

Penerapan Algoritma C4.5 dengan Reduced Error Pruning untuk Klasifikasi Tingkat Quarter Life Crisis Mahasiswa

Elsa Valen Yunita Wijaya

Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

INFO ARTIKEL

Histori Artikel

Diterima : 16 Juni 2026

Direvisi : 29 Juli 2026

Disetujui : 7 Agustus 2026

Kata Kunci

Quarter Life Crisis;

Algoritma C4.5;

Reduced Error Pruning;

Klasifikasi;

Pohon Keputusan

Keywords

Quarter Life Crisis;

C4.5 Algorithm;

Reduced Error Pruning;

Classification;

Decision Tree

ABSTRAK

Fase transisi dari dunia perkuliahan menuju dunia profesional seringkali memicu quarter life crisis pada mahasiswa tingkat akhir akibat tingginya tekanan akademik, sosial, dan ketidakpastian karir. Penelitian ini bertujuan membangun model klasifikasi tingkat quarter life crisis mahasiswa serta menganalisis efektivitas algoritma C4.5 yang dioptimalkan dengan teknik Reduced Error Pruning (REP) untuk mengatasi overfitting dan menyederhanakan struktur pohon keputusan. Data dikumpulkan dari 952 responden mahasiswa aktif melalui kuesioner yang mencakup variabel demografis dan psikologis, kemudian diproses melalui tahap persiapan data, pembentukan model dengan rasio pembagian data 70:15:15, dan evaluasi menggunakan confusion matrix. Atribut kesiapan karir, kejelasan karir, dan reaksi terhadap tekanan kerja teridentifikasi sebagai prediktor paling dominan berdasarkan nilai gain ratio tertinggi. Penerapan REP meningkatkan akurasi model dari 82,3% menjadi 87,4% dan menyederhanakan pohon dari 14 menjadi 10 simpul keputusan, sehingga aturan klasifikasi lebih mudah dibaca oleh praktisi konseling. Kebaruan penelitian ini terletak pada integrasi REP ke dalam algoritma C4.5 untuk data kuesioner psikologis mahasiswa Indonesia. Model yang dihasilkan terbukti efektif memberikan rekomendasi objektif bagi unit layanan konseling kampus dalam deteksi dini masalah psikologis mahasiswa.

The transition from university life to the professional world often triggers a quarter-life crisis among final-year students due to high academic and social pressure as well as career uncertainty. This study aims to build a classification model of students' quarter-life crisis levels and to analyse the effectiveness of the C4.5 algorithm optimised with the Reduced Error Pruning (REP) technique in addressing overfitting and simplifying the decision tree structure. Data were collected from 952 active student respondents through a questionnaire covering demographic and psychological variables, then processed through data preparation, model building with a 70:15:15 data split, and evaluation using a confusion matrix. Career readiness, career clarity, and reaction to work pressure were identified as the most dominant predictors based on the highest gain ratio values. Applying REP raised model accuracy from 82.3% to 87.4% and simplified the tree from 14 to 10 decision nodes, making the classification rules easier to read for counselling practitioners. The novelty of this study lies in integrating REP into the C4.5 algorithm for Indonesian students' psychological questionnaire data. The resulting model proved effective in providing objective recommendations for campus counselling units in the early detection of students' psychological problems.

* Corresponding author: elsav.y.wijaya@gmail.com

<https://doi.org/10.56427/jcbd.v5i3.1041>

© 2026 The Authors. Published by PT. Delitekno Media Mandiri. This is an open access article under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

1. Pendahuluan

Banyak peneliti saat ini tengah berupaya untuk mengembangkan kecerdasan buatan hingga mencapai tingkat sebanding dengan kecerdasan manusia. Alur kerja dari Machine Learning adalah mengambil data mentah sebagai input dan menghasilkan prediksi sebagai output [1]. Kebanyakan penelitian terdahulu mengenai Quarter Life Crisis (QLC) pada dewasa awal dan mahasiswa cenderung terfokus pada pendekatan psikologi deskriptif atau analisis statistik konvensional [2]. Meskipun beberapa studi machine learning mulai menerapkan algoritma klasifikasi seperti Decision Tree C4.5 pada domain pendidikan dan krisis psikologis [3], masih terdapat keterbatasan mendasar (research gap) dalam penanganan masalah overfitting pada struktur pohon keputusan yang terbentuk dari data survei psikologis yang kompleks dan berisiko mengalami kecenderungan bias (noise). Algoritma C4.5 standar sering kali menghasilkan pohon keputusan yang terlalu kompleks (overfitted) dengan jumlah aturan (rules) yang berlebihan, sehingga menurunkan daya generalisasi model saat diterapkan pada data baru [4]. Meskipun metode pemangkasan pasca-pembentukan (post-pruning) seperti Reduced Error Pruning (REP) telah terbukti mampu menyederhanakan struktur pohon pada dataset generik, efektivitas dan efisiensi algoritma REP dalam mengoptimalkan akurasi klasifikasi serta mereduksi kompleksitas aturan khusus pada prediksi tingkat krisis emosional/QLC mahasiswa belum

dieksplorasi secara mendalam. Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan mengimplementasikan dan menganalisis secara komprehensif performa Reduced Error Pruning (REP) pada algoritma C4.5 dalam meningkatkan efisiensi dan akurasi prediksi tingkat QLC mahasiswa dimana model Tree akan menghasilkan rule dan disederhanakan [5].

