

Klasifikasi Sentimen Ulasan Aplikasi Spotify Di *Google Play Store* Menggunakan Algoritma C4.5

Renata Elfrin S Wau^{1*}, Fatmawati²

Sistem Informasi, Universitas Bina Sarana Informatika
Jakarta, Indonesia

e-mail: renataelfrinwau@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna aplikasi Spotify berdasarkan teks ulasan di *Google Play Store* dengan menggunakan algoritma C4.5. Sebanyak 298 ulasan dibagi secara seimbang antara sentimen positif dan negatif, yang sebelumnya telah diberi label berdasarkan rating ulasan. Proses pra-pemrosesan dilakukan melalui tahapan *cleaning*, *case folding*, *tokenizing*, *stemming*, dan *stopword removal* dengan bantuan aplikasi RapidMiner. Dengan menggunakan algoritma *decision tree* C4.5, model klasifikasi dibangun dan diuji berdasarkan metode *cross validation*. Hasil akhir penelitian menunjukkan bahwa model mampu mengklasifikasikan data dengan akurasi 81,56%, *precision* 90,92%, *recall* 71,03%, *F1-score* 79,03%, dan *AUC* sebesar 0,732. Temuan ini mengindikasikan bahwa C4.5 efektif dalam membedakan antara sentimen positif dan negatif pada teks ulasan berbahasa Indonesia. Penelitian ini memberikan kontribusi berupa evaluasi kinerja algoritma C4.5 pada analisis sentimen ulasan aplikasi digital, sehingga dapat menjadi dasar dalam memantau tingkat kepuasan pengguna serta mendukung pengambilan keputusan berbasis data ulasan sebuah aplikasi dan dapat menjadi referensi dalam pemilihan metode klasifikasi untuk data teks.

Katakunci: klasifikasi sentimen, c4.5, spotify, *google play store*, ulasan pengguna

ABSTRACT

This research aims to classify user sentiment toward the Spotify application based on review text from the Google Play Store using the C4.5 algorithm. A total of 298 reviews were evenly divided between positive and negative sentiment, which had been previously labeled based on the review ratings. The preprocessing stage is carried out thru cleaning, case folding, tokenizing, stemming, and stopwords removal steps with the help of the RapidMiner application. Using the C4.5 decision tree algorithm, a classification model is built and tested based on the cross-validation method. The final results of the study show that the model is able to classify data with an accuracy of 81.56%, precision of 90.92%, recall of 71.03%, F1-score of 79.03%, and an AUC of 0.732. This finding indicates that C4.5 is quite effective in distinguishing between positive and negative sentiment in Indonesian-language review texts. This study provides an empirical evaluation of the C4.5 algorithm for sentiment analysis of digital application reviews. The results may serve as a foundation for monitoring user satisfaction and supporting data-driven decision-making based on user reviews, as well as a reference for selecting suitable classification methods for textual datasets.

Keywords: sentiment classification, c4.5, spotify, *google play store*, user reviews

1. PENDAHULUAN

Musik adalah bentuk seni yang bersifat universal (Setyoko et al., 2021). Cara Masyarakat dalam menikmati musik telah mengalami perubahan yang cukup signifikan seiring berkembangnya teknologi digital. Musik telah menjadi bagian tak terpisahkan dari kehidupan manusia selama ribuan tahun, berfungsi tidak hanya sebagai hiburan, tetapi juga sebagai sarana ekspresi, komunikasi, dan pelestarian budaya. Dengan kemajuan teknologi digital, cara masyarakat menikmati musik telah mengalami transformasi besar. Salah satu platform yang paling berpengaruh dalam revolusi ini adalah Spotify (Yolanda et al., 2024).



Pada tahun 2008, Daniel Ek berhasil meluncurkan Spotify di Swedia. Ini merupakan langkah awal dalam memberikan platform layanan musik digital (Fadryona, 2021). Spotify kini menjadi salah satu platform streaming musik terpopuler di kalangan pengguna, dengan menyediakan akses ke jutaan lagu, podcast, dan konten video dari berbagai musisi serta kreator global (Marlia et al., 2024). Layanan ini memiliki antarmuka yang mudah digunakan dan berbagai fitur inovatif yang memudahkan pengguna mencari musik baru, menyusun playlist, serta membagikan pengalaman mendengarkan kepada orang lain.

