

Studi Perbandingan: Algoritma Random Forest, Naive Bayes Dan Support Vector Machine Dalam Analisis Sentimen Pada Aplikasi Capcut Di Google Play Store

Indra Irawan¹, Wardianto², M.Hizbul Wathan³, M. Bagus Prayogi⁴,

^{1,2,4}Informatika, Universitas Nurul Huda, ³Teknologi Rekayasa Perangkat Lunak, Politeknik Manufaktur Negeri Bangka Belitung

Penulis Korepondensi : Indrairawan@unuha.ac.id¹,
wardianto@unuha.ac.id², mhizbul@polman-babel.ac.id³, mhdjesen212@gmail.com⁴

Article Information:

Received: 07 Agustus, 2024, Accepted: 11 September, 2024, Published: 17 Oktober 2024

Abstract

CapCut, a highly popular video editing tool, boasts millions of users worldwide across various age groups. Posting reviews on the Google Play Store can provide valuable insights into this application. This study aims to evaluate the effectiveness of three classification algorithms Random Forest, Naïve Bayes, and Support Vector Machine in performing sentiment analysis on Google Play Store reviews of the CapCut application. User reviews are identified and categorized into positive, negative, and neutral labels using sentiment analysis methods. A total of three thousand user review datasets were employed in this investigation. The research procedure involved data preprocessing, feature extraction, and model training. The results show that the Random Forest classification method achieved 83% accuracy, the Naïve Bayes method 70% accuracy, and the Support Vector Machine method 86% accuracy, indicating user sentiment towards the CapCut application. With an accuracy of 0.86, the SVM algorithm is found to yield the best results.

Keywords: Sentiment Analysis, Random Forest, Naïve Bayes, Support Vector Machine, Capcut

1. Pendahuluan

Di era modern yang terus berkembang ini aplikasi untuk mengedit video menjadi salah satu kunci utama untuk mewujudkan kehidupan kreatif pengguna internet. Aplikasi capcut telah menjadi salah satu platform untuk untuk pengeditan video, dengan banyak user yang telah memberikan ulasan dan tanggapan tentang pengalaman mereka ketika menggunakan aplikasi capcut. Google Play merupakan salah satu paltfrom yang disediakan pihak google dengan fasilitas berbagai produk seperti musik, buku, aplikasi, game, dan pemutar meida. Pengguna dapat mengakses goole play melalui website resmi maupun aplikasi baik di android maupun google tv dengan berbagai konten yang dapat dibeli [1].

Pada zaman sekarang banyak cara yang dapat digunakan untuk melakukan analisis sebuah aplikasi salah satu nya yaitu analisis sentiment. Analisis sentimen adalah proses komputasi otomatis yang memahami, mengekstrak, dan memproses data teks untuk mengidentifikasi sentimen yang terkandung dalam opini, pendapat, perilaku, dan emosi terhadap suatu subjek. Analisis sentiment berfungsi untuk memahami opini, sikap, emosi yang terkandung dalam teks, ulasan, tanggapan, postingan maupun sebuah komentar. Pada penelitian ini menggunakan algoritma klasifikasi random forest, naïve bayes, dan support vector machine untuk membandingkan analisis sentiment berupa ulasan aplikasi capcut yang didapat dari google play, dengan berbagai kelebihan dan kekurangan nya. Algoritma Random Forest termasuk dalam kategori ensemble learning. Ensemble learning menggabungkan beberapa model pembelajaran untuk menghasilkan prediksi yang lebih akurat dan stabil dibandingkan dengan model tunggal [2]. Algoritma naïve bayes bekerja dengan cara membuat visualisasi bahwa variable-variabel yang ada dalam data bersifat independent, Algoritma ini berjalan cepat dan efisien, dengan akurasi yang dipengaruhi oleh korelasi antar variabel[3]. Algoritma Support Vector Machine merupakan algoritma yang bekerja dengan cara mencari hyperplane yang dapat memisahkan dua kelas data, dengan menggunakan algoritma SVM akan mendapatkan akurasi prediksi yang tinggi tetapi memerlukan waktu yang sangat lama[4].