Manusia pada tahap kehidupannya akan mengalami suatu masa transisi yang disebut Emerging Adulthood [5]. Yakni suatu periode ketika individu tidak lagi dapat disebut sebagai remaja namun belum dapat pula disebut sebagai dewasa awal. Krisis emosional yang terjadi pada masa ini dikenal dengan istilah Quarter Life Crisis [5], [6]. Hal ini merupakan gejala psikologis yang dialami pada saat memasuki usia 20 sampai 30 tahunan awal. Secara akademik, usia ini merupakan usia yang sesuai dengan masa studi individu di perguruan tinggi [3]. Penelitian yang bertujuan untuk memprediksi Tingkat Quarter Life Crisis dikalangan mahasiswa menggunakan Metode Reduced Error Pruning (REP) dapat memberikan manfaat dalam mengetahui cabang-cabang yang informatif dalam klasifikasi tingkatan Quarter Life Crisis menggunakan metode pemangkasan pada Algoritma C4.5. Fase ini merupakan tahapan algoritma klasifikasi untuk membentuk aturan klasifikasi data yang diambil melalui analisis data training. Tahap kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari aturan klasifikasi [7].

2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan penelitian campuran (mixed methods) yang menggabungkan bentuk kuantitatif dan kualitatif secara simultan demi memahami isu penelitian serta data numerik secara mendalam [8]. Pendekatan kuantitatif diterapkan melalui pemodelan komputasi statistik berbasis Machine Learning menggunakan pohon keputusan (Decision Tree) dengan Algoritma C4.5 dan teknik pemangkasan Reduced Error Pruning (REP). Data kuantitatif penelitian ini diperoleh dari angka hasil skor kuesioner psikologis yang disebarkan kepada responden dimana Algoritma C4.5 memiliki metode dan fungsi yang ketika diterapkan pada algoritma pohon keputusan tunggal memiliki kecepatan dan akurasi klasifikasi yang tinggi. Sementara itu, pendekatan kualitatif diaplikasikan melalui metode studi literatur, observasi langsung, serta wawancara mendalam guna memperkuat analisis interpretasi hasil klasifikasi tingkatan Quarter Life Crisis (QLC) pada subjek penelitian [2], [9].

2.1 Jenis Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif dengan pendekatan eksperimen komputasional (computational experimental research) berbasis algoritma machine learning. Pendekatan kuantitatif diterapkan untuk mengukur dan mengklasifikasikan tingkat Quarter Life Crisis (QLC) mahasiswa berdasarkan data numerik dan variabel kategorikal skala Likert yang dihimpun melalui kuesioner terstruktur [10]. Desain penelitian dirancang secara kuasi-eksperimental komputasi (quasi-computational design) yang berfokus pada analisis perbandingan performa (comparative performance analysis) antara model klasifikasi Decision Tree C4.5 standar dengan model Decision Tree C4.5 yang dikombinasikan dengan metode pemangkasan pasca-pembentukan (post-pruning) Reduced Error Pruning (REP).

2.2 Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah kuesioner dan studi dokumentasi [11]. Kuesioner disebarkan secara online menggunakan Google Forms kepada mahasiswa yang menjadi sampel penelitian. Kuesioner terdiri dari dua bagian, yaitu data demografis responden (usia, jenis kelamin, fakultas, semester) dan pernyataan terkait tujuh variabel bebas yang diukur menggunakan skala Likert 5 poin, mulai dari 1 (sangat tidak setuju) hingga 5 (sangat setuju) Skala Likert dipilih karena mampu mengukur sikap, persepsi, dan pendapat responden secara kuantitatif.

2.3 Instrumen Penelitian

Instrumen penelitian yang digunakan adalah kuesioner Quarter Life Crisis Scale yang diadaptasi dari penelitian Atwood dan Scholtz [12] dan dimodifikasi sesuai konteks mahasiswa Indonesia. Kuesioner terdiri dari 19 butir pernyataan tidak termasuk Jenis kelamin dan Rentang Umur yang mewakili lima variabel bebas. Kisi-kisi instrumen disajikan pada tabel 1 untuk memudahkan pemahaman hubungan antara indikator dan butir pernyataan.