Dengan jumlah pengguna aktif yang terus bertambah, Spotify menjadi salah satu aplikasi yang banyak diulas di Google Play Store oleh pengguna. Ulasan-ulasan tersebut mencerminkan pengalaman pengguna, kepuasan, maupun berbagai keluhan yang berulang, seperti iklan yang mengganggu pada versi gratis, kualitas rekomendasi lagu yang kurang akurat, keterbatasan fitur dalam mode gratis, serta bug teknis yang sering terjadi (As'ari et al., 2025). Informasi yang terkandung dalam ulasan tersebut sangat berharga bagi pengembang aplikasi dalam meningkatkan kualitas layanan. Namun, ulasan pengguna yang tersedia dalam bentuk teks bebas sering kali sulit dianalisis secara manual, terutama jika jumlahnya sangat banyak. Oleh karena itu, diperlukan metode yang tepat untuk mengelompokkan atau mengklasifikasikan ulasan tersebut berdasarkan sentimen, seperti sentimen positif atau negatif. Proses ini dikenal dengan istilah klasifikasi sentimen. Klasifikasi sentimen adalah proses untuk mengidentifikasi dan mengelompokkan sentimen atau opini bentuk teks ke dalam kategori positif, negatif, maupun netral (Fitriana et al., 2024). Penelitian ini menggunakan algoritma C4.5 sebagai metode klasifikasi. J. Ross Quinlan mengembangkan algoritma ini sebagai penyempurnaan dari algoritma ID3 (Faizah & Jananto, 2021). C4.5 termasuk metode yang banyak digunakan dalam praktik data mining, terutama di berbagai sektor industri, dirancang untuk membangun model klasifikasi berbasis pohon keputusan (Swastika et al., 2023, hlm. 49). Dalam konteks data mining, proses klasifikasi merupakan salah satu teknik utama yang digunakan untuk mengekstraksi informasi atau pola tersembunyi dari kumpulan data yang besar dan kompleks (Rahayu et al., 2024, hlm. 02). Keunggulan algoritma C4.5 dengan teknik pohon keputusan ini, misalnya dapat menangani nilai atribut yang hilang, dapat memproses data diskrit dan numerik (kontinu) yang menghasilkan aturan untuk diinterpretasikan dengan mudah, dan memiliki kemampuan paling cepat di antara algoritma lain yang menggunakan memori utama pada komputer (Afrianto et al., 2025). Dalam lingkup analisis sentimen terhadap aplikasi Spotify, penerapan C4.5 berpotensi besar untuk mengelompokkan opini pengguna, mengidentifikasi kecenderungan atau preferensi mereka, serta membantu pengembang dalam meningkatkan kualitas layanan. Sebagai bagian dari algoritma *decision tree*.

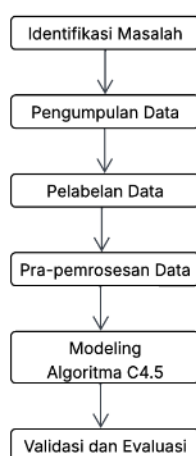
Terkait dengan analisis sentimen pada ulasan aplikasi Spotify telah diteliti oleh beberapa peneliti sebelumnya dengan menggunakan berbagai metode algoritma. Penelitian "Analisis Sentimen Terhadap Pemutar Musik Online Spotify dengan Algoritma Naive Bayes dan Support Vector Machine" menunjukkan bahwa kedua algoritma tersebut memiliki tingkat akurasi yang cukup baik, dengan SVM mencapai akurasi sebesar 82,42% dan Naive Bayes sedikit lebih unggul dengan akurasi sebesar 84,73% (Ginabila & Fauzi, 2023). Penelitian dari Rosmawati dkk "Penerapan Algoritma Random Forest untuk Menganalisis Ulasan Aplikasi Spotify pada Google Play Store" (Rosmawati et al., 2024). Model yang dihasilkan dievaluasi menggunakan *confusion matrix* dengan hasil akurasi sebesar 88,4% untuk data berbahasa Indonesia dan 93,6% untuk data berbahasa Inggris. Penelitian lainnya yaitu "Analisis Sentimen Aplikasi Spotify Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Support Vector Machine" dengan hasil pengujian akurasi sebesar 85%, precision 86%, recall 92% dan f1-score 89% (Wulandari & Sunardi, 2024). Terdapat pula penelitian yang mengkaji klasifikasi sentimen namun pada aplikasi yang berbeda. Salah satunya adalah "Analisis Sentimen Pengguna Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45" (Rahmaddeni et al., 2024). Ketiga penelitian tersebut menunjukkan bahwa berbagai algoritma klasifikasi seperti *Naive Bayes*, *SVM*, dan *Random Forest* telah banyak diterapkan dalam analisis sentimen terhadap ulasan aplikasi Spotify. Namun, belum ditemukan adanya penelitian yang secara khusus menerapkan algoritma C4.5 dalam lingkup yang sama, sehingga membuka peluang untuk mengeksplorasi kinerja dan potensi algoritma tersebut dalam penelitian ini. Penggunaannya secara spesifik untuk klasifikasi sentimen pada ulasan Spotify masih relatif terbatas. Oleh karena itu, penelitian ini menjadi kesempatan untuk mengkaji lebih dalam efektivitas algoritma C4.5 dalam mengolah dan mengategorikan sentimen pengguna secara otomatis berdasarkan data ulasan dari *Google Play Store*.