Pada penelitian sebelumnya yang telah mengkaji berbagai analisis sentiment dengan menggunakan berbagai algoritma, seperti pada penelitian yang ditulis oleh VF. Anindya dan Andri Wijaya pada tahun 2023 dengan judul “Analisis Sentimen Ulasan Aplikasi Capcut Menggunakan Metode Naive Bayes”. Hasilnya menunjukkan efektivitas metode Naïve Bayes dalam mengidentifikasi sentimen dengan tingkat akurasi yang signifikan. Klasifikasi sentimen ini memberikan pandangan berharga bagi pengembang CapCut untuk meningkatkan kualitas aplikasi [1]. Kemudian penelitian yang ditulis oleh Ayu Sagita, Ahmad Faqih, Gifthera Dwilestari, Bambang Siswoyo, Denni Pratama pada tahun 2023 yang berjudul Penerapan Metode Random Forest Dalam Menganalisis Sentimen Pengguna Aplikasi Capcut Di Google Play Store menunjukan hasil bahwa bahwa metode Random Forest memberikan tingkat akurasi yang tinggi dalam menganalisis sentimen pengguna CapCut, dengan akurasi sebesar 86%, presisi 89%, recall 81%, dan f1-score 85% [5].

Penelitian ini bertujuan untuk membandingkan efektivitas Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine dalam menganalisis sentiment terhadap ulasan aplikasi Capcut di Google Play Store. Memberikan nilai performa dan akurasi dengan masing-masing algoritma dalam mengidentifikasi tanggapan positif, negatif dan netral yang terkait dengan penggunaan aplikasi Capcut. Dengan menggunakan analisis sentiment pengguna dapat mengetahui bagaimana ulasan mengenai capcut pada google play store sehingga pengguna lebih bijaksana dalam menggunakan aplikasi tersebut.

2. Tinjauan Literatur¹

2.1 Data Mining

Data mining merupakan sebuah ilmu yang menggabungkan berbagai teknik dari machine learning, pengenalan pola, statistik, manajemen basis data, dan visualisasi. Tujuan utama dari data mining adalah untuk menggali informasi berharga dari kumpulan data besar yang terdapat

dalam database.[6]

2.2 Analisis Sentimen

Analisis sentimen merupakan sebuah studi yang menggunakan komputasi untuk memahami pendapat, perasaan, dan emosi (positif dan negatif) orang-orang. Hal ini dilakukan dengan menganalisis atribut dan entitas yang mereka ungkapkan secara tekstual. Analisis sentimen menghasilkan kalimat opini yang terbentuk dari proses ekstraksi dan pengolahan data tekstual secara otomatis. Secara keseluruhan, analisis sentimen dapat membantu organisasi dalam berbagai aspek, seperti memahami publik, meningkatkan kepuasan pelanggan, meningkatkan penjualan dan pemasaran, memantau media sosial, meningkatkan penelitian dan pengembangan, serta meningkatkan keamanan dan intelijen[2].

2.3 Algoritma Random Forest

Algoritma Random Forest adalah sebuah metode pembelajaran yang menggabungkan banyak pohon keputusan sebagai pembuat klasifikasi dasar. Algoritma Random Forest merupakan pengembangan dari Algoritme Decision Tree. Setiap pohon dalam Random Forest dilatih dengan sampel individu, menghasilkan banyak pohon seperti hutan. Seiring bertambahnya data, pohon-pohon ini pun berkembang. Algoritme ini membuat pohon klasifikasi dan regresi dengan memisahkan nodenya berdasarkan fungsi yang dioptimalkan untuk meminimalkan squared-error loss [7].

