Inf.*, vol. 2, no. 1, Art. no. 1, Feb. 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [14] G. Cahyani, W. Widayani, S. D. Anggita, Y. Pristyanto, I. Ikmah, and A. Sidauruk, "Klasifikasi Data Review IMDb Berdasarkan Analisis Sentimen Menggunakan Algoritma Support Vector Machine," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 3, Art. no. 3, Jul. 2022, doi: 10.30865/mib.v6i3.4023.
- [15] H. Setiawan, E. Utami, and S. Sudarmawan, "Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes," *J. Komtika Komputasi Dan Inform.*, vol. 5, no. 1, Art. no. 1, Jul. 2021, doi: 10.31603/komtika.v5i1.5189.
- [16] H. Sujadi, "Analisis Sentimen Pengguna Media Sosial Twitter terhadap Wabah Covid-19 dengan Metode Naive Bayes Classifier dan Support Vector Machine," *INFOTECH J.*, vol. 8, no. 1, Art. no. 1, Mar. 2022, doi: 10.31949/infotech.v8i1.1883.
- [17] G. K. Locarso, "Analisis Sentimen Review Aplikasi Pedulilindungi pada Google Play Store Menggunakan NBC," no. 2, p. 9, 2022.
- [18] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier dan Support Vector Machine pada Analisis Sentimen terhadap Dampak Virus Corona di Twitter," *J. Inform. Dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, Art. no. 2, Oct. 2022, doi: 10.33365/jatika.v3i2.2026.
- [19] E. R. Indriyani, P. Paradise, and M. Wibowo, "Perbandingan Metode Naïve Bayes dan Support Vector Machine Untuk Analisis Sentimen terhadap Vaksin Astrazeneca di Twitter," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 3, Art. no. 3, Jul. 2022, doi: 10.30865/mib.v6i3.4220.
- [20] S. Nurwahyuni, "Analisis Sentimen Aplikasi Transportasi Online KRL Access Menggunakan Metode Naive Bayes," *Swabumi Suara Wawasan Sukabumi Ilmu Komput. Manaj. Dan Sos.*, vol. 7, no. 1, Art. no. 1, May 2019, doi: 10.31294/swabumi.v7i1.5575.
- [21] R. Rismayani, H. Sy, T. Darwansyah, and I. Mansyur, "Implementasi Algoritma Text Mining dan Cosine Similarity untuk Desain Sistem Aspirasi Publik Berbasis Mobile," *Komputika J. Sist. Komput.*, vol. 11, no. 2, pp. 169–176, Aug. 2022, doi: 10.34010/komputika.v11i2.6501.
- [22] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin Elektro Dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, Oct. 2021, doi: 10.24176/simet.v11i2.4568.