

Analisis Sentimen Aplikasi Youtube di Google Play Store Menggunakan Machine Learning

Jimmy Alga, Cindi Wulandari*, Bunga Intan

Fakultas Ilmu Teknik, Program Studi Sistem Informasi, Universitas Bina Insan, Lubuklinggau, Indonesia
Email: ¹2002030047@mhs.univbinainsan.ac.id, ^{2,*}cindi_wulandari@univbinainsan.ac.id, ³bungaintan@univbinainsan.ac.id
Email Penulis Korespondensi: cindi_wulandari@univbinainsan.ac.id

Abstrak—Pengguna YouTube dapat membuat, menonton, dan berbagi video secara gratis. Interaksi antara penonton terjadi melalui fitur komentar, yang dapat bersifat positif atau negatif. Seringnya muncul komentar-komentar negatif pada aplikasi youtube di google play store ndapt memberikan efek kepada kaun-akun tersebut. Namuan untuk mengetahui berapa besar komentar-komnetar negatifa pada akun tersebut dibuthkan sebut dibutuhkan sebuah algoritma aitu SVM. Penelitian ini bertujuan untuk mengetahui sentimen terhadap aplikasi youtube di google play store menggunakan Machine Learning dengan algoritma SVM. Data yang di ambil adalah 4996 data ulasan yang kemudian dilakukan preprocessing data sehingga tersisa menjadi 4993 data yang bisa diolah. Lebellng data dilakukan otomatis berdasarkan score rating ulasan. Hasil labelling data dibagi menjadi 3 kelas yaitu kelas positif sebanyak 1083, kelas negatif sebanyak 3365 dan netral sebanyak 545. Klasifikasi dan evaluasi dilakukan menggunakan metode SVM. Berdasarkan nilai perbandingan data training dan testing sebesar 9:1, memperoleh hasil tingkat akurasi sebesar 75% kemudian presisi kelas negatif 76% dan recall kelas negatif 97% serta pengujian K-Fold Cross Validation menggunakan nilai K=10 dengan rata-rata Akurasi sebesar 0.75 atau 75%.

Kata Kunci: Youtube; Anlisis Sentimen; Google Play Store; Machine Learning

Abstract—YouTube users can create, watch, and share videos for free. Interaction between viewers occurs through the comment feature, which can be positive or negative. The frequent appearance of negative comments on the youtube application on the google play store can have an effect on these accounts. But to find out how much negative comments on the account are needed, an SVM algorithm is needed. This study aims to determine the sentiment towards the youtube application on the google play store using Machine Learning with the SVM algorithm. The data taken is 4996 review data which is then preprocessed so that the remaining data becomes 4993 data that can be processed. Data labelling is done automatically based on the review rating score. The results of data labelling are divided into 3 classes, namely positive classes as many as 1083, negative classes as many as 3365 and neutral as many as 545. Classification and evaluation are carried out using the SVM method. Based on the training and testing data comparison value of 9: 1, the results obtained an accuracy rate of 75% then negative class precision of 76% and negative class recall of 97% and K-Fold Cross Validation testing using a value of K = 10 with an average accuracy of 0.75 or 75%.

Keywords: Youtube; Sentiment Analysis; Google Play Store; Machine Learning

1. PENDAHULUAN

Era digital telah membuat manusia memasuki gaya hidup baru yang tidak bisa lepas dari perangkat yang serba elektronik. Teknologi menjadi alat yang membantu kebutuhan manusia, dengan teknologi, apapun dapat dilakukan dengan lebih mudah. Begitu pentingnya peran teknologi inilah yang mulai membawa peradaban memasuki ke era digital. Meningkatnya kebutuhan akan data dan informasi mendorong manusia untuk mengembangkan teknologi baru agar pengolahan data dan informasi dapat dilakukan dengan mudah dan cepat[1]. Era Internet of Things (IoT) menjadikan perkembangan teknologi tersebut dapat dinikmati melalui gadget. Apple, Windows dan Android menyediakan berbagai platform untuk menikmati konten-konten digital yang dapat diakses dengan aplikasi dan kuota data internet. Konten visual seperti berita teks atau info grafis dapat dinikmati melalui web browser, konten audio dapat didengarkan melalui Spotify atau konten audio visual yang dapat ditonton melalui YouTube. Youtube merupakan sebuah media sosial yang memiliki konten berbentuk video[2]. Dimana, videonya bertujuan sebagai hiburan, pengetahuan dan lain sebagainya. Pengguna youtube juga dapat membuat, melihat, serta berbagi video secara gratis.

Di youtube juga disediakan fitur komentar sebagai penyalur respon viewers (yang menonton video) terhadap konten video yang di tontonnya. Komentar tersebut bisa bersifat positif atau negatif. Disisi lain situs youtube juga dapat memberikan umpan balik (feedback) berupa like, dislike, dan komentar terhadap video yang ditayangkan di media youtube. Namun, aplikasi sosial media tentu memiliki kelebihan dan kekurangan, termasuk youtube. Pengalaman pengguna youtube sebagai aplikasi media sosial berbentuk video sering diungkapkan di kolom komentar di google play Store baik berupa kritik maupun kepuasan[3]. Perasaan negatif atau positif pengguna tentang suatu hal yang berhubungan dengan berbagai aspek dari sosial media ini. Fitur komentar di YouTube memainkan peran penting sebagai saluran komunikasi antara pemirsa dan pembuat konten. Hal ini memungkinkan penonton untuk memberikan respons, baik positif maupun negatif, terhadap konten yang mereka tonton.