Tabel 1. Atribut dan Variabel

Variabel	Indikator	Butir	Jumlah
Akademik	Kecocokan Prodi, IPK, Kepuasan Prodi, Perubahan Tujuan Pendidikan, lama kuliah	1-6	6
Karir	Kesiapan karir, Kejelasan Karir, Peluang Karir, Reaksi kerja	7-10	4
Kondisi Ekonomi	Kestabilan Ekonomi, Pengaruh Ekonomi, Kepuasan Ekonomi	11-13	3
Kepercayaan Diri	Kompetensi, Penampilan Fisik, Perasaan IPK, Prestasi	14-17	4
Dukungan Sosial	Dukungan Keluarga, Jumlah Prestasi	18-19	2

2.4 Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh mahasiswa aktif perguruan tinggi yang berada pada fase rentang usia dewasa awal (18 hingga 25 tahun) dan berpotensi mengalami tekanan psikologis Quarter Life Crisis (QLC). Teknik pengambilan sampel dilakukan menggunakan pendekatan non-probability sampling melalui metode purposive sampling [13], di mana responden dipilih berdasarkan kriteria inklusi spesifik yang mencakup status keaktifan sebagai mahasiswa, kesesuaian kelompok usia, serta

kelengkapan pengisian instrumen kuesioner. Melalui pendekatan ini, penelitian berhasil mengumpulkan data primer sebanyak 952 responden terverifikasi yang mencakup 19 indikator variabel penentu QLC yang terbagi ke dalam lima dimensi utama, yaitu akademik, karir, ekonomi, Kepercayaan diri, dan dukungan sosial. Dari keseluruhan basis data tersebut, sebanyak 952 data responden difokuskan sebagai sub-sampel eksperimental untuk pemodelan komputasi Decision Tree C4.5 dan pengujian validasi algoritma Reduced Error Pruning (REP) dengan pembagian proporsi data latih (training data) sebesar 70%, data validasi (validation data) sebesar 15%, dan data uji (testing data) sebesar 15% [14].

3. Hasil dan Pembahasan

Analisis data merupakan tahapan awal dalam proses pengolahan dataset sebelum dilakukan pembentukan model klasifikasi menggunakan algoritma C4.5. Tahapan ini bertujuan untuk memastikan bahwa data yang diperoleh dari hasil penyebaran kuesioner telah memenuhi kebutuhan proses klasifikasi sehingga mampu menghasilkan model yang akurat [15]. tahapan ini bertujuan untuk memastikan bahwa data yang diperoleh dari hasil penyebaran kuesioner telah memenuhi kebutuhan proses klasifikasi sehingga mampu menghasilkan model yang akurat. Proses analisis data pada penelitian ini meliputi seleksi atribut, data cleaning, transformasi data ke dalam skala numerik, pembentukan variabel target Quarter Life Crisis (QLC), analisis distribusi kelas, serta pembagian data menjadi data latih, data validasi, dan data uji [16].

3.1 Pengolahan Dataset

Atribut awal yang digunakan sebelum proses seleksi ditampilkan pada tabel 2.

Tabel 2. Atribut Data Set Sebelum Seleksi

No	Atribut	Keterangan
1	rentang_umur	18-21, 22-25
2	jenis_kelamin	Pria, Wanita
3	program_studi	Ilmu Komputer, Sistem Informasi, Matematika, Manajemen Bisnis, Hukum, Agama, Kesehatan Masyarakat, Lainnya
4	kecocokan_prodi	Sesuai, Tidak Sesuai
5	perubahan_tujuan_pendidikan	Ya, Tidak
6	kepuasan_prodi	Puas, Cukup Puas, Tidak Puas
7	pd_kompetensi	Tinggi, Sedang, Rendah
8	lama_kuliah	<1 tahun, 1-2 tahun, 3-4 tahun, >4 tahun
9	jumlah_prestasi	Tidak ada, 1-3, 4-6, 8-10, >10
10	pendapat_prestasi	Penting, Cukup Penting, Tidak Penting
11	ipk	>3,5, 3,0-3,5, 2,5-3,0, <2,5
12	perasaan_ipk	Puas, Cukup Puas, Tidak Puas
13	kestabilan_ekonomi	Stabil, Cukup Stabil, Tidak Stabil
14	pengaruh_ekonomi	Besar, Sedang, Kecil
15	kepuasan_ekonomi	Puas, Cukup Puas, Tidak Puas
16	dukungan_keluarga	Tinggi, Sedang, Rendah
17	reaksi_kerja	Positif, Netral, Negatif
18	kesiapan_karir	Siap, Cukup Siap, Tidak Siap
19	peluang_karir	Tinggi, Sedang, Rendah
20	kejelasan_karir	Jelas, Cukup Jelas, Tidak Jelas
21	persepsi_diri_fisik	Positif, Netral, Negatif

Setiap responden memiliki satu nilai untuk setiap atribut yang direpresentasikan dalam bentuk data kategorikal. Data ini kemudian digunakan sebagai input dalam proses pembentukan pohon keputusan.

