The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 5, October 2024
https://doi.org/10.33022/ijcs.v13i5.4327

Analisis Sentimen Pengguna Aplikasi Gen Ai dari Appstore dan Googleplay Menggunakan Algoritma C45

Rahmaddeni[1], Widia Ningsih[2], Sukri Adrianto[3], Fadly Kurniawan[4], Baginda Alfianda[5]
rahmaddeni@sar.ac.id[1], widianingsih141101@gmail.com[2], sukriadrianto@gmail.com[3], 2310031802120@sar.ac.id[4], 2310031802022@sar.ac.id[5]
[1,2,4,5] Universitas Sains dan Teknologi Indonesia
[3] Universitas Dumai

Informasi Artikel

Diterima : 18 Jul 2024
Direvisi : 5 Agu 2024
Disetujui : 3 Sep 2024

Kata Kunci

Analisis Sentimen, Generative AI, Algoritma Decision Tree C4.5, AppStore, Google Play

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi Generative AI (ChatGPT, Bing AI, Microsoft Co-Pilot, Gemini AI, dan Da Vinci AI) yang tersedia di AppStore dan Google Play menggunakan algoritma C4.5. Data yang digunakan dalam penelitian ini diambil dari ulasan pengguna yang tersedia di kedua platform tersebut. Proses penelitian melibatkan pengumpulan data, pelabelan manual, dan pemrosesan awal data, termasuk cleansing, tokenisasi, transformasi kasus, dan penghapusan stopword. Data yang telah diproses kemudian dianalisis menggunakan algoritma C4.5 untuk mengklasifikasikan sentimen menjadi positif, negatif, atau netral. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan sentimen dengan akurasi, presisi, dan recall yang tinggi. Penelitian ini memberikan kontribusi dalam memahami opini pengguna terhadap aplikasi Generative AI, serta membantu pengembang aplikasi dalam meningkatkan kualitas dan kinerja aplikasi mereka berdasarkan umpan balik pengguna.

Keywords

Sentiment Analysis, Generative AI, Decision Tree C4.5 Algorithm, AppStore, Google Play

Abstract

This study aims to analyze the sentiment of users of Generative AI applications (ChatGPT, Bing AI, Microsoft Co-Pilot, Gemini AI, and Da Vinci AI) available on the AppStore and Google Play using the C4.5 algorithm. The data used in this research was taken from user reviews available on both platforms. The research process involves data collection, manual labeling, and initial data processing, including cleansing, tokenization, case transformation, and stopword removal. The processed data is then analyzed using the C4.5 algorithm to classify sentiment into positive, negative, or neutral. The results of the study show that the C4.5 algorithm can classify sentiment with high accuracy, precision, and recall. This research contributes to understanding user opinions on Generative AI applications and helps app developers improve the quality and performance of their applications based on user feedback.

A. Pendahuluan

Gen AI atau Generative Artificial Intelligence, adalah konsep yang mengacu pada perkembangan terbaru dalam teknologi kecerdasan buatan. Istilah ini merujuk pada evolusi AI yang lebih maju, yang mampu mengatasi kompleksitas yang lebih besar, memahami konteks, dan belajar secara mandiri dari data yang tersedia. Perkembangan Teknologi Gen AI yang signifikan, mencakup kemampuan seperti pembelajaran mendalam (deep learning), pemrosesan bahasa alami (natural language processing), dan pengambilan keputusan otomatis yang lebih canggih. Dalam konteks aplikasi Gen AI, teknik data mining dapat digunakan untuk menganalisis pola-pola pengguna, preferensi, atau perilaku dalam data yang terkumpul. Misalnya, teknik clustering dapat membantu dalam segmentasi pengguna berdasarkan perilaku mereka terhadap aplikasi Gen AI tertentu, sedangkan algoritma association dapat mengidentifikasi pola pembelian atau penggunaan aplikasi yang sering terjadi. Teknik klasifikasi dapat diterapkan untuk mengkategorikan ulasan pengguna menjadi positif, negatif, atau netral. Salah satu algoritma klasifikasi yang umum digunakan adalah algoritma Decision Tree, seperti C4.5. Algoritma ini dapat membangun model yang memprediksi sentimen dari ulasan berdasarkan fitur-fitur tertentu yang diekstraksi dari teks ulasan tersebut.