Komentar tersebut dapat menjadi sumber umpan balik berharga bagi pembuat konten untuk memahami preferensi dan kebutuhan audiens mereka. Namun, keberadaan komentar juga dapat menjadi sarana untuk menyebarkan konten yang tidak pantas atau tidak sesuai dengan etika tertentu, menciptakan lingkungan yang kurang aman dan nyaman bagi beberapa pengguna. Di sisi lain, fitur umpan balik seperti tombol "like" dan "dislike" serta komentar yang diberikan oleh pengguna YouTube juga memiliki kelebihan dan kekurangan. Umpan balik positif dalam bentuk like dapat memberikan dorongan moral kepada pembuat konten, menunjukkan bahwa konten mereka dihargai oleh audiens. Namun, umpan balik negatif dalam bentuk dislike atau komentar negatif juga dapat berdampak negatif pada motivasi dan kepercayaan diri

pembuat konten, terutama jika disertai dengan kritik yang tidak membangun atau menghina. Sebagai hasilnya, lingkungan media sosial seperti YouTube dapat menjadi tempat yang menantang secara emosional bagi para penggunanya, di mana perasaan positif dan negatif bisa saling berinteraksi dalam dinamika yang kompleks[4].

Dengan tersedianya youtube pada platform digital seperti google play store, hal inilah yang membuat setiap pengguna yang mengunduh aplikasi youtube dapat memberikan ulasan terhadap aplikasi ini melalui fitur yang sudah disediakan oleh google play store. Ulasan tersebut dapat diakses secara bebas oleh siapa saja. Data ulasan dari aplikasi youtube dapat bermanfaat apabila diolah dengan baik. Hasil pengolahan data akan menghasilkan masukan untuk perbaikan dan pengembangan aplikasi, karena kritik dan saran terbaik berasal dari pendapat langsung semua pengguna. Pendapat dari setiap pengguna pada google play dapat berpengaruh pada calon pengguna sebagai bahan pertimbangan penggunaan suatu aplikasi[5]. Namun, dengan banyaknya data ulasan yang ada, akan sulit diolah secara manual. Dengan demikian, sangat diperlukan metode untuk melihat kecenderungan pengguna terhadap aplikasi secara otomatis, apakah positif atau negatif. Melihat pentingnya ulasan pengguna terhadap keberlangsungan suatu aplikasi, maka dari itu dirasa perlu dilakukan analisis sentimen pada data ulasan aplikasi youtube di google play store[6]. Analisis sentimen atau opinion mining adalah studi komputasional dari opini pengguna internet, sentimen dan emosi melalui entitas dan atribut yang dimiliki dan diekspresikan dalam bentuk teks.

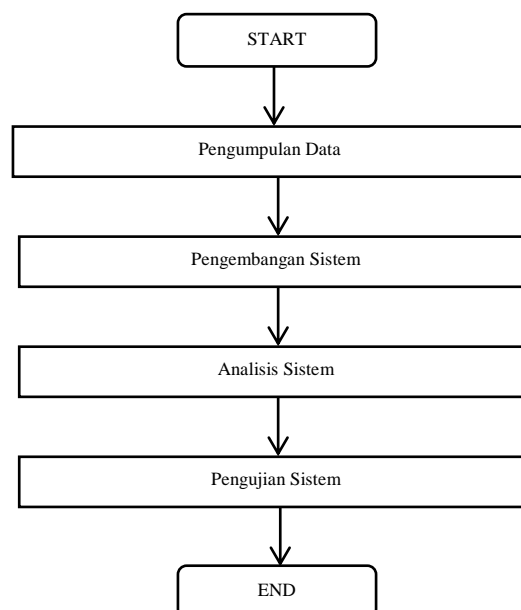
Analisis sentimen akan mengelompokkan (klasifikasi) polaritas dari teks ke dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam bentuk kalimat atau dokumen tersebut bersifat positif negatif atau netral[7]. Pada penelitian terdahulu yang dilakukan oleh [8] Savitri dkk, tentang Dari empat algoritma Supervised Machine Learning yang telah diuji, Logistic Regression menunjukkan akurasi tertinggi, yaitu 0,87. Analisis menunjukkan bahwa masyarakat cenderung netral terhadap pelaksanaan sekolah online. Pada Penelitian selanjutnya yang dilakukan oleh [9] Mualfah Ramadhoni dkk, tentang Nilai performa dari hasil pengujian menggunakan metode Support Vector Machine mendapatkan Akurasi 95.02%, Recall 95.02%, Precision 95.18% dan F1-Score 95.01%. Pada Penelitian selanjutnya yang dilakukan oleh [10] Idris Mustofa dkk, tentang algoritma Support Vector Machine terbukti mampu menghasilkan kinerja yang cukup baik dengan hasil akurasi sebesar 98% dan f1-score sebesar 0.98 atau sebesar 98%. Pada Penelitian selanjutnya yang dilakukan oleh [11] Hikmawan, tentang Support Vector Machine memiliki akurasi dan presisi tertinggi. Accuracy sebesar 92.93%, precision sebesar 95.70%, recall sebesar 89.17%.

Hasil dari penelitian ini adalah analisis sentimen pada salah satu aplikasi sosial media, yaitu youtube. Penelitian ini akan menjabarkan proses yang akan menghasilkan analisis sentimen menggunakan metode Machine Learning dengan menggunakan algoritma Support Vector Machine. Berdasarkan penelitian-penelitian sebelumnya, ditemukan bahwa metode Support Vector Machine dapat menghasilkan akurasi yang cukup tinggi dalam melakukan analisis sentimen. Tujuan dari penelitian ialah untuk mengetahui sentimen terhadap aplikasi youtube di google play store, dengan menggunakan Machine Learning dengan algoritma SVM diharapkan dapat memberikan hasil yang terbaik demi peningkatan pada aplikasi youtube di google play store.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian merupakan tahapan apa saja yang akan di lakukan pada penelitian[12]. Ada 4 tahap yang akan dilakukan dalam penelitian ini. Dapat dilihat pada gambar 1 Flowchart Tahapan Penelitian.