Dengan mengimplementasikan algoritma C4.5, diharapkan dapat diketahui seberapa baik kinerjanya dalam membuktikan bahwa metode pohon keputusan mampu mengidentifikasi pola-pola

dalam data teks ulasan pengguna dan menghasilkan model klasifikasi yang akurat serta mudah dipahami. Dari hasil penelitian diharapkan dapat menjadi referensi dasar pertimbangan bagi penelitian-penelitian selanjutnya dan juga dapat dimanfaatkan sebagai dasar dalam memantau tingkat kepuasan pengguna dan mendukung pengambilan keputusan berbasis data ulasan.

2. METODE PENELITIAN

Metode penelitian ini adalah metode kuantitatif dengan pendekatan data mining, khususnya pada teknik klasifikasi teks menggunakan algoritma C4.5. Rangkain tahapan-tahapan terstruktur dan terukur guna mencapai tujuan penelitian terlihat pada gambar 1 yang menunjukkan alur proses mulai dari pengumpulan data, pelabelan, pra-pemrosesan, pemodelan, hingga evaluasi. Setiap tahapan dilakukan secara berurutan untuk memastikan kualitas data dan akurasi model yang dihasilkan.



Gambar 1. Tahapan Penelitian

Data penelitian bersumber dari ulasan pengguna aplikasi Spotify pada Google Play Store. Pengumpulan data dilakukan menggunakan teknik web scraping dengan pendekatan web crawling untuk mengambil teks ulasan, rating, tanggal ulasan, dan informasi pendukung lainnya. Proses crawling dilakukan secara otomatis untuk menelusuri halaman ulasan dan menyimpannya dalam format .xlsx agar dapat diproses lebih lanjut (Talisman et al., 2024). Tahap pelabelan dilakukan secara semi-manual dengan mengacu pada rating yang diberikan pengguna. Ulasan dengan rating 4 dan 5 dikategorikan sebagai sentimen positif, sedangkan rating 1 dan 2 dikategorikan sebagai sentimen negatif. Ulasan dengan rating 3 tidak digunakan dalam penelitian untuk menghindari ambiguitas sentimen. Setelah proses kategorisasi berdasarkan rating, dilakukan pengecekan manual terhadap beberapa sampel data untuk memastikan kesesuaian antara isi teks dan label yang diberikan, sehingga meminimalkan potensi bias pelabelan.

Tahap berikutnya adalah pra-pemrosesan data untuk mengubah teks mentah menjadi data yang siap diproses oleh algoritma klasifikasi. Proses ini meliputi pembersihan karakter yang tidak relevan seperti tanda baca, emoji, angka, dan URL, kemudian seluruh teks diubah menjadi huruf kecil untuk menyeragamkan format. Selanjutnya dilakukan tokenisasi untuk memecah teks menjadi kata-kata tunggal, diikuti proses *stemming* untuk mengubah kata berimbuhan menjadi bentuk dasar. Tahap terakhir adalah penghapusan *stopword* untuk menghilangkan kata-kata yang tidak memiliki makna signifikan dalam analisis sentimen. Hasil pra-pemrosesan kemudian dikonversi menjadi representasi numerik menggunakan pembobotan term sebelum masuk ke tahap pemodelan.

Proses pemodelan dilakukan menggunakan algoritma C4.5 yang diimplementasikan melalui perangkat lunak RapidMiner. Dataset dibagi menjadi data latih dan data uji dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian. Selain itu, validasi model juga dilakukan menggunakan teknik *cross validation* untuk memastikan stabilitas performa model. Parameter default pada operator *Decision Tree* digunakan dengan kriteria pemilihan atribut berdasarkan nilai *gain ratio*. Proses ini menghasilkan struktur pohon keputusan yang digunakan untuk mengklasifikasikan sentimen ulasan.

Confusion matrix: matriks berukuran dua kali dua yang menunjukkan kinerja model klasifikasi berdasarkan hasil prediksi terhadap nilai aktual. Dalam prosesnya terdapat empat nilai, yaitu (1) True Positive (TP): diberi label positif dan menurut machine learning juga positif, (2) False Positive (FP): diberi

label positif sedangkan menurut machine learning negative, (3) True Negative (TN): diberi label negatif dan juga menurut machine learning negatif, (4) False Negative (FN) diberi label negatif tetapi machine learning menganggap positif.

Accuracy: Persentase prediksi yang benar terhadap seluruh data uji.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: Ketepatan model dalam memprediksi kelas tertentu (positif atau negatif).

$$\text{Precision} = \frac{TP}{TP + FP}$$

Kemampuan model dalam menemukan seluruh data yang benar-benar termasuk dalam suatu kelas.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-Score: rata-rata harmonic dari precision dan recall.

$$F1 - Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3. HASIL DAN PEMBAHASAN