2.4 Algoritma Naïve Bayes

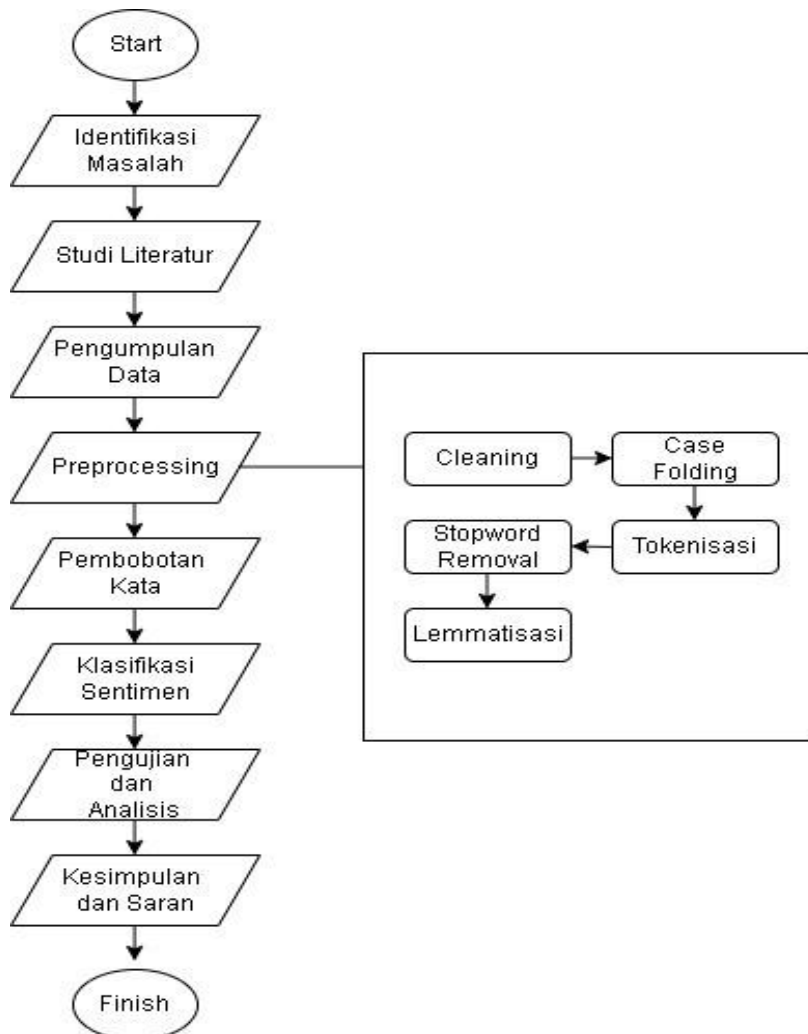
Algoritma Naïve Bayes menggunakan metode klasifikasi statistik yang mampu memperkirakan kemungkinan suatu sampel dengan memasukan sampel kedalam kelas tertentu. Metode ini akan digunakan untuk mengklasifikasikan data aplikasi capcut yang didapatkan dari ulasan play store menjadi kelas positif dan negatif. Algoritma Naive Bayes adalah algoritma yang sederhana dan mudah dipahami. Algoritma ini juga cukup cepat dan efisien dalam memproses data. Meskipun algoritma Naive Bayes memiliki beberapa kekurangan, seperti asumsi independensi antar fitur, algoritma ini tetap menjadi algoritma yang populer dan banyak digunakan dalam berbagai aplikasi[8].

2.5 Algoritma Support Vector Machine

Algoritma Support Vector Machines (SVM) merupakan metode klasifikasi dan regresi yang populer untuk menangani masalah linear dan nonlinear. Keunggulan SVM terletak pada kemampuannya untuk memisahkan data non-linear berdimensi tinggi secara linear dengan menggunakan fungsi kernel. Efektivitas SVM sangat bergantung pada jenis fungsi kernel yang dipilih dan disesuaikan dengan karakteristik data. Para peneliti yang mengatakan bahwa Algoritma Support Vector machine menjadi metode yang paling akurat dalam hal teks klasifikasi meskipun waktu yang dibutuhkan lebih lama daripada Algoritma lainnya[9].

3. Metode Penelitian

Penelitian ini menggunakan perbandingan algoritma Random Forest, Naïve Bayes, dan Support Vector Machine dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi CapCut di Google Play Store. Terdapat 9 tahapan penelitian, yaitu dimulai dari identifikasi masalah, studi literatur, pengumpulan data, pelabelan, preprocessing, pembobotan kata, klasifikasi sentimen, pengujian dan analisis, serta penarikan kesimpulan dan saran. Berikut diagram alur penelitian dapat dilihat pada Gambar 1



Gambar 1. Tahapan Metode Penelitian

3.1 Identifikasi masalah

Pada tahap ini peneliti akan mengidentifikasi permasalahan yang didapatkan dengan melihat ulasan mengenai aplikasi capcut yang ada di google play store. Berupa tanggapan negatif mengenai aplikasi capcut yang dijadikan sebagai pokok permasalahan.