Gambar 1. Flowchart Tahapan Penelitian

Adapun tahapan penelitian tersebut sebagai berikut :

- a. Pengumpulan Data
Pengumpulan data dengan cara melakukan pengamatan dan pencatatan langsung pada tempat penelitian atau pihak-pihak yang terkait dalam penelitian, Penulis melakukan observasi pengamatan langsung pada Toko Ratna Sihan, Serta pengamatan permasalahan yang berhubungan dengan pelayanan dan pengolahan data.
- b. Pengembangan Sistem
Pengembangan Sistem adalah proses membuat system yang akan dibuat lalu dikembangkan agar dapat digunakan secara efektif dan berguna.
- c. Analisis Sistem
Analisis Sistem Merupakan Proses untuk mengidentifikasi system yang telah dibuat apakah sudah sesuai dengan yang diharapkan.
- d. Pengujian Sistem
Pada tahapan ini merupakan pengujian system yang akan diuji coba apakah sudah bisa digunakan dan dipakai oleh user atau pengguna.

2.2 Support Vector Machine

Metode analisa yang digunakan dalam penelitian ini adalah metode klasifikasi menggunakan Support Vector Machine[13]. Analisis sentimen juga digunakan untuk melakukan pelabelan data kedalam kelas sentimen positif dan negatif. Selain itu metode Support Vector Machine digunakan untuk melakukan prediksi, baik itu khusus klasifikasi maupun khusus regresi[14]. Pada penelitian penulis menggunakan satu nilai perbandingan dari data training dan data testing, yaitu: Data Training 90% dan testing 10%. Setelah itu metode Support Vector Machine akan melakukan pelatihan dengan cara mendeteksi ciri-ciri data pada tiap kelas yang di bentuk pada data training. hasil pelatihan pada metode Support Vector Machine kemudian di evaluasi menggunakan data testing[15]. Evaluasi menggunakan library yang terdapat pada metode Support Vector Machine yaitu Confusion matrix untuk mencari nilai akurasi, presisi, dan recall.

2.3 Confusion Matrix

Confusion Matrix merupakan teknik yang digunakan untuk mengevaluasi klasifikasi model untuk memperkirakan objek yang benar atau salah. Sebuah matriks dari prediksi akan dibandingkan dengan kelas asli yang berisi informasi aktual dan prediksi nilai klasifikasi[16]. Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Nilai True Negative (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan False Positive (FP) merupakan data negatif namun terdeteksi sebagai data positif[8]. Seperti yang terlihat pada Tabel 3 berikut ini :

Tabel 1. Confusion Matrix

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan :

TP : True Positive (Jumlah prediksi benar pada kelas sentimen positif)

FP : False Positive (Jumlah prediksi salah pada kelas sentimen positif)

FN : False Negative (Jumlah prediksi salah pada kelas sentimen negatif)

TN : True Negative (Jumlah prediksi benar pada kelas sentimen negatif)

3. HASIL DAN PEMBAHASAN

Dari penelitian ini membahas analisis sentimen aplikasi youtube di google play store menggunakan machine learning. Langkah awal yaitu pengumpulan data ulasan pada google play store[17]. Pengumpulan data dilakukan menggunakan library snsrape, dengan menentukan youtube sebagai aplikasi yang akan diambil data ulasannya, kemudian dapatlah data ulasan mulai dari bulan desember 2022 sampai desember 2023. Data ulasan yang berhasil didapatkan berjumlah 4996 data, yang kemudian dilakukan preprocessing. Pada tahap preprocessing dari 4996 data yang berhasil didapatkan menghasilkan 4993 data yang berhasil lolos dari tahap preprocessing.

3.1 Pengumpulan Data

Pengumpulan data dilakukan menggunakan library snsrape yang ada pada bahasa pemrograman python. Langkah pertama dalam pengumpulan data adalah menginstall library snsrape di google colab, dengan menggunakan source code berikut ini “!pip install -q snsrape”. Kemudian dapat dilakukan pengambilan data sesuai jumlah yang penulis inginkan. Berikut adalah gambar source code yang penulis gunakan untuk melakukan pengambilan data[18]. Pada penelitian ini

data yang diambil melalui source code tersebut berjumlah 5000 data. Namun setelah data tersebut disimpan kedalam file csv jumlahnya berubah menjadi 4996 data. Data yang didapatkan tersebut memiliki keterangan waktu dari Desember 2022 sampai Desember 2023. Data yang dikumpulkan terdapat 4 kolom yaitu username, score, at, dan content[19]. Kemudian disimpan dengan format csv. Hasil pengumpulan data dapat dilihat pada gambar 2 berikut ini:

	userName	score	at	content
2063	Aditya Syafa	1	2023-12-28 02:04:55	Terlalu banyak bug , setiap detik pasti ada bu...
1480	Wahidin Jaguar	2	2023-12-27 20:10:01	YouTube. Memang Aneh Banget. Suka Update Sendi...
27	Mii won	1	2023-12-27 18:40:58	Sumpah gk guna lama-lama ni apk. Udah nyoba no...
443	Akhtar	1	2023-12-27 14:14:59	Ada bug di YouTube tepatnya di bagian komen sa...
4109	Deva Abi123	5	2023-12-27 12:30:47	Aku suka nonton tapi yang aku suka konten bang...