Tabel 3. Distribusi Tingkat Quarter Life Crisis Responden

Tingkat Quarter Life Crisis	Frekuensi	Persentase
Rendah	246	25,8%
Sedang	413	43,4%
Tinggi	293	30,8%
Total	952	100%

Tabel 3 menunjukkan bahwa mayoritas responden (43,4%) berada pada tingkat quarter life crisis sedang, diikuti oleh kategori tinggi sebesar 30,8% dan kategori rendah sebesar 25,8%. Temuan ini mengindikasikan bahwa sebagian besar mahasiswa mengalami tingkat quarter life crisis yang cukup signifikan, yang sejalan dengan penelitian sebelumnya yang menyatakan bahwa fenomena ini umum terjadi pada dewasa muda. Analisis lebih lanjut dilakukan untuk mengidentifikasi atribut-atribut yang paling berpengaruh dalam pembentukan pohon keputusan. Berdasarkan perhitungan gain ratio, atribut dengan nilai tertinggi menjadi root node dalam pohon keputusan.

Tahap terakhir pada analisis data adalah membagi dataset menjadi data latih (training data), data validasi (validation data), dan data uji (testing data) menggunakan metode train-test split dengan rasio 70% : 15% : 15%. Pembagian ini bertujuan agar model dapat dilatih, divalidasi, dan diuji menggunakan partisi data yang berbeda sehingga hasil evaluasi akhir bersifat objektif dan terhindar dari bias data. Proporsi pembagian dataset tersebut disajikan pada tabel 4.

Tabel 4. Proporsi Pembagian Dataset

Komponen Dataset	Jumlah Kasus (Baris)	Persentase Rasio
Data Latih	666	70%
Data Validasi	143	15%
Data Uji	143	15%
Total	952	100%

3.2 Representasi Data

Representasi data merupakan tahapan untuk mengubah dataset yang telah melalui proses analisis data menjadi informasi struktural berupa pohon keputusan (decision tree). Pada tahap ini dilakukan proses perhitungan matematis nilai Entropy, Information Gain, dan Gain Ratio secara manual untuk menentukan atribut mana yang paling kuat dijadikan sebagai simpul akar (root node) maupun simpul cabang (internal node).

3.3 Perhitungan Entropy Total

$$Entropy(S) = -\left(\frac{246}{952}\right) \log_2 \left(\frac{246}{952}\right) - \left(\frac{413}{952}\right) \log_2 \left(\frac{413}{952}\right) - \left(\frac{293}{452}\right) \log_2 \left(\frac{293}{452}\right) \quad (1)$$

Entropy (S) adalah 1,5504

Sebagai contoh perhitungan konkret pemilihan cabang, diambil atribut: "Sejauh mana kesiapan anda untuk berkarir setelah tamat kuliah?". Atribut ini memiliki dua sebaran nilai respon numerik dominan pada data latih. Nilai Kelas Rendah (Jumlah Data = 126), Nilai Kelas Sedang (Jumlah Data = 239) dan Nilai Kelas Tinggi (Jumlah Data = 587) Hasil perhitungan entropy pada masing-masing nilai atribut diperoleh pada nilai Kelas rendah adalah (0,1324), kelas sedang (0,2511) dan kelas tinggi (0,6166). Setelah mendapatkan nilai entropy sub-atribut, dilakukan perhitungan Entropy Terbobot untuk melihat akumulasi

$$Entropy\ Terbobot = \frac{587}{952} (0,6166) + \frac{239}{592} (0,2511) + \frac{126}{592} (0,1324) = 1,3169 \quad (2)$$

Selanjutnya nilai Information Gain dihitung menggunakan Persamaan (3).

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_i|}{|S|} Entropy(S_i) \quad (3)$$

$$Gain(S, A) = 1,5504 - 1,3169 = 0,2235 \quad (4)$$

Kemudian untuk menyempurnakan Algoritma C4.5 pada kelemahan ID3 dengan menggunakan Gain Ratio, agar pemilihan atribut tidak pincang pada atribut yang memiliki banyak variasi nilai unik. Langkah awal adalah mencari nilai pembagi berupa Split Information menggunakan Persamaan (5).

$$SplitInfo = -\sum \frac{|S_i|}{|S|} \log_2 \left(\frac{|S_i|}{|S|}\right) \quad (5)$$

Sehingga :

$$\begin{aligned} SplitInfo &= -\left(\frac{587}{952} \log_2 \frac{587}{952}\right) - \left(\frac{239}{952} \log_2 \frac{239}{952}\right) - \left(\frac{126}{952} \log_2 \frac{126}{952}\right) \\ &= 0,4301 + 0,5005 + 0,3861 \\ &= 1,3169 \end{aligned} \quad (6)$$

