

## Analisis Diagnosis Tingkat Kesehatan Mental Dengan Teknik Klasifikasi Algoritma C4.5

Erlin Elisa<sup>1</sup>, Pastima Simanjuntak<sup>2</sup>, Andi Maslan<sup>3</sup>

<sup>1</sup>Sistem Informasi, Teknik dan Komputer, Universitas Putera Batam, Kepulauan Riau

<sup>2</sup>Teknik Informatika, Teknik dan Komputer, Universitas Putera Batam, Kepulauan Riau

<sup>3</sup>Teknik Informatika, Teknik dan Komputer, Universitas Putera Batam, Kepulauan Riau

\*elin210110@gmail.com

### ABSTRAK

Kesehatan mental menjadi isu yang semakin penting di tengah meningkatnya tekanan dalam kehidupan modern, terutama di lingkungan kerja. Tuntutan pekerjaan, tekanan target, dan dinamika relasi sosial di tempat kerja dapat memicu stres yang berdampak negatif terhadap produktivitas dan kesejahteraan karyawan. Namun, rendahnya tingkat kesadaran serta stigma sosial terhadap gangguan mental seringkali menyebabkan stres tidak terdeteksi secara dini. Penelitian ini bertujuan untuk mengidentifikasi tingkat stres karyawan pada salah satu perusahaan di Kota Batam dengan menggunakan pendekatan data mining. Data diperoleh melalui penyebaran kuesioner berbasis DASS-21 (*Depression Anxiety Stress Scales*) yang mengukur tiga aspek utama: stres, kecemasan, dan depresi. Selanjutnya, data dianalisis menggunakan bahasa pemrograman *Python*, dengan tahapan meliputi preprocessing, transformasi nilai skala menjadi bentuk numerik, serta pembangunan model klasifikasi menggunakan algoritma C4.5 (*Decision Tree Classifier*). Hasil penelitian menunjukkan bahwa model klasifikasi mampu mengidentifikasi tingkat stres dengan akurasi sebesar 67%. Kinerja terbaik ditunjukkan pada kelas stres sedang (kelas 1), dengan nilai precision sebesar 0,71 dan recall 0,79. Sebaliknya, performa klasifikasi pada kelas minoritas seperti tidak stres (kelas 0) dan stres berat (kelas 2) masih tergolong rendah. Temuan ini menunjukkan bahwa algoritma C4.5 cukup efektif dalam mengenali pola stres dominan, namun diperlukan pengolahan data lanjutan dan teknik penyeimbangan kelas untuk meningkatkan akurasi keseluruhan model. Penelitian ini diharapkan dapat menjadi dasar dalam upaya deteksi dini dan penanganan stres di lingkungan kerja secara lebih tepat.

**Kata kunci:** Kesehatan mental, Data mining, DASS-21, Algoritma C4.5

### ABSTRACT

*Mental health has become an increasingly important issue amid the growing pressures of modern life, particularly in the workplace. Job demands, performance targets, and the dynamics of social relationships at work can trigger stress that negatively affects employee productivity and well-being. However, low awareness and social stigma surrounding mental health issues often result in stress going undetected at an early stage. This study aims to identify employee stress levels at a company in Batam City using a data mining approach. Data were collected through the distribution of questionnaires based on the DASS-21 (Depression Anxiety Stress Scales), which measures three main aspects: stress, anxiety, and depression. The data were analyzed using the Python programming language, with stages including preprocessing, transformation of scale values into numerical form, and the construction of a classification model using the C4.5 algorithm (Decision Tree Classifier). The results showed that the classification model was able to identify stress levels with an accuracy of 67%. The best performance was observed in the moderate stress class (class 1), with a precision value of 0.71 and a recall of 0.79. In contrast, the classification performance for minority classes such as no stress (class 0) and severe stress (class 2) was relatively low. These findings suggest that the C4.5 algorithm is reasonably effective in recognizing dominant stress patterns but requires further data processing and class-balancing techniques to improve overall model performance. This study is expected to serve as a foundation for early detection and more accurate handling of workplace stress.*

**Keywords:** Mental health, Data mining, DASS-21, C4.5 algorithm.

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Kesehatan mental kini menjadi isu global yang penting akibat tekanan hidup, gaya hidup modern, dan dampak pandemi COVID-19. Ketiganya menjadikan kesehatan mental kunci bagi kesejahteraan dan produktivitas, namun upaya pencegahan masih terhambat oleh stigma negatif dan rendahnya pemahaman masyarakat.