3.1. Dataset

	username	score	at	content
1	RUSDIANT	2	1/10/2025 12:33	ini kenapa Spotify nya selalu ada iklan terus pas saya mau liat lirik lagu itu ada yang tidak bisa dilihat, ini juga tbtd di suruh mendaftar ulang tolong ini di perbaiki dong
2	Arinda Syi	1	1/10/2025 14:53	iklannya tidak berputar bahkan saya tidak bisa mendengarkan lagu lagi di spotify aplikasinya sekarang berubah... kenapa ada premium padahal dulu nggak ada
3	Keltra	3	1/18/2025 14:16	Kualitas audio masih kurang, padahal udah berlangganan premium dan udah saya setting ke yang sangat tinggi semua. Tetap aja masih kurang kencang musiknya, aplikasi
4	Dimas Pra	1	1/18/2025 14:18	terlalu banyak bug, ketika mau mendaftarkan anggota selalu enggak ada jaringan padahal bagus sinyal emang ga worth, padahal udah sering langganan tidak ada pe
5	Martuti W	4	1/18/2025 23:03	Spotify keren!! Aku suka bisa dengerin lagu apa aja sama yaa lengkap gitu lah. Tapi tolong ya jangan bikin yg lain harus premium aku pas mau dengerin lagu
6	Aqila Pasir	5	1/19/2025 5:47	Dari say sendiri Spotify tuh nyaman banget dipake nya. Bisa disetel dimana mana dan kapan aja kita mau. Apalagi banyak tipe lagu disini. Dari seluruh aplikasi yang
7	Mystic Aet	1	1/19/2025 10:21	sekarang spotify gak sebebas kayak seperti sebelumnya segala harus premium denger 1 musik aja dapet iklan 2, mana 30 detik per iklan lagi download juga harus pr
8	Grastia Tia	4	1/19/2025 14:27	enak buat dengar lagu sambil kerja tugas, masak, membersihkan. Tapi kenapa sekarang spotify aku setiap masuk lama banget loadingnya padahal jaringan aku bagus
9	abdul gha	2	1/23/2025 1:23	Mau lanjut akun premium tapi sulit melakukan pembayaran karena eror tidak bisa terhubung ke web, selalu begitu terus dari dulu sampai detik ini. Mau bukti? Ada.
10	listiani	1	1/24/2025 0:39	ini kenapa updatenya kalau mau putar lagu mesti kali ada batas waktunya?! terus kalau mau dengar lagu di playlist sendiri harus juga keputar lagu lainnya yg sesua
11	Amanda P	1	1/24/2025 18:21	Spotify sekarang kenapa suka tiba-tiba berhenti si kalo play lagu sambil masuk ke apk lain?!! Padahal mau dengerin musik sambil nulis loh! Tapi setiap kali play musi
12	Rizki Maul	1	1/25/2025 14:44	iya iya bang, lu kikir banget masa mau putar lagu secara berurutan aja harus premium. Kalau nggak bisa di cepetin/pindahin menitnya masih mending lah. dari pada
13	Siti Hasnal	1	1/26/2025 12:42	Spotify tolong diperbaiki, bikin kesel tau kng!! Lagi enak-enak denger lagu malah mati sendiri, kalo mau denger lagu harus buka aplikasi lagi jadi tolong lah perbaiki bi
14	Fujiwara A	1	1/27/2025 10:37	Aplikasinya tiba-tiba suka ngelag ga jelas. Padahal dulu pas gua masih pake Spotify mod ga ada tuh kendala-kendala kek gini. Sok-sok2an ngebaned spotify mod tapi
15	Shirin Sub	2	1/28/2025 6:44	min tolonglah, ini mau nyetel lagu d playlist aja ga beraturan... udah gt juga kebanyakan premium nya, buat kaum gratisan kek aku bener' kurang bngt, mana iklannya
16	Shidqi Ryz	3	1/28/2025 8:58	Semuanya oke banget, kecuali beberapa hal berikut: (1) sistem shuffle yg buruk, sering kali sistem memberikan lagu-lagu yg sama dan tidak semua lagu-lagu dalam p
17	Akmal Ahi	1	1/28/2025 15:25	pls lah Spotify, kenapa lagu gua yang muncul jadi andmesh semua, cuma gara-gara temen gua pinjem akun gua buat dengerin lagu pas nongkrong dia nyetel andme
18	Novri	3	1/29/2025 12:14	ini aplikasi kok banyak kali iklan, terus kalo mau muter lagu yang mau dipilih harus premium dulu, dikit dikit premium, meskipun kualitas lagunya oke punya, tapi car
19	Aneylia	1	1/30/2025 4:00	Kenapa selalu muncul tulisan "kamu sedang offline" dan tidak bisa menyatel musik padahal internet saya lancar?!! Hmm, sudah berkali-kali uninstall dan install apli

Gambar 2. Hasil *crawling* data

Pada Gambar 2 menunjukkan bahwa data dikumpulkan berupa ulasan pengguna yang dipublikasikan dalam rentang waktu 10 Januari 2025 hingga 21 Juni 2025. Dari proses *crawling* yang dilakukan, diperoleh sebanyak 1000 data ulasan. Setiap entri ulasan memiliki empat atribut utama yaitu Username: nama pengguna yang memberikan ulasan; Score: nilai rating yang diberikan oleh pengguna dengan skala 1–5; At: tanggal dan waktu ulasan dikirim; Content: isi ulasan berupa teks opini pengguna terhadap aplikasi Spotify.