3.2 Studi Literatur

Pada tahapan ini penelitian akan mengkaji berbagai macam jurnal guna sebagai rujukan untuk bahan penelitian. Jurnal yang dikaji harus memiliki hubungan atau kedekatan mengenai penelitian yang akan dibuat.

3.3 Pengumpulan Data

Data yang didapat dari hasil scraping dengan menggunakan Bahasa pemrograman python, dengan jumlah data yang ambil yaitu sebanyak 3000 data dari komentar di google play store,

kemudian di unduh dan disimpan dalam format CSV.

3.4 Pelabelan

Pada proses pelabelan akan dilakukan dengan dasar nilai rating dan juga ulasan yang diberikan oleh pengguna aplikasi capcut. Pelabelan terdiri dari 3 buah kelas yaitu positif, negatif, dan netral.

3.5 Processing Data

Tahapan ini paling penting dalam proses mining karena dalam proses ini akan menentukan bagaimana data yang dihasilkan. Apakah data tersebut berkualitas atau justru sampah. Proses yang dilakukan dalam tahap ini yaitu cleaning, case folding, tokenizing, stopwords removal, dan lemmatization.

3.6 Pembobotan Kata

Pada tahap ini metode yang digunakan dalam pembobotan kata yaitu *TF-IDF* (Term Frequency – Inverse Document Frequency). Dimana frekuensi yang muncul dalam setiap ulasan akan menjadi sumber dalam pembobotan.

3.7 Klasifikasi Sentimen

Pada tahap ini dilakukan sebuah analisis untuk menentukan kategori sentiment dari setiap data ulasan yang didapat. Peneliti menggunakan 3 buah Algoritma klasifikasi yaitu Random Forest, Naïve Bayes, dan Support Vector Machine sebagai Algoritma perbandingan hasil.

3.8 Evaluasi Model

Pada tahapan ini ketiga buah Algoritma akan di evaluasi kinerjanya, hal ini untuk menentukan accuracy dan AUC setiap Algoritma guna membandingkan Algoritma yang terbaik.

3.9 Visualisasi

Visualisasi yang berfungsi untuk merepresentasikan hasil temuan setiap Algoritma dalam penelitian. Visualisasi yang digunakan disini yaitu matrix, grafik, table, dan diagram.

4. Hasil dan Pembahasan

4.1 Pemodelan

Scrapping

Proses pengumpulan data dari ulasan aplikasi capcut dengan menggunakan google colab. Data yang diperoleh dapat berupa username, userimage, reviewId, content, score dan lain sebagainya. Data tersebut bukan berarti data bersih, namun hanya berupa content yang akan di olah terlebih dahulu.

	score	content
2377	1	Banyak banget bug nya Padahal udah 4thn makek ...
2002	1	Cap cut udah gak asik lagi udah ada iklan"nya
2131	1	Stuk dimemuat efek biasanya hitungan detik lan...
244	3	Cap cut emng baguss banget tapi sekarang semen...
1827	1	kalo nge ekspor harus nonton iklan dulu, tp ik...

Gambar 2. Scrapping

Labelling

Data akan dikelompokan dengan memberikan label berupa kelas dari setiap ulasan yang di berikan. Pada penelitian ini menggunakan 3 buah kelas sebagai labelling yaitu positif, negatif, dan nertal.

	score	content	label
2377	1	Banyak banget bug nya Padahal udah 4thn makek ...	-1
2002	1	Cap cut udah gak asik lagi udah ada iklan"nya	-1
2131	1	Stuk dimemuat efek biasanya hitungan detik lan...	-1
244	3	Cap cut emng baguss banget tapi sekarang semen...	0
1827	1	kalo nge ekspor harus nonton iklan dulu, tp ik...	-1
1107	3	aplikasi sudah bagus sih tpi ada kekurangan ke...	0
290	3	Ini kenapa ya kok setiap di buka aplikasi.a ka...	0
1801	5	Baguss capcut bisa ngebantu saya banget trs ju...	1
1460	1	Buat pihak capcut tolong di fix bug nya ya, ka...	-1
1488	1	Tiap mo ngedit vidio buat tugas sekolah selalu...	-1
2396	5	Sukaaaa, foto jadi lebih bermakna dengan adany...	1
1947	1	Bis di update setiap mau ekspor harus nonton i...	-1
1951	2	Capcut, mohon benerin appnya karena aku gak bi...	-1
2646	5	Saya sangat menyukai aplikasi ini karena saya ...	1
1341	2	bagus sih bagus, tpi tiap mau ngekspor pasti k...	-1