Kesehatan mental merupakan aspek krusial dalam pembangunan sumber daya manusia yang berkualitas, khususnya bagi generasi muda yang menjadi kunci dalam mewujudkan Visi Indonesia Emas 2045 (Nisa et al., 2024). Di sisi lain, akses terhadap layanan kesehatan mental di Indonesia masih sangat terbatas, terutama di daerah-daerah berkembang. Banyak wilayah belum memiliki fasilitas layanan kesehatan jiwa yang memadai, dan jika pun tersedia, biayanya seringkali tidak terjangkau oleh sebagian besar masyarakat (Hendra et al., 2023). Hal ini berdampak pada rendahnya deteksi dan penanganan gangguan mental secara komprehensif.

Kota Batam sebagai kawasan industri strategis menghadapi tantangan serius dalam isu kesehatan mental pekerja akibat lingkungan kerja padat, target produksi ketat, dan beban kerja tinggi yang memicu stres hingga depresi. Data WHO (2021) mencatat prevalensi gangguan mental di Indonesia sebesar 9,8%, dengan depresi 6,6%, dan angka ini terus meningkat akibat dampak psikologis pandemi. Penelitian Onie et al. (2024) juga menunjukkan bahwa Provinsi Kepulauan Riau mencatat angka percobaan bunuh diri tertinggi secara nasional, serta tingkat bunuh diri tak tercatat tertinggi di dunia. Meningkatnya angka gangguan mental menuntut pendekatan yang lebih efektif dalam deteksi dini dan pengambilan keputusan berbasis data. Salah satu pendekatan yang menjanjikan adalah pemanfaatan data mining, yakni proses untuk menggali informasi tersembunyi dari data melalui identifikasi pola dan relasi yang tidak dapat dikenali secara manual (Nugraha et al., 2022). Dalam konteks ini, algoritma klasifikasi seperti C4.5 dapat digunakan untuk membangun

model prediktif yang membantu mengidentifikasi tingkat risiko stres dan gangguan mental pada pekerja. Algoritma C4.5 dipilih karena kemampuannya menghasilkan pohon keputusan yang interpretatif serta akurasi yang kompetitif dalam klasifikasi data kategorikal (Andarista & Jananto, 2022).

Penelitian ini bertujuan untuk mengkaji penerapan algoritma klasifikasi C4.5 pada data primer yang dikumpulkan dari pekerja di Kota Batam, dengan fokus pada identifikasi tingkat stres kerja. Melalui pendekatan ini, diharapkan dapat diperoleh model yang tidak hanya akurat, tetapi juga mudah dipahami oleh pengambil kebijakan dalam merancang intervensi kesehatan mental di lingkungan kerja secara lebih efektif dan berbasis data.

### 1.2 Landasan Teori

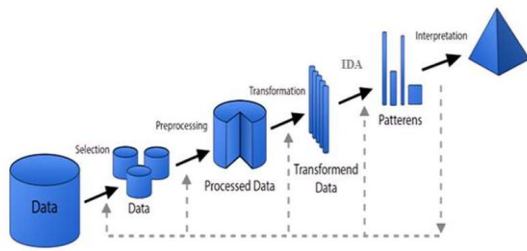
#### Kesehatan Mental

Kesehatan mental merupakan aspek krusial dalam mencapai kesehatan yang menyeluruh, namun masih kurang menjadi prioritas di negara berkembang dibandingkan penyakit menular (Ridlo, 2020). Kesehatan mental yang terganggu dapat menghambat individu menjalani kehidupan sehat sesuai harapan pribadi maupun sosial (Setyaningrum & Yanuarita, 2020). Menurut Federasi Kesehatan Mental Dunia, kesehatan mental adalah kondisi yang memungkinkan perkembangan optimal secara fisik, intelektual, dan emosional sesuai dengan lingkungan sosial (Supini et al., 2024).

#### Knowledge Discovery in Database (KDD)

KDD (*Knowledge Discovery in Database*) adalah proses terorganisir untuk menemukan pengetahuan yang valid, berguna, dan dapat dipahami dari kumpulan data besar yang saling terhubung dalam basis data. Proses ini mencakup pengumpulan dan analisis data untuk mendukung pengambilan keputusan. Inti dari KDD adalah data mining, yaitu penggunaan algoritma untuk mengevaluasi data, membangun model, dan mengidentifikasi pola tersembunyi yang sebelumnya tidak diketahui (Leza et al., 2024; Brodley et al., 2020; Al-khateeb

&Science,2021).