3.2. Pelabelan Data

Pelabelan data dalam penelitian ini dilakukan berdasarkan rating ulasan yang diberikan pengguna di *Google Play Store*. Proses pelabelan dilakukan semi manual di excel dengan menambahkan kolom baru untuk label sentimen. Kriteria yang digunakan adalah label positif untuk rating ulasan adalah 4 atau 5 sedangkan label negatif untuk rating ulasan adalah 1 atau 2. Ulasan dengan rating 3 tidak dimasukkan

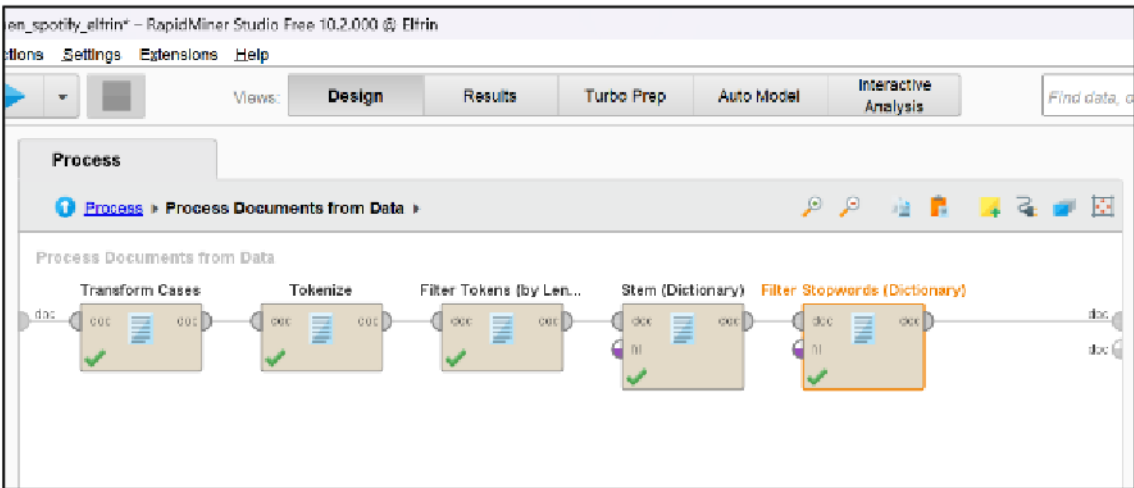
dalam dataset karena dianggap netral dan tidak merepresentasikan sentimen yang jelas. Setelah semua data (1000 ulasan) diberi label, peneliti kemudian mengambil masing-masing 50 % data berlabel positif dan 50% data berlabel negatif, sehingga total data yang digunakan dalam penelitian berjumlah 298 data ulasan yang seimbang (*balanced class*). Contoh hasil pelabelan data ditampilkan pada tabel 1, menunjukkan teks ulasan beserta label sentimen yang telah ditetapkan.

Tabel 1. Sampel pelabelan data

Ulasan	Score	Sentimen
lagunya lengkap, enak2, kalau lagunya yg dikepemngenin gk ada di apk ini bisa ditambahin dari apk lain	5	Positif
baguss apaknya, aku sering dengerin lagu disini 24/7	5	Positif
aku kasi reting 4 karena Spotify nya bagus dan enak di dengar terbiasa aku main game sampai denger lagi di Spotify	4	Positif
Aaaaaaa Spotify nya keren banget dan bagus tapi ada iklan nya	4	Positif
Review jujur. Menurutku spotify sudah menyediakan berbagai fitur-fitur dan konten yang menarik. namun ada sedikit kekurangan seperti, panel lirik yang terlalu monoton	4	Positif
ini kenapa Spotify kalau kita muter lagu lagu yang pertama belum selesai terus lagu yang ke dua udah jalan, mengganggu banget tolong di kasih waktu jeda antara lagu pertama dan lagu ke dua	2	Negatif
bug bug nya annoying BANGET, apa apa ada limit, dulu ngga gini padahal gw gak punya uang, tapi di suruh premium, apalagi gw gak punya dana, atau alat pembayaran online lainnya, tolong kalo pake tulisan hangeul(korea) kasih bacaan nya	1	Negatif
bug bug nya annoying BANGET, apa apa ada limit, dulu ngga gini padahal gw gak punya uang, tapi di suruh premium, apalagi gw gak punya dana, atau alat pembayaran online lainnya, tolong kalo pake tulisan hangeul(korea) kasih bacaan nya	1	Negatif
banyak banget bugnya sekarang, kenapa setelah diupdate bukan nya makin membaik? ini bukan soal device ya tapi memang banyak sekali perubahannya setelah diupdate, malah makin ga jelas,	1	Negatif
dulu, freeuser masih okey-okey aja dengan adanya iklan. (walau sempet entah sengaja atau engga. iklannya yang horor-horor) sekarang	1	Negatif

3.3. Pra-pemrosesan

Data ulasan yang telah dikumpulkan akan melalui serangkaian proses transformasi agar lebih mudah dianalisis oleh algoritma klasifikasi, yaitu algoritma C4.5.