Selanjutnya nilai Gain Ratio dihitung menggunakan Persamaan (7).

$$GainRatio = \frac{Gain}{SplitInfo} = \frac{0,2235}{1,3169} = 0,1697 \quad (7)$$

Hasil tersebut menunjukkan bahwa atribut "Menurut anda seberapa besar Ekonomi memengaruhi keberhasilan anda?" memiliki kemampuan yang baik dalam membedakan kategori Quarter Life Crisis setelah dilakukan normalisasi terhadap nilai Information Gain

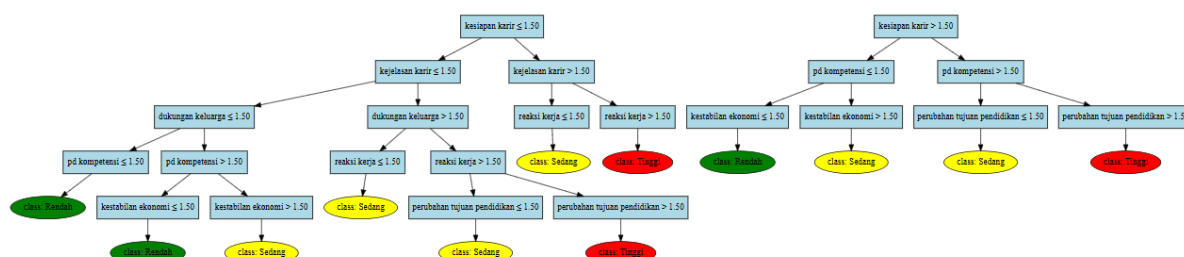
Tabel 5. Rekapitulasi Hasil Perhitungan Entropy, Information Gain, dan Gain Ratio

No	Perhitungan	Hasil
1	Entropy Total	1,5504
2	Entropy Kelas Rendah	0,1324
3	Entropy Kelas Sedang	0,2511
4	Entropy Kelas Tinggi	0,6166

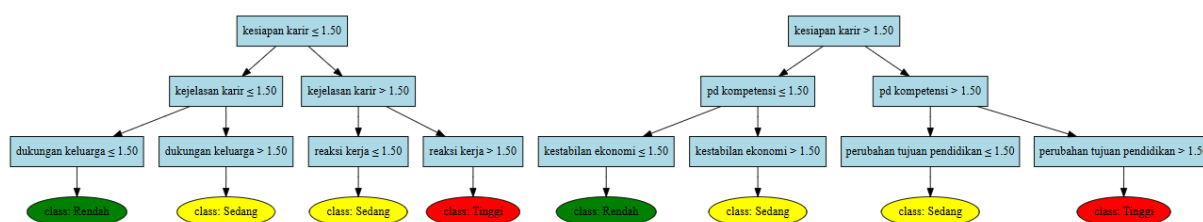
5	Entropy Terbobot	0,2235
6	Information Gain	0,1157
7	Split Information	1,3169
8	Gain Ratio	0,1697

3.4 Pembentukan Simpul Pokon Keputusan (Decision Tree)

Pohon keputusan sebelum REP memiliki struktur yang lebih kompleks dengan kedalaman yang lebih dalam dan jumlah node yang lebih banyak. Terdapat 14 node keputusan dan 15 daun (leaf) yang dihasilkan. Atribut kesiapan_karir menjadi root node yang pertama kali memisahkan data. Jika kesiapan karir rendah (nilai $\leq 1,50$), maka selanjutnya dilihat atribut kejelasan_karir. Jika kejelasan karir juga rendah, maka tingkat quarter life crisis ditentukan oleh atribut dukungan_keluarga dan kemudian pd_kompetensi atau kestabilan_ekonomi. Sebaliknya, jika kesiapan karir tinggi (nilai $> 1,50$), maka atribut pd_kompetensi menjadi penentu selanjutnya. Struktur pohon keputusan sebelum pruning tersebut ditampilkan pada gambar 1.



Gambar 1. Pohon Keputusan Sebelum Penerapan Pruning



Gambar 2. Pohon Keputusan Setelah Penerapan Pruning

Pohon keputusan sesudah REP memiliki struktur yang lebih sederhana dengan kedalaman yang lebih dangkal dan jumlah node yang lebih sedikit. Terdapat 10 node keputusan dan 8 daun (leaf) yang dihasilkan. Proses pruning menghilangkan cabang-cabang yang tidak signifikan, seperti cabang pd_kompetensi dan kestabilan_ekonomi pada subpohon kesiapan_karir rendah dengan kejelasan_karir rendah dan dukungan_keluarga rendah, serta cabang reaksi_kerja dan perubahan_tujuan_pendidikan pada subpohon yang sama. Penyederhanaan ini mengurangi kompleksitas model tanpa mengurangi akurasi secara signifikan, bahkan meningkatkan kemampuan generalisasi. Struktur pohon keputusan hasil pruning ditampilkan pada gambar 2.