Gambar 1. *Proses Knowledge Discovery in Database (KDD)*

### **Datamining**

Data mining adalah proses menganalisis dan mengekstraksi informasi bermanfaat dari kumpulan data berukuran besar untuk menemukan pola, hubungan, atau pengetahuan tersembunyi yang dapat mendukung pengambilan keputusan. Proses ini memanfaatkan data historis untuk meningkatkan efisiensi, mengurangi biaya, serta mengidentifikasi informasi yang sebelumnya tidak diketahui (Wahyudi et al., 2022; Manullang et al., 2021; Gaol, 2020).

### **Klasifikasi**

Mengukur tingkat akurasi pada dataset dapat dilakukan dengan teknik klasifikasi datamining, klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya kedalam kelas tertentu dari jumlah kelas yang tersedia (Utomo & Mesran, 2020). Klasifikasi merupakan salah satu teknik data mining yang bertujuan untuk menggambarkan suatu model pengetahuan berdasarkan ciri persamaan atau perbedaan dari suatu kelas data (Novika et al., 2021). Klasifikasi data adalah proses menemukan model atau fungsi yang menjelaskan dan membedakan kelas data serta konsepnya. Peneliti akan melakukan komparasi beberapa algoritma untuk menemukan akurasi tertinggi (Putri & Wijayanto, 2022).

### **Algoritma C4.5**

Algoritma C4.5 adalah metode supervised learning yang digunakan untuk membentuk pohon keputusan (decision tree) berdasarkan data, dan merupakan pengembangan dari algoritma ID3 yang diperkenalkan oleh J. Ross Quinlan.

Algoritma ini mampu menangani missing value, data kontinu, dan melakukan pruning. C4.5 bekerja dengan menghitung entropy dan gain untuk menentukan akar pohon dan mempartisi data berdasarkan tupel. Karena kemampuannya dalam mengklasifikasikan data secara efektif, C4.5 menjadi salah satu algoritma populer dalam pemodelan klasifikasi (Adriansa et al., 2022; Sepharni et al., 2022; Marlina & Bakri, 2021).

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Rumus 1. Menghitung *Gain*

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Rumus 2. Menghitung *Entropy*

### **Python**

*Python* adalah bahasa pemrograman tingkat tinggi berorientasi objek yang dikembangkan oleh Guido van Rossum dan dirilis pada 1991. *Python* bersifat multi-platform, mudah dipelajari, serta mendukung integrasi dengan bahasa dan alat lain. Bahasa ini populer karena fleksibilitasnya dalam berbagai bidang, termasuk pengembangan perangkat lunak, machine learning, dan deep learning. *Python* juga dilengkapi pustaka standar yang luas dan mendukung berbagai kebutuhan pemrograman modern (Prihatmono & Watratan, 2019; Alfarizi et al., 2023; Gunawan & Yelmi, 2021).

### **1.3 State Of Art**

Penelitian terkait penerapan algoritma data mining dalam bidang kesehatan mental terus berkembang seiring meningkatnya kebutuhan akan sistem deteksi dini yang akurat, efisien, dan berbasis data. Salah satu algoritma yang sering digunakan adalah C4.5 (*Decision Tree Classifier*). Suci & Milkhatun (2020) serta Treise et al. (2019) memanfaatkan algoritma C4.5 untuk menganalisis rekam medis pasien di Rumah Sakit Jiwa Atma Husada Mahakam Samarinda. Penelitian ini menunjukkan bahwa pasien dengan risiko kekerasan dan gangguan waham paling

dominan ditemukan pada kelompok usia dewasa muda dan berjenis kelamin laki-laki. Kedua studi ini menegaskan kemampuan algoritma C4.5 dalam menghasilkan pohon keputusan yang mudah diinterpretasikan dari data rekam medis.