Gambar 3. Pra-pemrosesan Data

Gambar 3 menunjukkan rancangan proses pra-pemrosesan data yang dilakukan menggunakan aplikasi RapidMiner. Proses ini difokuskan pada pengolahan atribut teks, yaitu ulasan, dengan menggunakan

operator *Process Documents from Data* untuk memungkinkan pelaksanaan beberapa tahapan pembersihan dan transformasi teks secara terstruktur sebelum data digunakan dalam tahap pemodelan.

Tabel 2. Sampel Data *case folding*

Sebelum	Sesudah
bug bug nya annoying BANGET, apa apa ada limit, dulu ngga gini padahal.	bug bug nya annoying banget, apa apa ada limit, dulu ngga gini padahal.

Tabel 3. Sampel Data *Tokenizing*

Sebelum	Sesudah
Spotify bagus, lagu lagunya baru rilis sudah ada dan tidak perlu pakai bayar" cukup Kouta dan internet music tetap menyala terus, terimakasih atas bantuannya	[spotify, bagus, lagu, baru, rilis, sudah, tidak, perlu, pakai, bayar, cukup, kouta, dan, internet, music, menyala, terus, terimakasih, atas, bantuan]

Tabel 4. Sampel Data *Stemming*

Setelah	Sebelum
Dengar	Mendengarkan
Lagu	Lagunya
Musik	Musiknya
Suka	Menyukai
Coba	Nyobain
Putar	Diputerin

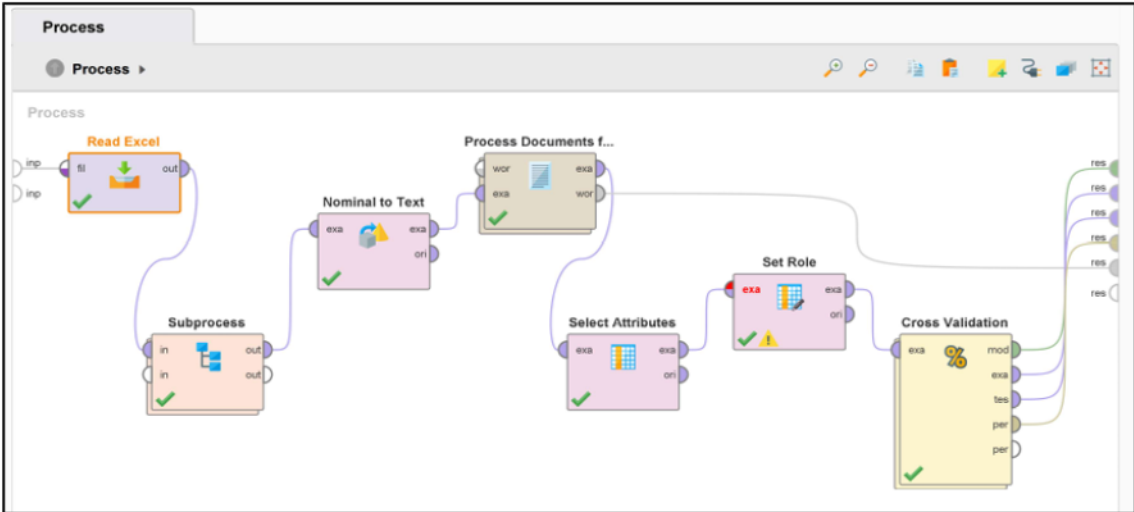
Tabel 5. Sampel Data *Stop Removal*

Sebelum	Setelah
[aku, kasi, reting, 4, karena, Spotify, nya, bagus, dan, enak, di, dengar, terbiasa, aku, main, game, sampai, denger, lagi, di, Spotify]	[kasi, reting, 4, Spotify, bagus, enak, dengar, main, game, denger, Spotify]

Pada penelitian ini, *case folding* dilakukan secara otomatis melalui operator *Process Documents* di RapidMiner, dengan mengaktifkan opsi *transform cases to lowercase* seperti yang ditampilkan pada tabel 2. Dengan melakukan tokenisasi, struktur kalimat dipecah menjadi kumpulan kata sehingga dapat diolah secara lebih spesifik oleh algoritma klasifikasi seperti C4.5 (Tabel 3). *Stemming* pada tabel 4 bertujuan untuk membuat variasi kata yang akar maknanya sama menjadi lebih sederhana., sehingga dapat mengurangi kompleksitas fitur dalam data teks. Untuk proses *Stopword Removal* berdasarkan tabel 5 dilakukan menggunakan *dictionary stopwords* Bahasa Indonesia yang diperoleh dari Kaggle. Daftar kata tersebut diimpor ke dalam RapidMiner melalui operator *Stopword Filter* dengan opsi "from file", kemudian diterapkan pada teks yang telah melalui proses tokenisasi.