Perbandingan visual antara gambar 1 dan gambar 2 menunjukkan bahwa penerapan REP berhasil mengurangi jumlah node dari 14 menjadi 10 node keputusan dan jumlah daun dari 15 menjadi 8 daun. Hal ini mengindikasikan bahwa teknik pruning efektif dalam menyederhanakan struktur pohon keputusan dengan tetap mempertahankan aturan klasifikasi yang paling penting.

3.5 Pengujian

Pengujian model dilakukan untuk mengevaluasi kinerja algoritma C4.5 dengan *Reduced Error Pruning* dalam mengklasifikasikan tingkat *quarter life crisis*. Pengujian dilakukan dengan menggunakan data testing yang terdiri dari 143 sampel (15% dari total data). Hasil pengujian disajikan dalam bentuk matriks klasifikasi dan matriks konfusi.

Tabel 6. Matriks Klasifikasi untuk Setiap Kategori

Kategori Aktual	Kategori Prediksi	Jumlah Sampel
Rendah	Rendah	31
Rendah	Sedang	4
Rendah	Tinggi	1
Sedang	Rendah	3
Sedang	Sedang	55
Sedang	Tinggi	5
Tinggi	Rendah	1

Tinggi	Sedang	4
Tinggi	Tinggi	39

Tabel 6 menunjukkan bahwa model berhasil mengklasifikasikan dengan benar sebanyak 31 sampel kategori rendah, 55 sampel kategori sedang, dan 39 sampel kategori tinggi. Kesalahan klasifikasi terjadi pada beberapa sampel, terutama pada kategori sedang yang terklasifikasi sebagai rendah (3 sampel) dan tinggi (5 sampel).

Tabel 7. Confusion Matrix

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Rendah	31	4	1
Sedang	3	55	5
Tinggi	1	4	39

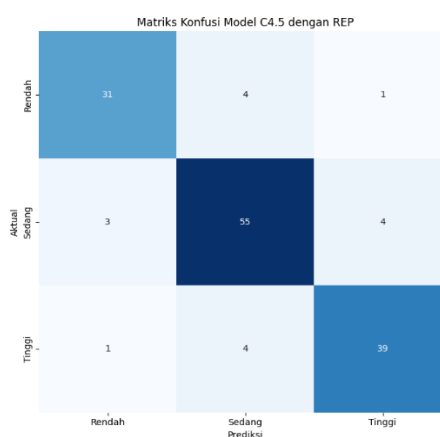
Berdasarkan matriks konfusi pada tabel 7, dapat dihitung metrik evaluasi kinerja model sebagai berikut:

- Akurasi: $(31 + 55 + 39) / 143 = 125 / 143 = 87,4\%$
- Presisi (Rendah): $31 / (31 + 3 + 1) = 31 / 35 = 88,6\%$
 - Presisi (Sedang): $55 / (4 + 55 + 4) = 55 / 63 = 87,3\%$
 - Presisi (Tinggi): $39 / (1 + 5 + 39) = 39 / 45 = 86,7\%$
 - Recall (Rendah): $31 / (31 + 4 + 1) = 31 / 36 = 86,1\%$
 - Recall (Sedang): $55 / (3 + 55 + 5) = 55 / 63 = 87,3\%$
 - Recall (Tinggi): $39 / (1 + 4 + 39) = 39 / 44 = 88,6\%$
 - F1-Score (Rendah): $2 \cdot (88,6\% \cdot 86,1\%) / (88,6\% + 86,1\%) = 87,3\%$
 - F1-Score (Sedang): $2 \cdot (87,3\% \cdot 87,3\%) / (87,3\% + 87,3\%) = 87,3\%$
 - F1-Score (Tinggi): $2 \cdot (86,7\% \cdot 88,6\%) / (86,7\% + 88,6\%) = 87,6\%$

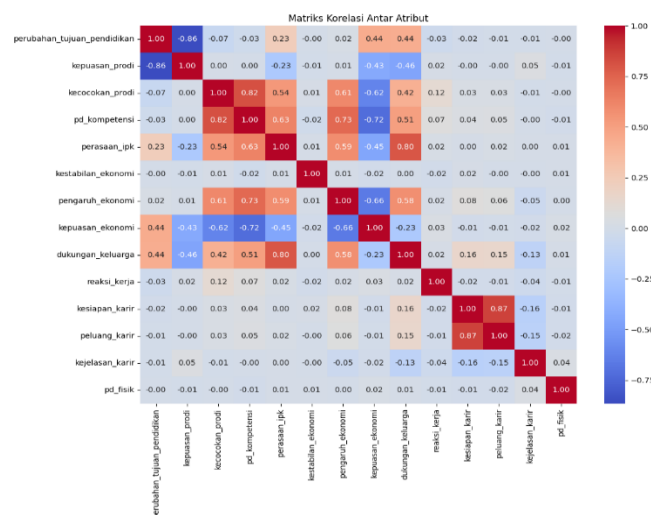
Tabel 8. Perbandingan Kinerja Model Sebelum dan Sesudah REP