Gambar 2. Hasil pengumpulan data

3.2 Preprocessing

Tahap preprocessing adalah mengubah data yang berhasil dikumpulkan menjadi format yang diperlukan. Tahap preprocessing dilakukan dengan bahasa pemrograman python yang dijalankan di google colab[20]. Adapun langkah-langkah tahap preprocessing yang dilakukan dalam penelitian ini yaitu Cleaning, Case Folding, Tokenizing, Normalisasi, Stopword Removal, Stemming

3.3 Cleaning

Cleaning yaitu pembersihan data ulasan dari kata yang tidak dibutuhkan seperti HTML, username (@username), emoticon, hastag(#), angka, url, dan lain-lain[21]. Adapun hasil dari proses Cleaning data ulasan yang sudah didapatkan dapat dilihat pada gambar 3 berikut ini :

	score	content	cleaning
0	4	Suaranya ?????? Kecil sekali kurang treble	Suaranya Kecil sekali kurang treble
1	3	Makan kuota internet aplikasinya banyak sekali...	Makan kuota internet aplikasinya banyak sekali...
2	3	Mohon maaf kami pengguna setia YouTube ingin l...	Mohon maaf kami pengguna setia YouTube ingin l...
3	2	Kenapa ya youtube ini aku baru mencet vidio la...	Kenapa ya youtube ini aku baru mencet vidio la...
4	5	fiturnya bagus, tolong tambahkan remix green s...	fiturnya bagus tolong tambahkan remix green sc...
...
4991	1	Makin kesini iklannya makin gila. Makin panjang...	Makin kesini iklannya makin gila Makin panjang...
4992	3	tolong diperbaiki bug translate nya ya, komen ...	tolong diperbaiki bug translate nya ya komen b...
4993	5	Bagus bisa melihat video tapi sayangnya ngebug...	Bagus bisa melihat video tapi sayangnya ngebug...
4994	1	YouTube main kesini makin kesana, menurutku si...	YouTube main kesini makin kesana menurutku sih...
4995	3	Iklannya banyak banget. Baru klik video yg mau...	Iklannya banyak banget Baru klik video yg mau ...

Gambar 3. Hasil proses Cleaning

3.4 Case Folding

Case Folding merupakan proses yang digunakan untuk menyeragamkan huruf yang terdapat pada ulasan, dari huruf kapital menjadi huruf kecil (Lower Case[22]). Adapun hasil proses case folding dapat dilihat pada gambar 4 berikut ini:

	score	content	cleaning	case folding
0	4	Suaranya ?????? Kecil sekali kurang treble	Suaranya Kecil sekali kurang treble	suaranya kecil sekali kurang treble
1	3	Makan kuota internet aplikasinya banyak sekali...	Makan kuota internet aplikasinya banyak sekali...	makan kuota internet aplikasinya banyak sekali...
2	3	Mohon maaf kami pengguna setia YouTube ingin l...	Mohon maaf kami pengguna setia YouTube ingin l...	mohon maaf kami pengguna setia youtube ingin l...
3	2	Kenapa ya youtube ini aku baru mencet vidio la...	Kenapa ya youtube ini aku baru mencet vidio la...	kenapa ya youtube ini aku baru mencet vidio la...
4	5	fiturnya bagus, tolong tambahkan remix green s...	fiturnya bagus tolong tambahkan remix green sc...	fiturnya bagus tolong tambahkan remix green sc...
...
4991	1	Makin kesini iklannya makin gila. Makin panjang...	Makin kesini iklannya makin gila Makin panjang...	makin kesini iklannya makin gila makin panjang ...
4992	3	tolong diperbaiki bug translate nya ya, komen ...	tolong diperbaiki bug translate nya ya komen b...	tolong diperbaiki bug translate nya ya komen b...
4993	5	Bagus bisa melihat video tapi sayangnya ngebug...	Bagus bisa melihat video tapi sayangnya ngebug...	bagus bisa melihat video tapi sayangnya ngebug...
4994	1	YouTube main kesini makin kesana, menurutku si...	YouTube main kesini makin kesana menurutku sih...	youtube main kesini makin kesana menurutku sih...
4995	3	Iklannya banyak banget. Baru klik video yg mau...	Iklannya banyak banget Baru klik video yg mau ...	iklannya banyak banget baru klik video yg mau ...

Gambar 4. Hasil proses Case Folding

3.5 Tokenizing

Tokenizing adalah proses memecahkan atau memotong kata pada dokumen menjadi term-term berdasarkan spasi. Adapun hasil dari proses tokenisasi dari data ulasan yang sudah didapatkan dapat dilihat pada gambar 5 berikut ini :