Suatu teknik yang dapat digunakan untuk mengolah komentar-komentar dan mengetahui sentimen dari penonton video tersebut yaitu analisis sentimen. Analisis sentimen (sentiment analysis) atau disebut juga opinion ming merupakan suatu teknik klasifikasi teks ke dalam kelompok-kelompok sentimen yang biasanya berupa sentimen positif, negative dan netral dengan mendeteksi polaritas yang terkandung dalam suatu opinil[1]. Analisis sentimen biasa digunakan untuk mengetahui reaksi pengguna, atau pembeli suatu jasa atau produk.

Penerapan analisis sentimen telah dilakukan oleh beberapa peneliti sebelumnya, salah satu penelitian analisis sentimen dilakukan untuk mengetahui opini masyarakat terkait penggunaan suatu produk yaitu analisis sentimen dilakukan dengan algoritma Naïve Bayes dan C.45 yang menggambarkan sentimen masyarakat atau pengguna twitter terhadap chatgpt dengan hasil perbandingan performa accuracy 77.33%, precision 100.00%, dan recall 30.18%[2].

Selain itu, studi mengenai “Analisis Sentimen Gofood” di Twitter yang memanfaatkan SVM dan algoritma Naïve Bayes dilakukan untuk mengetahui pendapat masyarakat umum terhadap kinerja Gojek (Gofood) di Indonesia. Penilaian Analisis Sentimen Gofood sebesar 92,8% bersifat netral, 5,2% positif, dan 2,0% negatif. Setelah membandingkan hasil akurasi, ditemukan bahwa algoritma SVM mengungguli algoritma Naïve Bayes dalam hal akurasi. Pendekatan Naïve Bayes menghasilkan skor akurasi sebesar 74,6% dan 91,5%, sedangkan teknik SVM menghasilkan hasil akurasi sebesar 83% dan 98,5%[3].

Analisis sentimen di YouTube dapat digunakan untuk mengatasi masalah termasuk leksikon sentimen kecil, gaya bahasa informal pengguna, dan memberi label pada nilai setiap kata, menurut penelitian sebelumnya [4] yang menerbitkan Analisis Sentimen di YouTube: Survei Singkat. Sementara itu, penelitian [5] dengan judul Perancangan Sistem Analisis Sentimen Komentar Pelanggan dengan Metode Naïve Bayes Classifier menyimpulkan bahwa pengklasifikasi Naïve Bayes dapat mengembalikan dokumen dan data yang cocok karena dapat menganalisis

sentimen dari 175 data dan terbukti memiliki nilai akurasi sentimen sebesar 75,42%. meningkat.

Berdasarkan latar belakang tersebut, tujuan dari penelitian ini adalah untuk melakukan analisis sentimen terhadap pengguna aplikasi gen AI (ChatGPT, Bing AI, Microsoft Co-Pilot, Gemini AI, dan Da Vinci AI) di AppStore dan Google Play menggunakan algoritma C4.5. Metode ini digunakan dengan menggabungkan informasi dari dua sumber utama, yaitu nilai rating aplikasi dan teks ulasan pengguna, untuk menganalisis pola sentimen positif, negatif, atau netral dari ulasan pengguna tersebut.

B. Metode Penelitian

1. Analisis Sentimen

Analisis sentimen mengkategorikan sentimen ke dalam tiga kategori: positif, negatif, dan netral berdasarkan polaritas bahasa dalam sebuah frasa[6][7]. Analisis sentimen, disebut juga analisis sentimen dalam bahasa Indonesia, adalah suatu proses atau metodologi yang digunakan untuk menentukan bagaimana suatu sentimen dikomunikasikan melalui bahasa dan apakah itu positif atau negatif[8].

2. Gen AI

Kecerdasan buatan yang dikenal sebagai "AI generatif" memungkinkan algoritma menghasilkan konten teks, gambar, audio, dan video secara otomatis. AI Generatif berpotensi mempercepat pengembangan kode, debugging, dan bahkan pembuatan laporan analitis di bidang analisis data[9]. Sebuah instruksi memulai AI generatif, instruksi ini dapat berupa teks, gambar, video, gambar, melodi, atau jenis masukan lainnya yang mampu ditangani oleh sistem AI. Berdasarkan instruksi tersebut, algoritma AI lainnya kemudian menghasilkan materi baru. Hasilnya bisa berupa artikel, solusi pemecahan masalah, atau replika gambar atau suara seseorang yang hidup.