Metrik Evaluasi	C4.5 Tanpa REP	C4.5 dengan REP
Akurasi	82,3%	87,4%
Presisi	81,5%	86,9%
Recall	80,8%	86,7%
F1-Score	81,1%	86,8%

Tabel 8 menunjukkan bahwa penerapan Reduced Error Pruning berhasil meningkatkan seluruh metrik evaluasi kinerja model. Akurasi meningkat dari 82,3% menjadi 87,4%, yang berarti model mampu mengklasifikasikan tingkat quarter life crisis dengan lebih tepat sebanyak 5,1% lebih baik. Peningkatan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa teknik pruning dapat mengurangi overfitting dan meningkatkan kemampuan generalisasi model. Visualisasi confusion matrix hasil pengujian ditampilkan pada gambar 3.



Gambar 3. Tampilan Visual Confusion Matrix Quarter Life Crisis



Gambar 4. Tampilan Matrix Korelasi

Kemudian setelah masuk ke confusion matrix masukan ke pengujian Matrix Korelasi dimana hasil matriks korelasi menunjukkan bahwa atribut kesiapan_karir memiliki korelasi tertinggi dengan tingkat quarter life crisis ($r = 0,45$), diikuti oleh kejelasan_karir ($r = 0,38$) dan reaksi_kerja ($r = 0,32$). Hal ini mengonfirmasi bahwa ketiga atribut tersebut merupakan prediktor utama dalam klasifikasi tingkat quarter life crisis. Hasil matriks korelasi tersebut ditampilkan pada gambar 4.

4. Kesimpulan

Berdasarkan hasil pengujian, model klasifikasi algoritma C4.5 dengan Reduced Error Pruning mampu mengklasifikasikan tingkat quarter life crisis mahasiswa dengan akurasi 87,4%, disertai nilai presisi, recall, dan F1-Score yang konsisten di atas 86% pada seluruh kategori. Penerapan REP terbukti efektif karena meningkatkan akurasi sebesar 5,1% dibandingkan model tanpa pruning. Atribut kesiapan karir menjadi prediktor paling dominan, diikuti kejelasan karir dan reaksi terhadap tekanan kerja, sehingga faktor persiapan dan kejelasan karir merupakan penentu utama tingkat quarter life crisis yang dialami mahasiswa.

Model ini dapat dimanfaatkan sebagai alat deteksi dini bagi mahasiswa untuk mengenali risiko krisis sejak awal masa perkuliahan, sekaligus sebagai dasar bagi konselor dan unit layanan bimbingan karir dalam merancang intervensi yang lebih tepat sasaran, seperti workshop perencanaan karir, pengembangan soft skills, dan konseling kelompok bagi kelompok rentan. Penelitian selanjutnya disarankan memperluas atribut dengan variabel psikologis lain seperti dukungan sosial, stres akademik, dan kondisi ekonomi keluarga, menggunakan data longitudinal untuk menangkap dinamika krisis secara temporal, serta membandingkan kinerja C4.5 dengan algoritma lain seperti Random Forest, Support Vector Machine, dan Naïve Bayes maupun teknik ensemble learning guna meningkatkan akurasi dan stabilitas model.