score	content	cleaning	case folding	tokenisasi
0	4	Suaranya ?????? Kecil sekali kurang treble	Suaranya Kecil sekali kurang treble	suaranya kecil sekali kurang treble [suaranya, kecil, sekali, kurang, treble]
1	3	Makan kuota internet aplikasinya banyak sekali...	makan kuota internet aplikasinya banyak sekali...	[makan, kuota, internet, aplikasinya, banyak, ...]
2	3	Mohon maaf kami pengguna setia YouTube ingin l...	mohon maaf kami pengguna setia YouTube ingin l...	[mohon, maaf, kami, pengguna, setia, youtube, ...]
3	2	Kenapa ya yuotube ini aku baru mencet vidio la...	kenapa ya yuotube ini aku baru mencet vidio la...	[kenapa, ya, yuotube, ini, aku, baru, mencet, ...]
4	5	fiturnya bagus, tolong tambahkan remix green s...	fiturnya bagus tolong tambahkan remix green sc...	[fiturnya, bagus, tolong, tambahkan, remix, gr...]
...
4991	1	Makin kesini iklannya makin gila. Makin panjang...	makin kesini iklannya makin gila Makin panjang...	[makin, kesini, iklannya, makin, gila, makin, ...]
4992	3	tolong diperbaiki bug translate nya ya, komen ...	tolong diperbaiki bug translate nya ya komen b...	[tolong, diperbaiki, bug, translate, nya, ya, ...]
4993	5	Bagus bisa melihat video tapi sayangnya ngebug...	bagus bisa melihat video tapi sayangnya ngebug...	[bagus, bisa, melihat, video, tapi, sayangnya, ...]
4994	1	YouTube main kesini makin kesana, menurutku si...	youtube main kesini makin kesana menurutku sih...	[youtube, main, kesini, makin, kesana, menurut...]
4995	3	Iklannya banyak banget. Baru klik video ya mau...	iklannya banyak banget Baru klik video yg mau ...	[iklannya, banyak, banget, baru, klik, video, ...]

Gambar 5. Hasil proses Tokenizing

3.6 Normalisasi

Pada tahap ini merupakan tahap untuk mengubah kalimat yang tidak baku atau slangword menjadi kalimat baku[23]. Adapun hasil dari proses normalisasi dari data ulasan yang sudah didapatkan dapat dilihat pada gambar 6 berikut ini :

score	content	cleaning	case folding	tokenisasi	normalisasi
0	4	Suaranya ?????? Kecil sekali kurang treble	suaranya kecil sekali kurang treble	[suaranya, kecil, sekali, kurang, treble]	[suaranya, kecil, sekali, kurang, treble]
1	3	Makan kuota internet aplikasinya banyak sekali...	makan kuota internet aplikasinya banyak sekali...	[makan, kuota, internet, aplikasinya, banyak, ...]	[makan, kuota, internet, aplikasinya, banyak, ...]
2	3	Mohon maaf kami pengguna setia YouTube ingin l...	mohon maaf kami pengguna setia youtube ingin l...	[mohon, maaf, kami, pengguna, setia, youtube, ...]	[mohon, maaf, kami, pengguna, setia, youtube, ...]
3	2	Kenapa ya yuotube ini aku baru mencet vidio la...	kenapa ya yuotube ini aku baru mencet vidio la...	[kenapa, ya, yuotube, ini, aku, baru, mencet, ...]	[kenapa, ya, yuotube, ini, saya, baru, mencet, ...]
4	5	fiturnya bagus, tolong tambahkan remix green s...	fiturnya bagus tolong tambahkan remix green sc...	[fiturnya, bagus, tolong, tambahkan, remix, gr...]	[fiturnya, bagus, tolong, tambahkan, remix, gr...]
...
4991	1	Makin kesini iklannya makin gila. Makin panjang...	makin kesini iklannya makin gila Makin panjang...	[makin, kesini, iklannya, makin, gila, makin, ...]	[makin, kesini, iklannya, makin, gila, makin, ...]
4992	3	tolong diperbaiki bug translate nya ya, komen ...	tolong diperbaiki bug translate nya ya komen b...	[tolong, diperbaiki, bug, translate, nya, ya, ...]	[tolong, diperbaiki, bug, translate, nya, ya, ...]
4993	5	Bagus bisa melihat video tapi sayangnya ngebug...	bagus bisa melihat video tapi sayangnya ngebug...	[bagus, bisa, melihat, video, tapi, sayangnya, ...]	[bagus, bisa, melihat, video, tetapi, sayangnya, ...]
4994	1	YouTube main kesini makin kesana, menurutku si...	youtube main kesini makin kesana menurutku sih...	[youtube, main, kesini, makin, kesana, menurut...]	[youtube, main, kesini, makin, kesana, menurut...]
4995	3	Iklannya banyak banget. Baru klik video ya mau...	iklannya banyak banget Baru klik video yg mau ...	[iklannya, banyak, banget, baru, klik, video, ...]	[iklannya, banyak, banget, baru, klik, video, ...]

Gambar 6. Hasil proses Normalisasi

3.7 Stopword Removal

Stopword Removal adalah proses penyaringan dan penghapusan kata yang tidak penting dari hasil tokenisasi. Pada penelitian ini penulis menggunakan stopwords bahasa Indonesia yang di dapatkan dari library NLTK yang terdapat pada bahasa pemrograman Python kemudian ditambah dengan stopwords yang penulis buat sendiri. Adapun hasil dari proses Stopword Removal data ulasan yang sudah didapatkan pada gambar 7 sebagai berikut :