3. Naïve Bayes

Berdasarkan teorema Bayes[10], Naive Bayes (NB) merupakan teknik klasifikasi langsung yang menghitung semua probabilitas dengan menggabungkan sejumlah nilai frekuensi dari database. Karena kata-kata tersebut dianggap mempunyai arti tertentu, maka kumpulan kata-kata dalam suatu dokumen dapat digunakan untuk mengidentifikasi kategorinya[11][12]. Memprediksi suatu kasus berdasarkan hasil klasifikasi adalah bagaimana metode klasifikasi Naive Bayes digunakan untuk mengambil keputusan. Dalam penelitian ini, sentimen ditentukan dari dokumen tweet dengan menggunakan metode Naive Bayes. Teknik probabilitas dan statistik adalah dasar dari prosedur ini[13][14]. Perhitungan probabilitas dan statistik digunakan dalam pengoperasian algoritma ini[15]. Pendekatan ini merupakan bentuk langsung dari persamaan (1)[16].

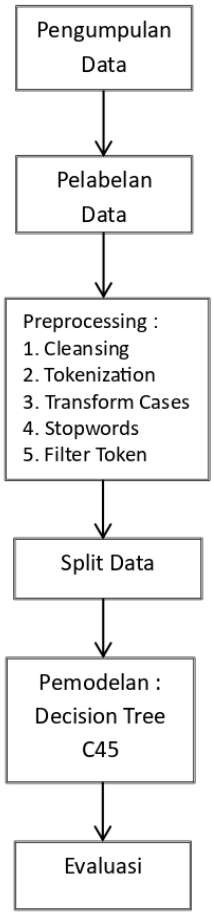
$$P(x|y) = P(y|x)\ P(x)P(y) \quad (1)$$

Dimana y adalah data dari kelas yang tidak diketahui. x = hipotesis data y P(x|y) adalah probabilitas hipotesis x dan y benar. P(x) adalah probabilitas hipotesis x.

P(y|x) adalah probabilitas y berdasarkan kondisi hipotesis. x P(y) adalah probabilitas bahwa y.
Aspek positif dari Naive Bayes[17]:
1. Cocok untuk data kualitatif dan kuantitatif
2. tidak memerlukan banyak data.
3. Tidak perlu mengumpulkan banyak data pelatihan.
4. Perhitungan dapat mengabaikan nilai apa pun yang hilang.
5. Perhitungannya mudah dimengerti dan cepat.

4. Tahapan Penelitian

Metode untuk mengklasifikasikan sentimen dari ulasan yang dikumpulkan adalah fokus utama penelitian ini. Layanan Google Colab dan Kaggle digunakan untuk memproses data ulasan, dan alat RapidMiner digunakan untuk mengkategorikan data. Gambar 1 menggambarkan tahapan penelitian.



Gambar 1. Tahap Penelitian

a. Pengumpulan Data

Pada tahap ini, peneliti mengambil data publik dari kaggle dengan jumlah data 7736 yang berisi 3 kolom yaitu text, ratting dan sentimen. Tampilan data ulasan pengguna Gen Ai dapat dilihat pada Tabel berikut :

b. Pelabelan Data

Setiap review diberi label secara manual dan diproses dalam file xlsx menggunakan aplikasi Microsoft Excel setelah data diperoleh. Tujuan pemberian tag pada kumpulan data adalah untuk membuat sentimen pada data lebih mudah diidentifikasi dan mengidentifikasi objek data dalam tinjauan.

c. *Preprocessing* Data

Pada tahap ini, label sentimen akan diterapkan pada data yang telah dikumpulkan. Langkah selanjutnya adalah mengolah data melalui tahap preprocessing setelah diberi label sentimen. Pra-pemrosesan data melalui beberapa tahapan:

a. *Cleaning* Proses pembersihan dokumen atau data bertujuan untuk menghilangkan unsur-unsur yang tidak relevan seperti nama pengguna (@), simbol, tanda baca, dan angka. Tujuan pembersihan data adalah untuk memastikan bahwa data tersebut hanya berisi informasi yang sesuai dan relevan[18].
b. *Tokenization* proses kumpulan kata dari suatu halaman, paragraf, atau kalimat dipecah menjadi satu frasa atau kata yang berfungsi sebagai satu kesatuan, tahap ini dikenal sebagai "tokenisasi". Angka, karakter pembatas, tanda baca, dan karakter non-abjad semuanya dapat dihilangkan dengan metode ini.
c. *Transform Cases* biasanya digunakan dalam tahap preprocessing. Parameter TF-IDF RapidMiner sangat bergantung pada prosedur ini, yang menyebabkan kalimat dalam dokumen menggunakan huruf kapital.
d. *Stopwords* untuk membuang kata-kata yang tidak berguna dan tidak penting adalah proses yang dikenal sebagai stopwords. Namun ada beberapa kata yang tidak relevan dengan penelitian ini dimasukkan kembali ke dalam data, seperti “kok”, “lah”, dan seterusnya.
e. Filter token menghapus kata-kata yang panjangnya melebihi jumlah karakter tertentu untuk meningkatkan kejelasan kalimat[19].

d. Split Data

Split data adalah metode yang digunakan untuk mempartisi kumpulan data dan merupakan salah satu dari beberapa perspektif yang memengaruhi seberapa baik kinerja demonstrasi klasifikasi dalam penghitungan pembelajaran mesin[20]. Split data merupakan bagian informasi dapat menjadi pegangan untuk mempartisi informasi penyiapan dan informasi pengujian[21].

e. Pemodelan

Pembentukan model dilakukan terhadap data ulasan yang telah diolah pada saat ini. Data analisis akan dibagi dengan rasio training-to-testing sebesar 80:20 pada tahap split data. Algoritma Decision Tree C4.5 digunakan peneliti untuk menghasilkan akurasi.

5. Evalusi

Pada tahap ini model yang dikembangkan dievaluasi untuk mengetahui apakah analisis yang dilakukan berhasil. Matriks konfusi digunakan untuk menghitung

pengukuran akurasi, presisi, dan recall dalam evaluasi ini. Matrix Precision adalah jenis akurasi yang berfokus pada keakuratan prediksi yang dibuat oleh model, sedangkan akurasi adalah metrik yang mengukur sejauh mana suatu model dapat mengklasifikasikan data dengan benar. Dengan membagi jumlah prediksi yang benar dengan jumlah sampel maka dapat ditentukan nilai akurasinya. Namun, recall adalah perbandingan antara jumlah hasil positif yang diprediksi dan observasi kelas sebenarnya[22].

C. Hasil dan Pembahasan

1. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah kumpulan data ulasan pengguna aplikasi Gen AI yang didapat dari laman Kaggle Open Datasets (kaggle.com) sebanyak 7761 data. Kumpulan data ini memiliki 3 atribut, yaitu: text, ratting, dan sentimen.

2. Pelabelan Data

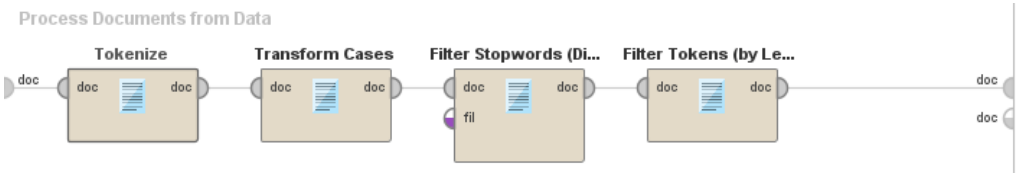
Proses pelabelan dilakukan secara manual untuk menetukan tweet tersebut sentimen positif yang berisi pujian seperti senang, bahagia, atau puas, sedangkan sentimen negatif berisi kritik, keluhan, hinaan, kesal, dan kecewa. Pelabelan dapat dilihat pada Tabel 1.

Tabel 1. Tampilan Data Ulasan

No	Text	Rating	Sentimen
1	Quite refreshing and impressive but not sure why the search function stuck halfway, failed to load any hit even after retrying, not sure if it's the issue with my phone or internet connection but when this happen and i try Chrome it works well and fast. keep up the good job... But why it keeps force closing when I'm streaming video as well as tab freeze.	3	Positif
2	nice ☺	1	Negatif
3	good ▯	5	Positif
4	very good and intresting .i lov rashiya	5	Positif
5	In Bing news, there is no way to translate it, in Google news, there is option to translate it just like I do on browser. Please add basic features it's a shame	2	Negatif