Referensi

- [1] Y. Wulandari, E. Haerani, S. Kurnia Gusti, and D. S. Ramadhani, "Klasifikasi Berita Menggunakan Algoritma C4.5," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 2, 2022, doi: [10.32672/jnkti.v5i2.4194](https://doi.org/10.32672/jnkti.v5i2.4194).
- [2] K. Valentino and D. Hendrawan, "Tinjauan Sistematis: Gambaran Quarter-life Crisis, Dampak, serta Faktor-faktor yang Memengaruhinya," *Buletin Psikologi*, vol. 33, no. 1, p. 26, Jun. 2025, doi: [10.22146/buletinpsikologi.98848](https://doi.org/10.22146/buletinpsikologi.98848).
- [3] F. Nur Anifah and A. Rohman, "Penerapan Algoritma C4.5 dalam Memprediksi Kepuasan Siswa Terhadap Kinerja Guru SMK," *Jurnal Algoritma*, vol. 23, no. 1, May 2026, doi: [10.33364/algoritma/v.23-1.3203](https://doi.org/10.33364/algoritma/v.23-1.3203).
- [4] D. W. S. Azahra, W. E. Setyowati, and B. Febriana, "The Relationship between Self-Disclosure and Quarter-Life Crisis among Nursing Students," *Journal of Medical Practice and Research*, no. 2, pp. 61–69, Dec. 2025, doi: [10.65310/r4f7ka29](https://doi.org/10.65310/r4f7ka29).
- [5] E.-T. R. Widaad, Arbin Janu Setiyowati, and Diniy Hidayatur Rahman, "Hubungan Dukungan Sosial dan Regulasi Emosi dengan Quarter Life Crisis Mahasiswa," *Buletin Konseling Inovatif*, vol. 3, no. 3, pp. 203–216, Oct. 2023, doi: [10.17977/um059v3i32023p203-216](https://doi.org/10.17977/um059v3i32023p203-216).
- [6] S. Pazer, "Prevalence, predictors, and psychosocial correlates of the quarter-life crisis in early adulthood," *Discover Psychology*, vol. 6, no. 1, Dec. 2026, doi: [10.1007/s44202-026-00781-w](https://doi.org/10.1007/s44202-026-00781-w).
- [7] H. Santoso, D. Desliani, P. Studi Ilmu Komputer, and F. Sains dan Teknologi, "Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naïve Bayes Classifier Student Sentiment Analysis about Face-To-Face Learning using Naïve Bayes Classifier," Aug. 2022, doi: [10.33633/tc.v21i3.6262](https://doi.org/10.33633/tc.v21i3.6262).
- [8] F. O. D. S. Budi, Y. Wandansari, and D. Tedjawidjaja, "Florentia Olivia Dermawan Setia Budi, Yettie Wandasari, Detricia Tedjawidjaja : Quarter Life Crisis dan Self-Efficacy pada Lulusan Perguruan Tinggi Usia Dewasa Awal yang Mengganggu Relationship Dewasa Awal," *Experientia: Jurnal Psikologi Indonesia*, vol. 13, no. 2, pp. 195–210, Dec. 2025, doi: [10.33508/exp.v13i2.7889](https://doi.org/10.33508/exp.v13i2.7889).
- [9] A. Kautsar, "Classification of Korean Drama Popularity Based on Ratings Using Naïve Bayes," *Journal of Computers and Digital Business*, vol. 5, no. 1, pp. 23–30, Jan. 2026, doi: [10.56427/jcbd.v5i1.814](https://doi.org/10.56427/jcbd.v5i1.814).

- [10] L. P. Simanjuntak, "Penerapan Data Mining Dalam Proses Hukum Pidana Bagi Pelaku Kekerasan Pada Wanita Menggunakan Algoritma C4.5," *ADA Journal of Information System Research*, vol. 1, no. 2, pp. 79–92, Feb. 2024, doi: [10.64366/adajisr.v1i2.41](https://doi.org/10.64366/adajisr.v1i2.41).
- [11] Rizki Nur Saadah Lingga and Nita Maharani Harahap, "Data Mining Analisis Kecendrungan Minat Pembeli Menggunakan Algoritma Apriori Pada Outlet Mao Mao Thai Tea," *Journal of Computers and Digital Business*, vol. 3, no. 2, pp. 50–58, May 2024, doi: [10.56427/jcbd.v3i2.398](https://doi.org/10.56427/jcbd.v3i2.398).
- [12] J. D. Atwood and C. Scholtz, "The Quarter-life Time Period: An Age of Indulgence, Crisis or Both?," *Contemp. Fam. Ther.*, vol. 30, no. 4, pp. 233–250, 2008, doi: [10.1007/s10591-008-9066-2](https://doi.org/10.1007/s10591-008-9066-2).
- [13] D. Made Aryadi Mertha Sanjaya, A. A. Istri Ita Paramitha, and N. Widya Utami, "Penerapan Data Mining untuk Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma C4.5: Studi Kasus STMIK Primakara," Jun. 2022, doi: [10.22437/jiutuj.v6i1.19600](https://doi.org/10.22437/jiutuj.v6i1.19600).
- [14] M. H. Sukri and Y. Handrianto, "Penerapan Algoritma C4.5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta," Feb. 2024, doi: [10.31294/iccj.v4i1.2582](https://doi.org/10.31294/iccj.v4i1.2582).
- [15] A. N. Aziziah, A. I. Purnamasari, and I. Ali, "PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI STOK BAHAN MINUMAN DI CAFE SEMANIS," Feb. 2024, doi: [10.36040/jati.v8i1.8347](https://doi.org/10.36040/jati.v8i1.8347).
- [16] E. Devia, "Penerapan Decision Tree Dengan Algoritma C4.5 Untuk Menentukan Rekomendasi Kenaikan Jabatan Karyawan," *JURNAL INFORMATION SYSTEM*, vol. 3, no. 1, pp. 028–037, May 2023, doi: [10.61488/jis.v3i1.257](https://doi.org/10.61488/jis.v3i1.257).