score	content	cleaning	case folding	tokenisasi	normalisasi	stopword removal	
0	4	Suaranya ?????? Kecil sekali kurang treble	Suaranya Kecil sekali kurang treble	suaranya kecil sekali kurang treble	[suaranya, kecil, sekali, kurang, treble]	[suaranya, kecil, sekali, kurang, treble]	[suaranya, treble]
1	3	Makan kuota internet aplikasinya banyak sekali...	Makan kuota internet aplikasinya banyak sekali...	makan kuota internet aplikasinya banyak sekali...	[makan, kuota, internet, aplikasinya, banyak, ...]	[makan, kuota, internet, aplikasinya, banyak, ...]	[makan, kuota, internet, aplikasinya, buka, ap...]
2	3	Mohon maaf kami pengguna setia YouTube ingin l...	Mohon maaf kami pengguna setia YouTube ingin l...	mohon maaf kami pengguna setia youtube ingin l...	[mohon, maaf, kami, pengguna, setia, youtube, ...]	[mohon, maaf, kami, pengguna, setia, youtube, ...]	[mohon, maaf, pengguna, setia, youtube, lapor...]
3	2	Kenapa ya youtube ini aku baru mencet vidio la...	Kenapa ya youtube ini aku baru mencet vidio la...	kenapa ya youtube ini aku baru mencet vidio la...	[kenapa, ya, youtube, ini, aku, baru, mencet, ...]	[kenapa, ya, youtube, ini, saya, baru, mencet...]	[youtube, mencet, vidio, langsung, selesai, an...]
4	5	fiturnya bagus, tolong tambahkan remix green s...	fiturnya bagus tolong tambahkan remix green sc...	fiturnya bagus tolong tambahkan remix green sc...	[fiturnya, bagus, tolong, tambahkan, remix, gr...]	[fiturnya, bagus, tolong, tambahkan, remix, gr...]	[fiturnya, bagus, tolong, tambahkan, remix, gr...]
...
4991	1	Makin kesini iklannya makin gila. Makin panjang...	Makin kesini iklannya makin gila. Makin panjang...	makin kesini iklannya makin gila. makin panjang...	[makin, kesini, iklannya, makin, gila, makin, ...]	[makin, kesini, iklannya, makin, gila, makin, ...]	[kesini, iklannya, gila, nggak, skip, maksa, b...]
4992	3	tolong diperbaiki bug translate nya ya, komen ...	tolong diperbaiki bug translate nya ya komen b...	tolong diperbaiki bug translate nya ya komen b...	[tolong, diperbaiki, bug, translate, nya, ya, ...]	[tolong, diperbaiki, bug, translate, nya, ya, ...]	[tolong, diperbaiki, bug, translate, komen, ba...]
4993	5	Bagus bisa melihat video tapi sayangnya ngebug...	Bagus bisa melihat video tapi sayangnya ngebug...	bagus bisa melihat video tapi sayangnya ngebug...	[bagus, bisa, melihat, video, tapi, sayangnya...]	[bagus, bisa, melihat, video, tetapi, sayangnya...]	[bagus, video, sayangnya, ngebug, mendownload...]

Gambar 7. Hasil proses Stopword Removal

3.8 Stemming

Stemming adalah proses untuk menghilangkan awalan atau akhiran kata yang terdapat kata sambung, kata depan, kata ganti, menjadi kata dasar. Proses Stemming dilakukan dengan library sastrawi yang terdapat pada bahasa pemrograman python. Adapun hasil dari proses stemming data ulasan yang sudah didapatkan pada gambar 8 adalah sebagai berikut :

score	content	cleaning	case folding	tokenisasi	normalisasi	stopword removal	stemming
0	4	Suaranya ?????? Kecil sekali kurang treble	Suaranya Kecil sekali kurang treble	suaranya kecil sekali kurang treble	[suaranya, kecil, sekali, kurang, treble]	[suaranya, kecil, sekali, kurang, treble]	[suara, treble]
1	3	Makan kuota internet aplikasinya banyak sekali...	Makan kuota internet aplikasinya banyak sekali...	makan kuota internet aplikasinya banyak sekali...	[makan, kuota, internet, aplikasinya, banyak, ...]	[makan, kuota, internet, aplikasinya, banyak, ...]	[makan, kuota, internet, aplikasi, buka, aplik...]
2	3	Mohon maaf kami pengguna setia YouTube ingin l...	Mohon maaf kami pengguna setia YouTube ingin l...	mohon maaf kami pengguna setia youtube ingin l...	[mohon, maaf, kami, pengguna, setia, youtube, ...]	[mohon, maaf, kami, pengguna, setia, youtube, ...]	[mohon, maaf, guna, setia, youtube, lapor, you...]
3	2	Kenapa ya youtube ini aku baru mencet vidio la...	Kenapa ya youtube ini aku baru mencet vidio la...	kenapa ya youtube ini aku baru mencet vidio la...	[kenapa, ya, youtube, ini, aku, baru, mencet, ...]	[kenapa, ya, youtube, ini, saya, baru, mencet...]	[youtube, mencet, vidio, langsung, selesai, an...]
4	5	fiturnya bagus, tolong tambahkan remix green s...	fiturnya bagus tolong tambahkan remix green sc...	fiturnya bagus tolong tambahkan remix green sc...	[fiturnya, bagus, tolong, tambahkan, remix, gr...]	[fiturnya, bagus, tolong, tambahkan, remix, gr...]	[fiturnya, bagus, tolong, tambah, remix, green...]
...
4991	1	Makin kesini iklannya makin gila. Makin panjang...	Makin kesini iklannya makin gila. Makin panjang...	makin kesini iklannya makin gila. makin panjang...	[makin, kesini, iklannya, makin, gila, makin, ...]	[makin, kesini, iklannya, makin, gila, makin, ...]	[kesini, iklan, gila, nggak, skip, maksa, bang...]
4992	3	tolong diperbaiki bug translate nya ya, komen ...	tolong diperbaiki bug translate nya ya komen b...	tolong diperbaiki bug translate nya ya komen b...	[tolong, diperbaiki, bug, translate, nya, ya, ...]	[tolong, diperbaiki, bug, translate, nya, ya, ...]	[tolong, baik, bug, translate, komen, ba...]