Gambar 3. Labelling

Data Cleaning

Menjabarkan Langkah-langkah proses pembersihan data yang didapatkan dari hasil scrapping yang berguna untuk pemodelan dan pengolahan lebih lanjut. Pembersihan dapat meliputi penghapusan data yang terdapat duplikasi, penggunaan emoji dan karakter special, penggunaan angka, kata berulang atau Salinan, dan spasi berlebih. Pada tahap ini akan menjadi tahap yang sangat penting karena jika data yang dibersihkan masih kotor hal itu akan menjadikan hasil analisis menjadi bias dan tidak akurat.

Tokenizing

Proses pemisahan kalimat/teks kedalam bentuk kata, frasa atau elemen lain yang lebih fundamental yang dari data yang akan di uji. Proses tokenizing berfungsi untuk memberi kemudahan dalam melakukan analisis dengan menjabarkan dan mengelompokkan unit dasar dalam text.

score	content	label	text_token
0	1 banyak banget bug nya padahal udah 4thn makek ...	-1	[banyak, banget, bug, nya, padahal, udah, 4thn...
1	1 cap cut udah gak asik lagi udah ada iklan nya	-1	[cap, cut, udah, gak, asik, lagi, udah, ada, i...
2	1 stuk dimemuat efek biasanya hitungan detik lan...	-1	[stuk, dimemuat, efek, biasanya, hitungan, det...
3	3 cap cut emng baguss banget tapi sekarang semen...	0	[cap, cut, emng, baguss, banget, tapi, sekaran...
4	1 kalo nge ekspor harus nonton iklan dulu tp ik...	-1	[kalo, nge, ekspor, harus, nonton, iklan, dulu...

Gambar 4. Tokenizing

Stopword Removal

Proses penghapusan kata-kata umum yang dianggap tidak membawa informasi dalam text analisis. Kata-kata yang dihapus biasanya berupa kata penghubung, atau kata imbuhan.

score	content	label	text_token
0	1 banyak banget bug nya padahal udah 4thn makek ...	-1	[banget, bug, nya, udah, 4thn, makek, apk, gak...
1	1 cap cut udah gak asik lagi udah ada iklan nya	-1	[cap, cut, udah, gak, asik, udah, iklan, nya]
2	1 stuk dimemuat efek biasanya hitungan detik lan...	-1	[stuk, dimemuat, efek, hitungan, detik, langsung]

Gambar 5. Stopword Removal

Stemming

Proses untuk menyamakan kata-kata untuk menjadikan data latih menjadi lebih sedikit, hal ini dapat mengurangi beban analisis text. Stemming membantu mencari akar dari setiap kata dengan mengabaikan keanekaragaman sehingga kata yang memiliki kemiripan akan dianggap sebagai sama.

score		content	label	text_token	steamed
0	1	banyak banget bug nya padahal udah 4thn makek ...	-1	[banget, bug, nya, udah, 4thn, makek, apk, gak...]	[banget, bug, nya, udah, 4thn, makek, apk, gak...]
1	1	cap cut udah gak asik lagi udah ada iklan nya	-1	[cap, cut, udah, gak, asik, udah, iklan, nya]	[cap, cut, udah, gak, asik, udah, iklan, nya]
2	1	stuk dimemuat efek biasanya hitungan detik lan...	-1	[stuk, dimemuat, efek, hitungan, detik, langsung]	[stuk, muat, efek, hitung, detik, langsung]
3	3	cap cut emng baguss banget tapi sekarang semen...	0	[cap, cut, emng, baguss, banget, semenjak, dah...]	[cap, cut, emng, baguss, banget, semenjak, dah...]
4	1	kalo nge ekspor harus nonton iklan dulu tp ik...	-1	[kalo, nge, ekspor, nonton, iklan, tp, iklanny...]	[kalo, nge, ekspor, nonton, iklan, tp, iklan, ga]