3. *Preprocessing*

Pre-processing merupakan langkah yang digunakan untuk menghapus kesalahan atau faktor lain yang dideteksi tidak sesuai dan dapat membuat nilai dari hasil proses data yang diolah menurun seperti melakukan *Cleansing* kata RT, url, hastag, mention, simbol, dan menghapus data duplikat. *Tokenizing* bertujuan untuk membagi teks dokumen dan menghilangkan data yang bukan huruf. Selanjutnya proses *Transform Case* yaitu untuk menyeragamkan huruf menjadi huruf kecil (*lower case*) dan dilanjutkan dengan proses Filter *Stopwords* yaitu menghapus katakata yang tidak relevan dengan memasukan file yang berisi list stopword. Proses *Filter Tokens (by Length)* juga diperlukan untuk menghapus kata dengan jumlah huruf yang kurang dari empat. *Prepocessing Text* ditunjukan pada gambar berikut.



Gambar 2. *Preprocessing* data

a. Untuk mempermudah proses Stopwords, tokenisasi adalah langkah yang memecah banyak kalimat menjadi dokumen kata yang lebih kecil. Representasi visual beberapa contoh komentar sebelum dan sesudah tahap tokenisasi akan ditampilkan pada tabel 1.

Tabel 1. Ilustrasi *Tokenization*

Sebelum *Tokenization*	**Sesudah *Tokenization***
Quite refreshing and impressive but not sure why the search function stuck halfway, failed to load any hit even after retrying, not sure if it's the issue with my phone or internet connection but when this happen and i try Chrome it works well and fast keep up the good job But why it keeps force closing when Im streaming video as well as tab freeze	'Quite', 'refreshing', 'and', 'impressive', 'but', 'not', 'sure', 'why', 'the', 'search', 'function', 'stuck', 'halfway', 'failed', 'to', 'load', 'any', 'hit', 'even', 'after', 'retrying', 'not', 'sure', 'if', 'its', 'the', 'issue', 'with', ' my', 'phone', 'or', 'internet', 'connection', 'but', 'when', 'this', 'happen', 'and', 'i', 'try', 'Chrome', 'it', 'works', 'well', 'and', 'fast', 'keep', 'up', 'the', 'good', 'job', 'But', 'why', 'it', 'keeps', 'force', 'closing', 'when', 'Im', 'streaming', 'video', 'as', 'well', 'as', 'tab', 'freeze'

b. Dalam review, langkah yang disebut "Transform Cases" digunakan untuk mengubah kata yang dimulai dengan huruf kapital menjadi huruf kecil. String Putska (perpustakaan) digunakan selama prosedur ini. Beberapa komentar sebelum dan sesudah tahap kasus transformasi ditunjukkan pada Tabel 2 yang dapat dilihat di bawah.

Tabel 2. Ilustrasi *Transform Cases*

Sebelum *Transform Cases*	**Sesudah *Transform Cases***
'Quite', 'refreshing', 'and', 'impressive', 'but', 'not', 'sure', 'why', 'the', 'search', 'function', 'stuck', 'halfway', 'failed', 'to', 'load', 'any', 'hit', 'even', 'after', 'retrying', 'not', 'sure', 'if', 'its', 'the', 'issue', 'with', ' my', 'phone', 'or', 'internet', 'connection', 'but', 'when', 'this', 'happen', 'and', 'i', 'try', 'Chrome', 'it', 'works', 'well', 'and', 'fast', 'keep', 'up', 'the', 'good', 'job', 'But', 'why', 'it', 'keeps', 'force', 'closing', 'when', 'Im', 'streaming', 'video', 'as', 'well', 'as', 'tab', 'freeze'	'quite', 'refreshing', 'and', 'impressive', 'but', 'not', 'sure', 'why', 'the', 'search', 'function', 'stuck', 'halfway', 'failed', 'to', 'load', 'any', 'hit', 'even', 'after', 'retrying', 'not', 'sure', 'if', 'its', 'the', 'issue', 'with', ' my', 'phone', 'or', 'internet', 'connection', 'but', 'when', 'this', 'happen', 'and', 'i', 'try', 'chrome', 'it', 'works', 'well', 'and', 'fast', 'keep', 'up', 'the', 'good', 'job', 'but', 'why', 'it', 'keeps', 'force', 'closing', 'when', 'im', 'streaming', 'video', 'as', 'well', 'as', 'tab', 'freeze'

c. Tahap menghilangkan seluruh konjungsi dari dataset adalah penghapusan stopword. Kamus stopwords yang dapat ditemukan di website Keagle dikutip oleh peneliti. Beberapa komentar akan ditampilkan pada tabel di bawah ini, baik sebelum maupun sesudah stopword pengolahan, seperti terlihat pada tabel 3.