Gambar 8. Hasil proses Stemming

3.9 Labeling

Proses Labelling dilakukan membagi ke dalam tiga kelas sentimen berdasarkan score rating ulasan pada google play store, yaitu jika data mendapat skor 1 atau 2 maka termasuk kelas negatif, jika mendapat skor 3 maka akan masuk ke kelas netral dan jika mendapat skor 4 dan 5, maka masuk pada kelas positif. Hasil labeling dapat dilihat pada gambar 9 dibawah ini :

score	stemming	label
0	4	['suara', 'treble'] positif
1	3	['makan', 'kuota', 'internet', 'aplikasi', 'bu...'] netral
2	3	['mohon', 'maaf', 'guna', 'setia', 'youtube', ...] netral
3	2	['youtube', 'mencet', 'vidio', 'langsung', 'se...'] negatif
4	5	['fiturnya', 'bagus', 'tolong', 'tambah', 'rem...'] positif
...
4998	1	['kesini', 'iklan', 'gila', 'nggak', 'skip', '...'] negatif
4999	3	['tolong', 'baik', 'bug', 'translate', 'komen'...] netral
4990	5	['bagus', 'video', 'sayang', 'ngebug', 'mendow...'] positif
4991	1	['youtube', 'main', 'kesini', 'kesana', 'turut...'] negatif
4992	3	['iklan', 'banget', 'klik', 'video', 'putar', ...] netral

4993 rows x 3 columns

Gambar 9. Hasil proses Labeling

3.10 Klasifikasi dan Evaluasi

Setelah dilakukan labelling data, proses selanjutnya yaitu klasifikasi dan evaluasi. Data yang sudah diberi label kemudian dibagi menjadi dua, yaitu data training dan data testing. Proses klasifikasi dan evaluasi pada penelitian ini dilakukan menggunakan metode support vector machine.

3.11 Menampilkan data ulasan positif, negatif dan netral

Setelah semua data ulasan diberi label maka didapatkan jumlah hasil dari data ulasan tersebut berdasarkan labelnya masing-masing. Adapun hasil yang didapatkan dari proses diatas dapat dilihat pada gambar 10 berikut :

```
['positif' 'netral' 'negatif']
negatif    3365
positif    1083
netral     545
Name: label, dtype: int64
```

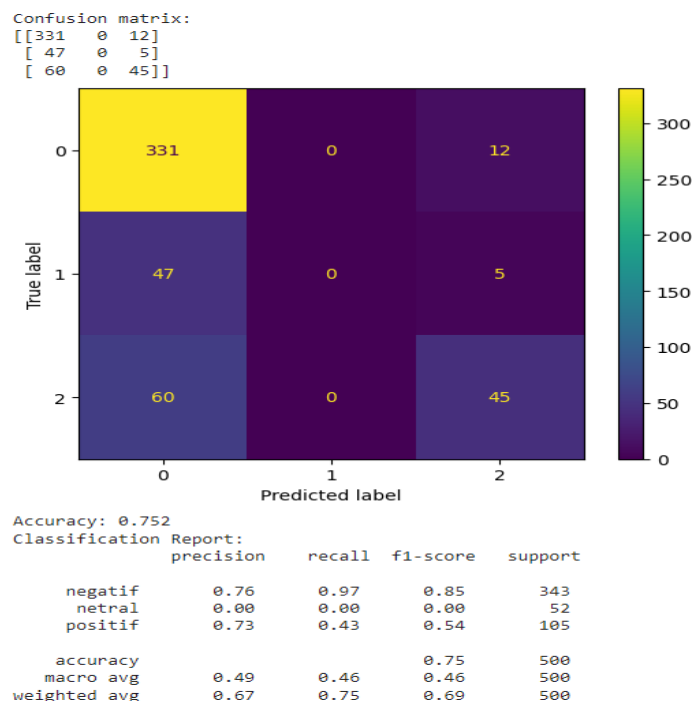
Gambar 10. Jumlah data ulasan berdasarkan label

3.12 Pembagian Data Training dan Data Testing

Langkah awal klasifikasi menggunakan metode support vector machine adalah membagi data menjadi dua yaitu data training dan data testing. Pada penelitian ini penulis menggunakan perbandingan 90% : 10%. Adapun hasil dari proses pembagian data ulasan yang sudah didapatkan hasil jumlah data training sebesar 4493 dan jumlah data testing sebanyak 500.

3.13 Klasifikasi dan Evaluasi menggunakan Support Vector Machine

Metode support vector machine akan melakukan pelatihan dengan cara mendeteksi ciri-ciri data pada tiap kelas yang dibentuk pada data training. Hasil pelatihan pada metode support vector machine kemudian dievaluasi menggunakan data testing. Proses evaluasi dilakukan menggunakan library yang terdapat pada metode support vector machine yaitu confusion matrix untuk mencari nilai akurasi, presisi, dan recall. Adapun hasil dari proses klasifikasi menggunakan metode Support vector machine dapat dilihat pada gambar 11 dibawah ini :



Gambar 11. Hasil klasifikasi

3.14 Performing K-Fold Cross Validation

Metode ini melibatkan pembagian data training menjadi K kelompok yang sama besar, atau biasa disebut "fold". Setiap fold kemudian bergiliran sebagai data validasi sementara K-1 fold yang lainnya digunakan sebagai data training. Proses ini diulang sebanyak K kali, sehingga setiap fold digunakan sebagai data validasi tepat satu kali. Pada penelitian ini akan menggunakan nilai K=10 yang berarti dataset akan dibagi menjadi 10 bagian atau fold untuk melakukan pengujian K-Fold Cross Validation sebanyak 10 kali pengulangan. Hasil pengujian berdasarkan nilai K-Fold dapat dilihat pada tabel 2 berikut :