Gambar 6. Sebelum Steaming

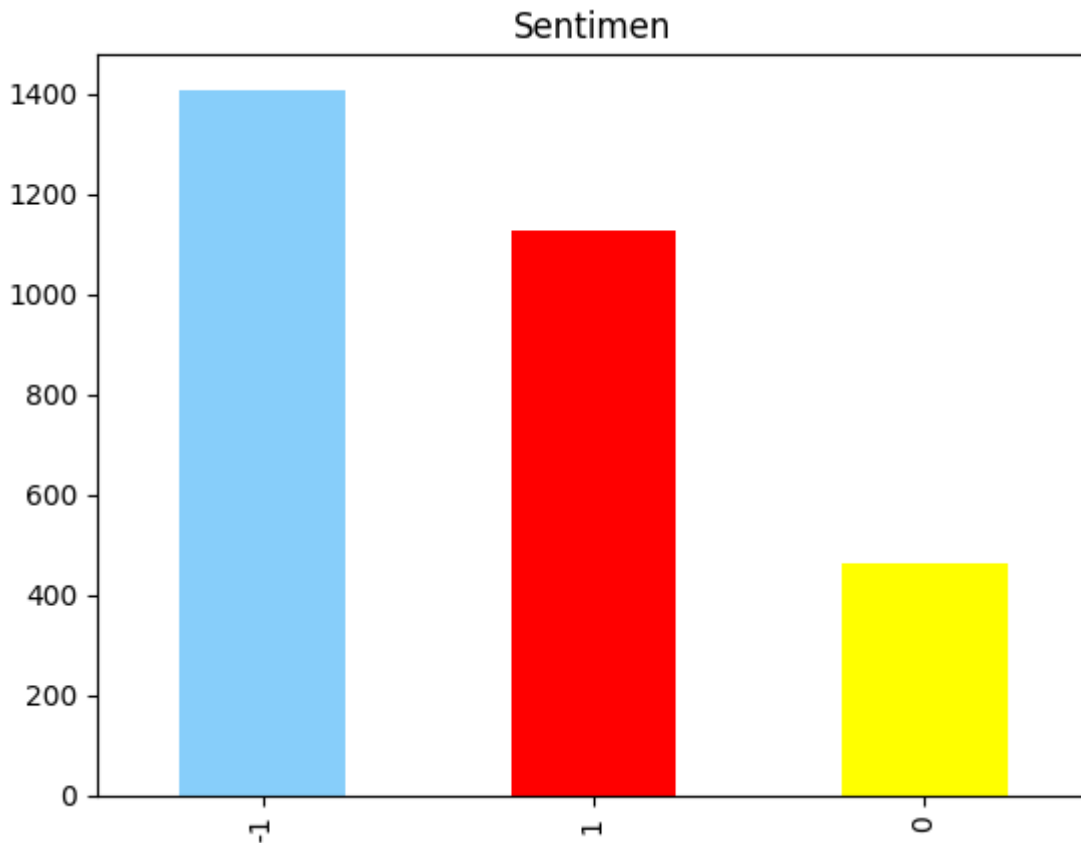


	score	content	label	text_token	steamed	text_string
0	1	banyak banget bug nya padahal udah 4thn makek ...	-1	[banget, bug, nya, udah, 4thn, makek, apk, gak...]	[banget, bug, nya, udah, 4thn, makek, apk, gak...]	banget bug nya udah 4thn makek apk gak baik bu...
1	1	cap cut udah gak asik lagi udah ada iklan nya	-1	[cap, cut, udah, gak, asik, udah, iklan, nya]	[cap, cut, udah, gak, asik, udah, iklan, nya]	cap cut udah gak asik udah iklan nya
2	1	stuk dimemuat efek biasanya hitungan detik lan...	-1	[stuk, dimemuat, efek, hitungan, detik, langsung]	[stuk, muat, efek, hitung, detik, langsung]	stuk muat efek hitung detik langsung
3	3	cap cut emng baguss banget tapi sekarang semen...	0	[cap, cut, emng, baguss, banget, semenjak, dah...]	[cap, cut, emng, baguss, banget, semenjak, dah...]	cap cut emng baguss banget semenjak dah update...
4	1	kalo nge ekspor harus nonton iklan dulu tp ik...	-1	[kalo, nge, ekspor, nonton, iklan, tp, iklanny...]	[kalo, nge, ekspor, nonton, iklan, tp, iklan, ga]	kalo nge ekspor nonton iklan iklan

Gambar 7. Setelah Steaming

Visualisasi Label

Pada visualisasi label akan menjelaskan tentang perbandingan data hasil scrapping yang terdiri dari 3 buah kelas label yaitu positif, negatif, dan netral.