Tabel 3. Ilustrasi *Stopword Removal*

Sebelum *Stopword Removal*	**Sesudah *Stopword Removal***
'quite', 'refreshing', 'and', 'impressive', 'but', 'not', 'sure', 'why', 'the', 'search', 'function', 'stuck', 'halfway', 'failed', 'to', 'load', 'any', 'hit', 'even', 'after', 'retrying', 'not', 'sure', 'if', 'its', 'the', 'issue', 'with', ' my', 'phone', 'or', 'internet', 'connection', 'but', 'when', 'this', 'happen', 'and', 'i', 'try', 'chrome', 'it', 'works', 'well', 'and', 'fast', 'keep', 'up', 'the', 'good', 'job', 'but', 'why', 'it', 'keeps', 'force', 'closing', 'when', 'im', 'streaming', 'video', 'as', 'well', 'as', 'tab', 'freeze'	'quite', 'refreshing', 'impressive', 'sure', 'why', 'search', 'function', 'stuck', 'halfway', 'failed', 'load', 'any', 'hit', 'even', 'after', 'retrying', 'sure', 'issue', 'with', ' my', 'phone', 'or', 'internet', 'connection', 'when', 'this', 'happen', 'try', 'chrome', 'works', 'well', 'fast', 'keep', 'good', 'job', 'why', 'keeps', 'force', 'closing', 'when', 'streaming', 'video', 'well', 'tab', 'freeze'

d. Filter Token adalah jenis token yang berbentuk kata-kata dengan panjang karakter tertentu untuk mengidentifikasi kalimat. Dalam proses ini, minimal tiga dan maksimal sepuluh karakter digunakan untuk panjang. Prosedur Filter Token berikut dan selanjutnya ditunjukkan pada Tabel 4 di sebelah kanan.

Tabel 4. Ilustrasi *Filter Token*

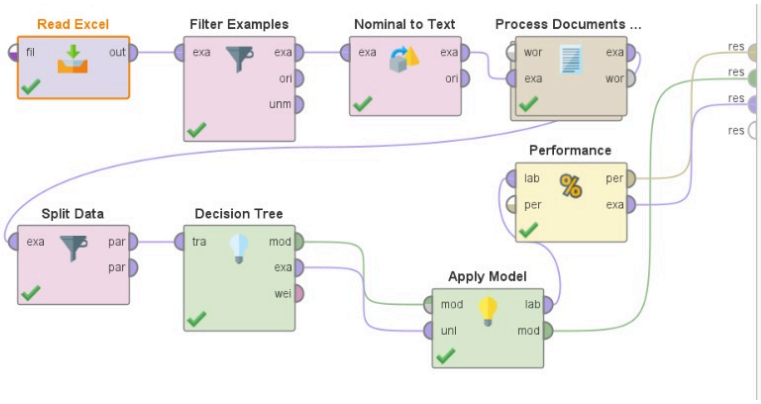
Sebelum *Filter Token*	**Sesudah *Filter Token***
'quite', 'refreshing', 'impressive', 'sure', 'why', 'search', 'function', 'stuck', 'halfway', 'failed', 'load', 'any', 'hit', 'even', 'after', 'retrying', 'sure', 'issue', 'with', ' my', 'phone', 'or', 'internet', 'connection', 'when', 'this', 'happen', 'try', 'chrome', 'works', 'well', 'fast', 'keep', 'good', 'job', 'why', 'keeps', 'force', 'closing', 'when', 'streaming', 'video', 'well', 'tab', 'freeze'	'quite', 'refreshing', 'impressive', 'sure', 'why', 'search', 'function', 'stuck', 'halfway', 'failed', 'load', 'any', 'hit', 'even', 'after', 'retrying', 'sure', 'issue', 'with', 'phone', 'internet', 'connection', 'when', 'happen', 'try', 'chrome', 'works', 'well', 'fast', 'keep', 'good', 'job', 'why', 'keeps', 'force', 'closing', 'when', 'streaming', 'video', 'well', 'tab', 'freeze'

4. Split Data

Setelah dataset sudah melalui preprocessing tahap selanjutnya yang dilakukan yaitu split data. Pada tahap ini data training yang digunakan yaitu 80:20.