Selain algoritma C4.5, metode machine learning lainnya juga telah dievaluasi untuk mendeteksi gangguan mental. Furqan et al. (2020) mengevaluasi performa Support Vector Machine (SVM) dalam mendiagnosis penyakit autisme, bipolar, dan skizofrenia berdasarkan data Genome-Wide Association Studies (GWAS), dengan hasil bahwa kernel Gaussian menunjukkan kinerja terbaik. Di sisi lain, penelitian oleh Nanda & Sari (2020) menerapkan algoritma Naive Bayes dalam mendeteksi gangguan kepribadian manusia, dengan hasil akurasi tertinggi sebesar 88,8% pada pengujian 10-fold cross validation. Sagala & Tampubolon (2018) juga menunjukkan keunggulan algoritma Naive Bayes dalam klasifikasi data konsumsi alkohol pada siswa setelah dilakukan pemilihan fitur menggunakan Gain Ratio.

Pendekatan lain ditunjukkan oleh Hikmatulloh et al. (2019) yang menggunakan algoritma ID3 dalam mendiagnosis kesehatan kehamilan dan berhasil mencapai akurasi 80,33%, serta telah diimplementasikan dalam aplikasi mobile berbasis Android. Hal ini menunjukkan potensi algoritma pohon keputusan tidak hanya dalam klasifikasi data tetapi juga dalam pengembangan sistem pendukung keputusan yang aplikatif.

Nisa et al. (2024) menerapkan algoritma LightGBM dalam mengklasifikasikan gangguan mental berdasarkan data survei OSMI. Hasilnya menunjukkan akurasi pelatihan sebesar 93% dan pengujian sebesar 82%, membuktikan efektivitas LightGBM dalam klasifikasi data berskala besar. Studi oleh Studi et al. (2023) juga menunjukkan tren serupa, yaitu dominasi gangguan mental pada laki-laki berusia 12–25 tahun, menekankan pentingnya intervensi dini terhadap kelompok usia remaja.

Secara umum, studi-studi terdahulu menunjukkan bahwa berbagai algoritma klasifikasi memiliki potensi dalam mengolah data kesehatan mental, masing-masing dengan kelebihan dalam konteks data dan tujuan yang berbeda. Namun demikian, masih terdapat tantangan dalam hal keterbatasan data primer, ketidakseimbangan kelas, dan interpretasi hasil yang aplikatif untuk pengambilan keputusan kesehatan di lapangan. Oleh karena itu, penelitian ini hadir untuk mengkaji lebih lanjut pemanfaatan algoritma C4.5 terhadap data kesehatan mental pekerja di Kota Batam guna menghasilkan model klasifikasi yang akurat dan mudah diterapkan di lingkungan kerja.

## 2. METODE PELAKSANAAN

Berikut adalah desain penelitian yang akan dijalankan dalam penelitian ini:



Gambar 2. Desain Penelitian

Penelitian ini dirancang untuk menganalisis dampak kesehatan mental terhadap pekerja di Kota Batam, khususnya pada sektor industri. Permasalahan utama yang diangkat adalah bagaimana kondisi kesehatan mental memengaruhi produktivitas dan keseimbangan hidup antara pekerjaan dan kehidupan pribadi. Pendekatan yang digunakan adalah metode data mining dengan teknik klasifikasi, memanfaatkan algoritma C4.5 untuk membangun model prediksi tingkat gangguan mental. Pengumpulan data dilakukan melalui penyebaran kuesioner berbasis DASS-21 (*Depression Anxiety Stress Scales*), yang mengukur tiga indikator utama: depresi, kecemasan, dan stres. Objek penelitian

difokuskan pada para pekerja pabrik di Kota Batam. Analisis data akan dilakukan menggunakan bahasa pemrograman Python, yang digunakan dalam tahap preprocessing, klasifikasi, dan evaluasi akurasi model. Hasil yang diharapkan adalah model klasifikasi yang mampu memprediksi tingkat gangguan mental pekerja berdasarkan indikator-indikator DASS-21, sehingga dapat digunakan sebagai dasar deteksi dini dan pengambilan keputusan dalam manajemen kesehatan mental di lingkungan kerja.

### 3. HASIL DAN PEMBAHASAN HASIL

Hasil analisis data dari kuesioner DASS-21 yang disebarkan kepada karyawan di salah satu perusahaan di Kota Batam diolah menggunakan Python untuk preprocessing, transformasi data, dan pembuatan model klasifikasi dengan algoritma C4.5. Tujuannya adalah untuk mengidentifikasi tingkat stres karyawan berdasarkan respons terhadap indikator-

indikator stres. Hasil analisis disajikan dalam bentuk pohon keputusan, metrik evaluasi model, dan interpretasi pola klasifikasi yang terbentuk.