Gambar 8. Visualisasi Label

Wordcloud Sentimen

Wordcloud adalah representasi visual dari kata-kata yang paling sering muncul dalam sebuah teks. Ini merupakan alat yang populer dalam analisis teks dan membantu dalam memahami pola serta tren yang ada dalam kumpulan kata-kata [11]. Wordcloud memvisualisasikan kata-kata berdasarkan kategori sentimennya, memperlihatkan kata-kata yang sering digunakan dalam masing-masing sentimen. Ukuran font dalam wordcloud menandakan frekuensi kemunculan kata tersebut, di mana semakin besar ukuran font menunjukkan frekuensi yang lebih tinggi.





Page 199

Naive Bayes					
	precision	recall	f1-score	support	
-1	0.73	0.60	0.66	141	
0	0.59	0.84	0.70	143	
1	0.86	0.65	0.74	139	
accuracy			0.70	423	
macro avg	0.73	0.70	0.70	423	
weighted avg	0.73	0.70	0.70	423	

SVM					
	precision	recall	f1-score	support	
-1	0.77	0.91	0.83	141	
0	0.96	0.84	0.90	143	
1	0.89	0.83	0.86	139	
accuracy			0.86	423	
macro avg	0.87	0.86	0.86	423	
weighted avg	0.87	0.86	0.86	423	

5. Kesimpulan

Berdasarkan hasil analisis sentimen pengguna aplikasi Capcut di Google Play Store, yang dievaluasi menggunakan metrik akurasi, recall, dan skor f1, kami membandingkan kinerja tiga algoritma pembelajaran mesin (Random Forest, Naive Bayes, dan SVM). Hasil menunjukkan bahwa SVM mencapai kinerja terbaik dengan akurasi tertinggi sebesar 0,86, diikuti oleh Random Forest dengan akurasi 0,83, dan Naive Bayes dengan akurasi terendah sebesar 0,70. Dari hasil perhitungan metrik evaluasi tersebut, dapat disimpulkan bahwa SVM adalah pilihan terbaik di antara ketiga algoritma tersebut.

Referensi

- [1] A. W. VF. Anindya, "Analisis Sentimen Ulasan Aplikasi Capcut Menggunakan Metode Naive Bayes," *Statistika*, vol. 23, no. 2, hal. 164–175, 2023, doi: 10.29313/statistika.v23i2.2353.
- [2] C. G. Indrayanto, D. E. Ratnawati, dan B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, hal. 1131–1139, 2023.
- [3] W. Yulita, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *J. Data Min. dan Sist. Inf.*, vol. 2, no.

- 2, hal. 1, 2021, doi: 10.33365/jdmsi.v2i2.1344.
- [4] R. Wahyudi dan G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, hal. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [5] A. Sagita dkk., “PENERAPAN METODE RANDOM FOREST DALAM MENGANALISIS SENTIMEN PENGGUNA APLIKASI CAPCUT DI GOOGLE PLAY STORE,” vol. 7, no. 6, hal. 3307–3313, 2023.
- [6] I. Irawan, M. R. Qisthiano, M. Syahril, dan P. M. Jakak, “Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO,” *J. Pengemb. Sist. Inf. dan Inform.*, vol. 4, no. 4, hal. 26–35, 2023, doi: 10.47747/jpsii.v4i4.1374.
- [7] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, dan F. Syafria, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, hal. 122–130, 2022.
- [8] B. Mas Pintoko dan K. Muslim, “Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier,” *e-Proceeding Eng.*, vol. 5, no. 3, hal. 8121–8130, 2018.
- [9] E. Indrayuni, “Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film,” *J. Pilar Nusa Mandiri*, vol. 14, no. 2, hal. 175, 2018, doi: 10.33480/pilar.v14i2.918.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)