5. Pemodelan

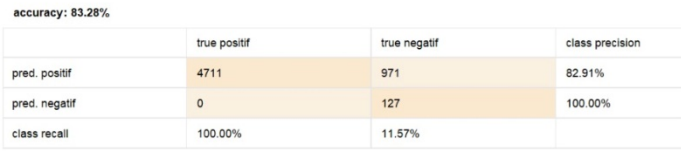
Setelah dataset sudah melalui preprocessing dan split data tahap selanjutnya yang dilakukan yaitu pemodelan. Pada tahap ini pemodelan yang digunakan yaitu klasifikasi Decision Tree C45. Proses pemodelan dapat dilihat pada Gambar 2.



Gambar 3. Model Klasifikasi Decision Tree C45

6. Evaluasi Hasil dan Pembahasan

Tahap selanjutnya yaitu tahap evaluasi yang bertujuan untuk melihat hasil validasi nilai *Accuracy, Class Precission, dan Class Recall* yang dihasilkan dari proses klasifikasi Decision Tree C45. Hasil pengujian algoritma Decision Tree C45 dapat dilihat pada Gambar 3.



accuracy: 83.28%

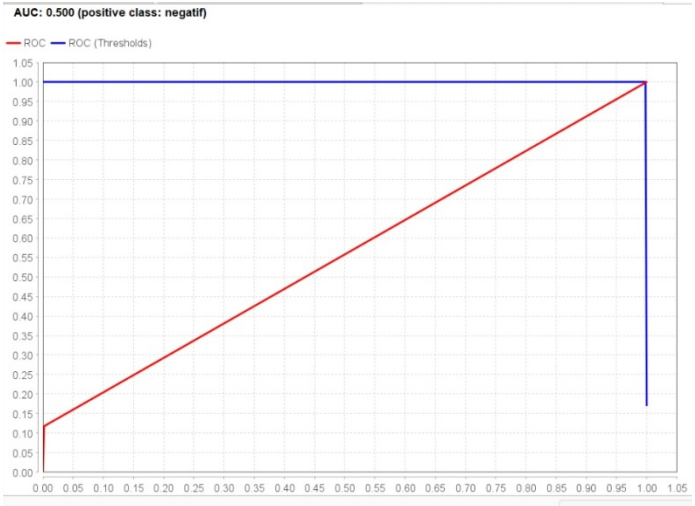
	true positif	true negatif	class precision
pred. positif	4711	971	82.91%
pred. negatif	0	127	100.00%
class recall	100.00%	11.57%	

Gambar 4. Hasil Pengujian Algoritma Decision Tree C45

Tabel 5. Nilai Hasil Validasi Decision Tree C45

Klasifikasi	Precision	Recall	Accuracy
Positif	82.91%	100.00%	83.28%
Negatif	100.00%	11.57%	

Berdasarkan hasil validasi menggunakan algoritma Decision Tree C45 pada Gambar 4 dan Tabel 5 dalam analisis sentimen twitter pada pengguna Gen AI yang telah di uji memiliki nilai *accuracy* 83.28%, *precision* positif 82.91%, *precision* negatif 100.00%, *recal* positif 100.00%, *recall* negatif 11.57%. Dalam bentuk grafik dapat dilihat pada Gambar 6.



Gambar 6. Grafik AUC Algoritma Decision Tree C45

Berdasarkan grafik AUC pada Gambar 6, maka dapat dilihat seberapa baik model Decision Tree C45 dalam mengklasifikasikan sentimen positif dan negatif. Dengan jumlah sentimen positif yang jauh lebih besar (5607) dibandingkan dengan sentimen negatif (1654), dapat disimpulkan bahwa mayoritas ulasan pengguna aplikasi Gen AI adalah positif.

D. Simpulan

Dapat diambil kesimpulan, berdasarkan seluruh hasil tahapan penelitian yang telah dilakukan pada analisis sentimen algoritma pohon keputusan pada review pengguna aplikasi Gen AI, bahwa analisis sentimen dapat dilakukan secara otomatis dan cepat dengan melakukan scraping data. Telah dibuktikan bahwa pra-pemrosesan komentar menghasilkan kalimat yang diperlukan untuk proses analisis sentimen. Algoritma Decision Tree C45 digunakan untuk menguji analisis sentimen pada komentar, dan hasilnya menunjukkan akurasi sebesar 83,28 persen, presisi positif sebesar 82,91 persen, presisi negatif sebesar 100,00 persen, recall positif sebesar 100,00 persen, dan recall negatif sebesar 11,57 persen. Kesimpulan yang diambil dari temuan ini adalah positif, yang menunjukkan bahwa Gen AI menerima tanggapan positif dari para pemberi komentar di Appstore dan Google Playstore.