#### 1. Analisis Data

Penelitian ini menggunakan kuesioner DASS-21 untuk mengidentifikasi tingkat stres karyawan di salah satu perusahaan di Kota Batam, dengan fokus pada 7 item pertanyaan terkait stres. Respon dikodekan menggunakan skala Likert 4 poin dan diolah secara numerik. Setelah melalui proses pembersihan dan transformasi data menggunakan Python, skor stres dikategorikan ke dalam lima tingkat sesuai pedoman DASS-21. Data tersebut kemudian digunakan untuk membangun model klasifikasi dengan algoritma C4.5 guna memprediksi tingkat stres berdasarkan pola jawaban responden.

Adapun pertanyaan-pertanyaan untuk setiap variable dapat dilihat pada tabel di bawah ini:

**Tabel 1.** Pertanyaan Kuisisioner DASS-21

Variable	Code	Question
Depression	QD-01	<i>I couldn't seem to experience any positive feeling at all.</i>
	QD-02	<i>I found it difficult to work up the initiative to do things.</i>
	QD-03	<i>I felt that I had nothing to look forward to.</i>
	QD-04	<i>I felt down-hearted and blue.</i>
	QD-05	<i>I was unable to become enthusiastic about anything.</i>
	QD-06	<i>I felt I wasn't worth much as a person.</i>
	QD-07	<i>I felt that life was meaningless.</i>
Anxiety	QA-01	<i>I was aware of dryness of my mouth.</i>
	QA-02	<i>I experienced breathing difficulty.</i>
	QA-03	<i>I experienced trembling (eg, in the hands).</i>



Variable	Code	Question
Stress	QA-04	<i>I was worried about situations in which I might panic and make a fool of myself.</i>
	QA-05	<i>I felt I was close to panic.</i>
	QA-06	<i>I was aware of the action of my heart in the absence of physical exertion.</i>
	QA-07	<i>I felt scared without any good reason.</i>
	QS-01	<i>I found it hard to wind down.</i>
	QS-02	<i>I tended to over-react to situations.</i>
	QS-03	<i>I felt that I was using a lot of nervous energy.</i>
	QS-04	<i>I found myself getting agitated.</i>
	QS-05	<i>I found it difficult to relax.</i>
	QS-06	<i>I was intolerant of anything that kept me from getting on with what I was doing.</i>
	QS-07	<i>I felt that I was rather touchy.</i>

**Preprocessing Data**

Pada tahap pertama, dilakukan pemeriksaan terhadap data hasil kuesioner untuk mengidentifikasi adanya data kosong (*missing values*), duplikasi, atau entri yang tidak valid. Setelah data dibersihkan, dilakukan proses transformasi jawaban kuesioner ke bentuk numerik, di mana skala Likert dari DASS-

21 dikonversi menjadi angka dengan ketentuan:

Tidak Pernah = 0

Kadang-kadang = 1

Sering = 2

Hampir Selalu = 3

Berikut adalah sampel dari data isian data kuesioner yang telah di kodekan:

**Tabel 2.** Representasi Proses Kuantifikasi

ID	Depression							Anxiety							Stress						
	QD-01	QD-02	QD-03	QD-04	QD-05	QD-06	QD-07	QA-01	QA-02	QA-03	QA-04	QA-05	QA-06	QA-07	QS-01	QS-02	QS-03	QS-04	QS-05	QS-06	QS-07
1	0	1	1	1	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1	1	0
2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3	2	3	2	3	2	3	3	0	3	2	2	1	3	3	3	2	3	3	3
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	2	2	2	2	1	1	2	1	2	2	2	0	1	2	2	2	2	2
6	1	1	2	2	2	2	2	1	1	1	2	2	1	2	1	1	2	2	2	2	1
7	0	0	0	1	0	2	1	3	1	3	1	2	3	3	0	1	3	0	1	0	3
8	1	3	0	1	1	1	2	0	2	1	0	1	3	3	0	1	3	0	0	1	1
9	1	1	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	1	1	1	1
10	1	1	1	1	1	0	0	1	0	0	0	0	1	0	1	1	0	1	1	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
369 9	1	0	0	0	1	0	0	2	0	1	1	0	1	0	2	1	1	1	0	0	1

```
# 1. Load data
df = pd.read_csv('tes.csv', sep=';')
# Coba ganti dengan sep=';' jika error
print("Data Loaded:")
print(df.head())
print(df.columns)
```