3.4. Modeling Algoritma C4.5

Tahap ini merupakan inti dari proses klasifikasi sentimen. Proses import data yang diambil dari file Excel yang berisi kumpulan ulasan pengguna aplikasi Spotify beserta label sentimennya (positif atau negatif) adalah awal dalam pembangunan model klasifikasi. Data yang telah diproses melalui tahap pra-pemrosesan kemudian dimasukkan ke dalam proses modeling di aplikasi RapidMiner. Berikut modeling algoritma C4.5 menggunakan Rapidminer ditampilkan pada gambar 3 di bawah ini.



Gambar 3. Modeling C4.5 pada Rapidminer

3.5. Evaluasi Model

Evaluasi model dilakukan menggunakan beberapa metrik pengukuran yang paling umum digunakan dalam klasifikasi, yaitu akurasi, *precision*, *recall*, *f-measure*, dan AUC.

accuracy: 81.56% +/- 5.66% (micro average: 81.54%)			
	true Negatif	true Positif	class precision
pred. Negatif	137	43	76.11%
pred. Positif	12	106	89.83%
class recall	91.95%	71.14%	

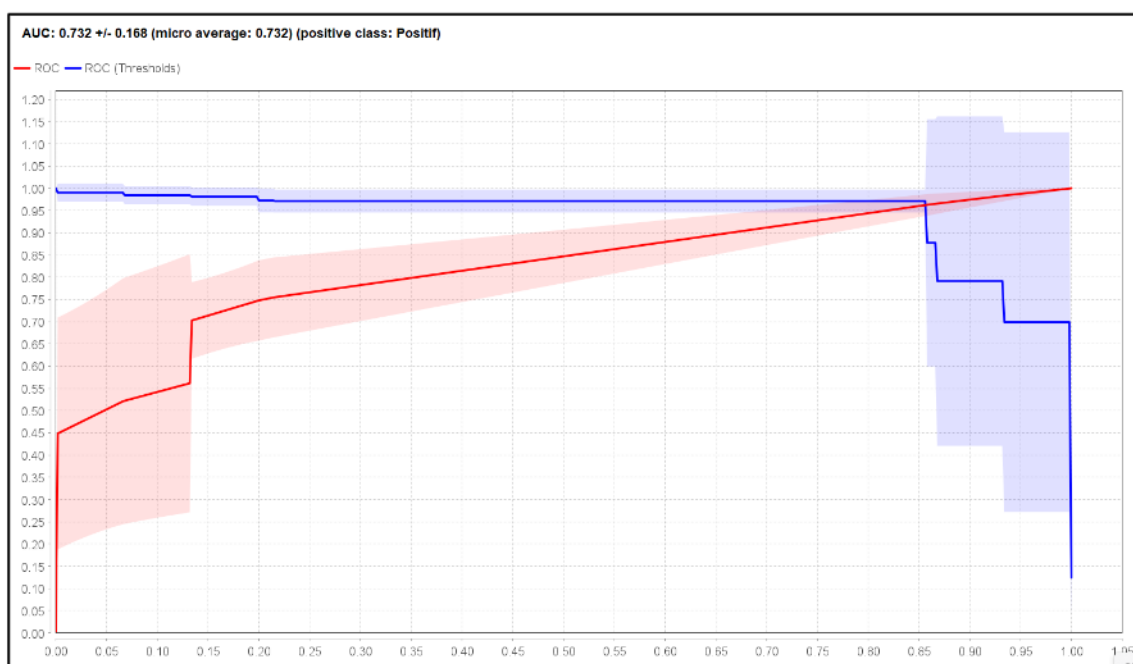
Gambar 4. Hasil Confusion Matrix

Tabel 6. Hasil Evaluasi Matriks

Evaluasi Matriks	Pemodelan C4.5
Accuracy	81.56%
Precision	90.92%
Recall	71.03%
F-measure	79.03%

Berdasarkan hasil pengujian yang ditunjukkan pada gambar 4 dan tabel 6, nilai akurasi yang diperoleh sebesar 81,56%, yang menunjukkan bahwa secara keseluruhan model mampu mengklasifikasikan data ulasan pengguna Spotify menjadi sentimen positif dan negatif dengan tingkat ketepatan yang cukup baik, meskipun ulasan memiliki variasi bahasa dan struktur kalimat. Pada metrik *precision*, model mencatat nilai 90,92%, yang berarti dari seluruh prediksi yang diklasifikasikan sebagai positif, sekitar 90,92% benar-benar positif. Hal ini sejalan dengan jumlah data pada *confusion matrix*, di mana terdapat 106 data *True Positive* (TP) dan 137 data *True Negative* (TN), sedangkan kesalahan prediksi ditunjukkan oleh 12 data *False Negative* (FN) dan 43 data *False Positive* (FP). Selanjutnya, nilai *recall* yang diperoleh sebesar 71,03%, menunjukkan bahwa dari seluruh data yang benar-benar positif, model berhasil mengenali sekitar 71% di antaranya, sementara sisanya terlewat sebagai negatif. Nilai *F1-score* atau *f-measure* yang dicapai sebesar 79,03%, yang mencerminkan keseimbangan antara *precision* dan *recall*.