E. Referensi

[1] M. Yusuf Rismanda Gaja, I. Maulana, and O. Komarudin, "Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan Playstore Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2767–2774, 2024, doi: 10.36040/jati.v7i4.7197.

[2] R. W. Permatasari, "Analisis Sentimen Masyarakat Indonesia Mengenai Vaksin COVID-19 Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier dan Support Vector …," vol. 5, no. 12, pp. 5680–5686, 2021, [Online]. Available: https://repository.its.ac.id/91280/

[3] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood

Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 542, 2022, doi: 10.30865/mib.v6i1.3530.

[4] S. Thomas, Yuliana, and Noviyanti. P, "Study Analisis Metode Analisis Sentimen pada YouTube," *J. Inf. Technol.*, vol. 1, no. 1, pp. 1–7, 2021, doi: 10.46229/jifotech.v1i1.201.

[5] E. M. Sipayung, H. Maharani, and I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *J. Sist. Inf.*, vol. 8, no. 1, pp. 2355–4614, 2016, [Online]. Available: http://ejournal.unsri.ac.id/index.php/jsi/index

[6] R. A. Husen, R. Astuti, L. Marlia, R. Rahmaddeni, and L. Efrizoni, "Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 211–218, 2023, doi: 10.57152/malcom.v3i2.901.

[7] A. F. Anees, A. Shaikh, A. Shaikh, and S. Shaikh, "Survey Paper on Sentiment Analysis : Techniques and Challenges," *EasyChair*, pp. 2516–2314, 2020, [Online]. Available: https://login.easychair.org/publications/preprint_download/Sc2h

[8] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.

[9] G. S. Mahendra *et al.*, *Tren Teknologi AI : Pengantar, Teori, dan Contoh Penerapan Artificial Intelligence di Berbagai Bidang*, no. April. 2024. [Online]. Available: https://books.google.co.id/books?id=qBsFEQAAQBAJ

[10] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.

[11] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. Gall, "ARdoc: App reviews development oriented classifier," *Proc. ACM SIGSOFT Symp. Found. Softw. Eng.*, vol. 13-18-Nove, no. November, pp. 1023–1027, 2016, doi: 10.1145/2950290.2983938.

[12] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "How can i improve my app? Classifying user reviews for software maintenance and evolution," *2015 IEEE 31st Int. Conf. Softw. Maint. Evol. ICSME 2015 - Proc.*, no. October, pp. 281–290, 2015, doi: 10.1109/ICSM.2015.7332474.

[13] A. R. Chrismanto and Y. Lukito, "Identifikasi Komentar Spam Pada Instagram," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 3, p. 219, 2017, doi: 10.24843/lkjiti.2017.v08.i03.p08.

[14] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews," *2020 Int. Conf. Contemp. Comput. Appl. IC3A 2020*, no. February, pp. 217–220, 2020, doi: 10.1109/IC3A48958.2020.233300.

[15] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi:

10.30864/eksplora.v9i1.237.
[16] M. E. Lasulika, "Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel," *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 11–16, 2019, doi: 10.33096/ilkom.v11i1.408.11-16.
[17] Rayuwati, Husna Gemasih, and Irma Nizar, "IMPLEMENTASI AlGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
[18] Amna *et al.*, *Data Mining*, vol. 2, no. January 2013. 2023. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
[19] I. Santoso, Windu Gata, and Atik Budi Paryanti, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 364–370, 2019, doi: 10.29207/resti.v3i3.1084.
[20] A. Nurhopipah and U. Hasanah, "Dataset Splitting Techniques Comparison For Face Classification on CCTV Images," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 4, p. 341, 2020, doi: 10.22146/ijccs.58092.
[21] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *J. Sains dan Inform.*, no. August, pp. 19–28, 2023, doi: 10.34128/jsi.v9i1.622.
[22] V. Muslimah *et al.*, "Kemajuan dalam Ilmu Informatika Dari Decision Support System Menuju Artificial Intelligence," 2024.