Gambar 4. Load Data

```
Data Loaded:
I couldn't seem to experience any positive feeling at all. \
0          Never / Tidak Pernah
1      Almost Always / Hampir Selalu
2      Almost Always / Hampir Selalu
3      Sometimes / Kadang-kadang
4      Sometimes / Kadang-kadang

I found it difficult to work up the initiative to do things. \
0      Sometimes / Kadang-kadang
1      Almost Always / Hampir Selalu
2      Almost Always / Hampir Selalu
3      Sometimes / Kadang-kadang
4      Sometimes / Kadang-kadang
```

Gambar 5. Data Loaded

## 2. Mapping Nilai

Mengubah atau mengganti nilai-nilai dalam suatu kolom (biasanya dalam DataFrame) berdasarkan aturan yang telah ditetapkan sebelumnya.

```
# 2. Mapping nilai
mapping = {
    'Never / Tidak Pernah': 0,
    'Sometimes / Kadang-kadang': 1,
    'Often / Sering': 2,
    'Almost Always / Hampir Selalu': 3
}
```

Gambar 5. Mapping Nilai

Sehingga hasil preprosesing data dapat dilihat sebagai berikut.

```
Cek data akhir:
X shape: (3699, 20)
y shape: (3699,)
X preview:
I couldn't seem to experience any positive feeling at all. \
0          0
1          3
2          3
3          1
4          1

I found it difficult to work up the initiative to do things. \
0          1
1          3
2          3
3          1
4          1
```

Gambar 6. Hasil preprosesing

## 2. Analisis Algoritma C4.5

Setelah tahap pre prosesing dan cleaning data dilakukan langkah selanjutnya yang dilakukan adalah melatih model dengan data agar bisa mempelajari pola hubungan antara fitur dan label. Penelitian ini akan menggunakan algoritma C4.5.

## # 7. Training model

```
clf = DecisionTreeClassifier(criterion='entropy')
clf = clf.fit(X, y)
```

Gambar 7. Training Model C4.5  
Adapun Output klasifikasi yang dihasilkan adalah.

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix, ConfusionMatrixDisplay
from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

# Split data: 80% train, 20% test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# Training model C4.5
clf = DecisionTreeClassifier(criterion='entropy')
clf.fit(X_train, y_train)

# Prediksi
y_pred = clf.predict(X_test)

# Classification report
print("Classification Report:")
print(classification_report(y_test, y_pred))

=== VISUALISASI CONFUSION MATRIX ===
cm = confusion_matrix(y_test, y_pred)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=clf.classes_)
disp.plot(cmap=plt.cm.Blues)
plt.title("Confusion Matrix")
plt.show()