Tabel 2. Hasil nilai K-Fold

Nilai K-Fold	Akurasi	Presisi	Recall
1	0.77	0.76	0.98
2	0.73	0.74	0.96
3	0.74	0.76	0.94
4	0.77	0.77	0.98
5	0.75	0.76	0.98
6	0.76	0.77	0.96
7	0.74	0.75	0.96
8	0.76	0.77	0.96
9	0.76	0.77	0.96
10	0.75	0.76	0.96
Rata-rata Akurasi	0.75		

4. KESIMPULAN

Kesimpulan dari penelitian yang telah dilakukan mengenai analisis sentimen aplikasi youtube di google play store menggunakan Machine Learning adalah data yang digunakan dengan memanfaatkan library snsrape, sebanyak 4996 data dan dilakukan preprocessing sehingga tersisa menjadi 4993 data yang bisa diolah. Data yang diambil merupakan data ulasan di google play store dari jangka waktu bulan desember 2022 sampai desember 2023. Hasil labelling 3 kelas data yaitu kelas positif sebanyak 1083, kelas negatif sebanyak 3365 dan netral sebanyak 545. Proses klasifikasi dan penerapan algoritma support vector machine menggunakan perbandingan data training dan data testing yang telah penulis lakukan berdasarkan data training 90% dan data testing 10% menghasilkan nilai akurasi yaitu 75% dengan nilai presisi kelas negatif 76% beserta recall kelas negatif 97%. Pengujian K-Fold Cross Validation menggunakan nilai K=10 mendapatkan nilai rata-rata Akurasi sebesar 0.75 atau 75%.

REFERENCES

- [1] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [2] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/teknol.v10i2.419.
- [3] S. A. Aaputra, Didi Rosiyadi, Windu Gata, and Syepri Maulana Husain, "Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019, doi: 10.29207/resti.v3i3.1118.
- [4] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [5] M. Hudha, E. Supriyati, and T. Listyorini, "Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 1–6, 2022, doi: 10.33387/jiko.v5i1.3376.
- [6] M. K. Khoirul Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373.
- [7] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.31539/intecom.v5i1.3708.
- [8] Karsito and S. Susanti, "Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia," *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.
- [9] Y. Septiani, E. Aribbe, and R. Diansyah, "ANALISIS KUALITAS LAYANAN SISTEM INFORMASI AKADEMIK UNIVERSITAS ABDURRAB TERHADAP KEPUASAN PENGGUNA MENGGUNAKAN METODE SEVQUAL (Studi Kasus : Mahasiswa Universitas Abdurrab Pekanbaru)," *J. Teknol. Dan Open Source*, vol. 3, no. 1, pp. 131–143, 2020, doi: 10.36378/jtos.v3i1.560.
- [10] H. Mukhtar, J. Al Amien, and M. A. Rucyat, "Filtering Spam Email menggunakan Algoritma Naïve Bayes," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 1, pp. 9–19, 2022, doi: 10.37859/coscitech.v3i1.3652.
- [11] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [12] N. Faridhotul Hidayah, K. Paranita Kartika R., and S. Nur Budiman, "Penerapan Metode Naive Bayes Dalam Analisis Sentimen Aplikasi Sentuh Tanahku Pada Google Play," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 679–683, 2022, doi: 10.36040/jati.v6i2.5610.
- [13] P. M. Nirmala Dharmapatni and N. L. P. Merawati, "Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan," *J. Bumigora Inf. Technol.*, vol. 2, no. 2, pp. 105–112, 2020, doi: 10.30812/bite.v2i2.904.
- [14] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi:

- 10.21107/edutic.v7i1.8779.
- [15] Y. Familia Nugraini, R. Rohmat Saedudin, and R. Andreswari, “Implementasi Data Mining Dalam Kasus Mental Health Pada Sosial Media Twitter Menggunakan Metode Naive Bayes Implementation of Data Mining in the Case of Mental Health on Social Media Twitter Using Naive Bayes Method,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9260–9265, 2021.
 - [16] P. P. A. Arsyia Monica Pravina, Imam Cholissodin, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” vol. 3, no. 3, pp. 2789–2797, 2019.
 - [17] R. Ardhani et al., “ANALISIS SENTIMEN TERHADAP LAYANAN APLIKASI GRAB INDONESIA MENGGUNAKAN METODE NAÏVE BAYES,” vol. 8, no. 1, pp. 303–309, 2024.
 - [18] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
 - [19] S. M. Fani, R. Santoso, and S. Suparti, “Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-Means Clustering,” *J. Gaussian*, vol. 10, no. 4, pp. 583–593, 2021, doi: 10.14710/j.gauss.v10i4.30409.
 - [20] S. N. Hakim, “ANALISIS SENTIMEN PERSEPSI PENGGUNA MYINDIHOME MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES CLASSIFIER (NBC) TUGAS,” p. 6, 2021.
 - [21] D. A. Agustina, S. Subanti, and E. Zukhronah, “Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine,” *Indones. J. Appl. Stat.*, vol. 3, no. 2, p. 109, 2021, doi: 10.13057/ijas.v3i2.44337.
 - [22] R. Risnantoyo, A. Nugroho, and K. Mandara, “Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm,” *J. Informatics Telecommun. Eng.*, vol. 4, no. 1, pp. 86–96, 2020, doi: 10.31289/jite.v4i1.3798.
 - [23] A. FATIHIN, “Analisis Sentimen Terhadap Ulasan Aplikasi Mobile Menggunakan Metode Support Vector Machine (Svm) Dan Pendekatan Lexicon Based,” p. 103, 2022.