Gambar 5 menunjukkan hasil kurva ROC dengan nilai AUC (Area Under Curve) dari model adalah sebesar 0.732. Nilai AUC yang berada di atas 0.7 menandakan bahwa model memiliki kemampuan yang cukup untuk membedakan antara kelas positif dan negatif.



Gambar 5. Hasil Kurva ROC

4. KESIMPULAN

Berdasarkan hasil penelitian, klasifikasi algoritma C4.5 mampu mengklasifikasikan sentimen ulasan pengguna aplikasi Spotify dengan tingkat ketepatan yang baik, terutama dalam mengidentifikasi sentimen positif. Model menghasilkan *precision* yang tinggi, yang menunjukkan bahwa prediksi sentimen positif cenderung akurat dan minim kesalahan klasifikasi. Penelitian ini membuktikan bahwa algoritma C4.5 dapat diterapkan secara efektif dalam klasifikasi sentimen ulasan pengguna aplikasi Spotify berbahasa Indonesia.

5. REFERENSI

- Afrianto, E., Suseno, J. E., & Warsito, B. (2025). Decision Tree Method with C4.5 Algorithm for Students Classification Who is Entitled to Receive Indonesian Smart Card (KIP). *IOP Conference Series: Materials Science and Engineering*, 879(1). <https://doi.org/10.1088/1757-899X/879/1/012072>
- As'ari, W. R., Arifin, M., Fithri, D. L., & Setiaji, P. (2025). Analisis Sentimen Ulasan Pengguna Aplikasi Spotify Di Google Play Store Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 3600–3607. <https://doi.org/10.36040/jati.v8i3.8739>
- Fadryona, R. (2021). The Spotify Effect: Peranan Spotify for Artists dalam Industri Musik 4.0. *Jurnal InterAct*, 9(2), 96–109. <https://doi.org/10.25170/interact.v9i2.2188>
- Faizah, T., & Jananto, A. (2021). Perbandingan Algoritma C4.5 Dan Id3 Untuk Prediksi Ketepatan Waktu Lulus Mahasiswa. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 980–990. <https://doi.org/10.35957/jatisi.v8i2.593>
- Fitrana, L. A., Linawati, S., & Herlinawati, N., Sa'adah, R. & Seimahuria, S. (2024). Analisis Sentimen Pengguna Twitter Terhadap Brand Indosat Menggunakan Metode Naïve Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 4291–4297. <https://doi.org/10.36040/jati.v8i3.9866>
- Ginabila & Fauzi, A. (2023). Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine. *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, 6(2), 111–122. <https://doi.org/10.47324/ilkominfo.v6i2.180>
- Marlia, S., Setiawan, K., & Juliane, C. (2024). Analisis Fitur Musik dan Tren Popularitas Lagu di Spotify menggunakan K - Means dan CRISP - DM. 13, 595–607.
- Rahayu, P. W., Sudipa, I. G. I., Suryani, S., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., ... & Maysanjaya, I. M. D. (2024). Buku Ajar Data Mining. PT. Sonpedia Publishing Indonesia.
- Rahmaddeni, Ningsih, W., Adrianto, S., Kurniawan, F., & Alfianda, B. (2024). Analisis Sentimen Pengguna

- Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45. *The Indonesian Journal of Computer Science*, 14(2), 8153–8164.
- Rosmawati, Kharisma, I. L., & Insany, G. P. (2024). Penerapan Algoritma Random Forest untuk Menganalisis Ulasan Aplikasi Spotify pada Google Play. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 369–378. <https://doi.org/10.29408/edumatic.v8i2.26394>
- Setyoko, A., Putra, B. A., & Rawanggalih, S. (2021). *Komparatif Terhadap Musik Sebagai “Bahasa Universal”*. 14(1), 1–11.
- Swastika, R., Mukofimah, S., Susanto, F., Muslihudin, M., & Ipnuwati, S. (2023). Implementasi data mining (clustering, association, prediction, estimation, classification). Indramayu: Adab.
- Talisman, B. R., Darmawan, I., & Pratiwi, O. N. (2024). *Perancangan Aplikasi Data Crawling Untuk Pencarian Buku Pada Toko Buku Online*. 11(4), 4333–4340.
- Wulandari, C., & Sunardi, L. (2024). *Analisis Sentimen Aplikasi Spotify Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Support Vector Machine*. 4(5), 2588–2595. <https://doi.org/10.30865/klik.v4i5.1762>
- Yolanda, M. N., Aini, H., & Sari, D. M. (2024). *Penggunaan Aplikasi Spotify Sebagai Media Baru Layanan Streaming Musik pada Generasi Z*. 2, 16–24.