=== VISUALISASI DECISION TREE ===
plt.figure(figsize=(20, 10))
plot_tree(clf, feature_names=X.columns, class_names=[str(cls) for cls in clf.classes_], filled=True)
plt.title("Visualisasi Pohon Keputusan (C4.5)")
plt.show()
```

Gambar 8. Training Model C4.5  
Dengan hasil Clasiffication Report :

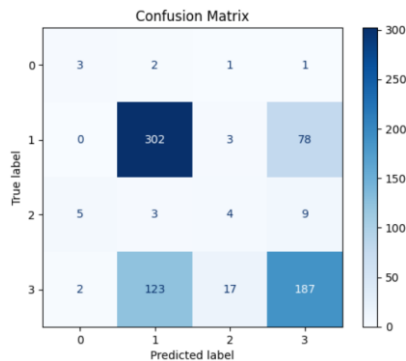
Classification Report:					
	precision	recall	f1-score	support	
0	0.30	0.43	0.35	7	
1	0.70	0.79	0.74	383	
2	0.16	0.19	0.17	21	
3	0.68	0.57	0.62	329	
accuracy			0.67	740	
macro avg	0.46	0.49	0.47	740	
weighted avg	0.67	0.67	0.67	740	

Gambar 9. Classification Report

Berdasarkan evaluasi model klasifikasi menggunakan algoritma C4.5 terhadap data DASS-21, diperoleh akurasi keseluruhan sebesar 67%, dengan performa terbaik pada kelas 1 (ringan) yang merupakan kategori terbanyak, ditunjukkan oleh precision 0.70, recall 0.79, dan f1-score 0.74. Model juga cukup baik dalam mengklasifikasikan kelas 3 (berat) dengan f1-score 0.62, namun menunjukkan performa rendah pada kelas 0 (normal) dan kelas 2 (sedang), dengan f1-score masing-masing hanya 0.35 dan

0.17, yang kemungkinan disebabkan oleh ketidakseimbangan data. Rata-rata makro menunjukkan bahwa performa model tidak merata antar kelas (f1-score 0.47), sedangkan weighted average yang memperhitungkan proporsi data tiap kelas konsisten dengan akurasi sebesar 0.67.

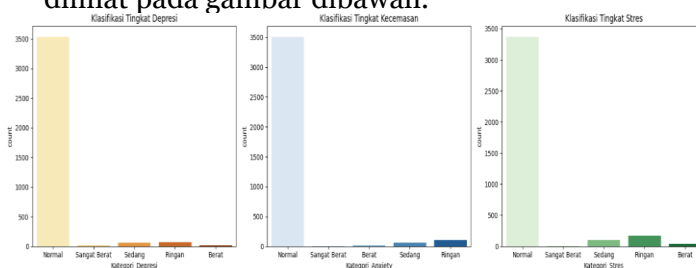
Untuk lebih jelasnya dapat dilihat pada grafik confusion matrix berikut.



Gambar 10. Confusion Matrix

## PEMBAHASAN

Sesuai dengan solusi awal pada penelitian ini bahwasanya Analisis akan dilakukan dengan mengklasifikasikan data berdasarkan kategori DASS-21 yaitu kelas Anxiety, Depresi dan Strees, sehingga output akhir dari klasifikasi dengan menambahkan coding untuk logika pengelompokan kasus untuk lebih jelasnya dilihat pada gambar dibawah.



Gambar 11. DASS-21 Classification

Berdasarkan hasil klasifikasi tingkat depresi, kecemasan, dan stres, mayoritas responden berada dalam kategori "Normal", menunjukkan kondisi kesehatan mental yang secara umum masih baik. Namun, terdapat kecenderungan peningkatan pada kategori ringan dan sedang, terutama pada aspek stres yang menunjukkan jumlah lebih tinggi dibandingkan depresi dan kecemasan. Meskipun kategori berat dan sangat berat muncul dalam jumlah kecil, temuan ini mengindikasikan adanya sebagian individu yang mulai mengalami

gangguan psikologis. Oleh karena itu, diperlukan perhatian dan dukungan psikologis dari pihak terkait untuk mencegah berkembangnya gangguan mental yang lebih serius.

## 4. KESIMPULAN

Berdasarkan hasil pengolahan dan analisis data dari kuesioner DASS-21 yang diterapkan pada karyawan di Kota Batam, serta pemanfaatan algoritma C4.5 berbasis pemrograman Python untuk klasifikasi kondisi mental, maka dapat disimpulkan:

Pemanfaatan algoritma C4.5 dengan implementasi Python berhasil mengolah data kuesioner DASS-21 dan mengklasifikasikan tingkat depresi, kecemasan, dan stres responden, dengan hasil mayoritas berada pada kategori Normal. Distribusi paling bervariasi terlihat pada kategori stres, yang mengindikasikan dampak tekanan kerja lebih signifikan. Model menunjukkan akurasi klasifikasi sebesar 67% dan cukup efektif dalam mengenali kategori dominan seperti Normal dan Ringan, khususnya pada dimensi kecemasan dan stres. Namun, performanya masih kurang optimal pada kategori Sedang dan Berat akibat ketidakseimbangan jumlah data, sehingga diperlukan perbaikan kualitas dan distribusi data untuk meningkatkan akurasi model.

## DAFTAR PUSTAKA

- Adriansa, M., Yulianti, L., & Elfianty, L. (2022). Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 07(21), 115–121. <https://doi.org/10.54367/jtiust.v7i1.1983>
- Al-khateeb, M. O., & Science, C. (2021). Intelligent Data Analysis approaches for Knowledge Discovery: Survey and challenges. 20(5), 1782–1792. <https://doi.org/10.17051/ilkonline.2021.05.196>
- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning.



- Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID), 2(1), 1–6.
- Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, 16(2), 29. <https://doi.org/10.33365/jtk.v16i2.1525>
- Atalya Angelus Leza, M., Widya Utami, N., & Anugrah Cahya Dewi, P. (2024). Prediksi Prestasi Siswa Smas Katolik Santo Yoseph Denpasar Berdasarkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery in Database Dan Algoritma Regresi Linier Berganda. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 373–379. <https://doi.org/10.36040/jati.v8i1.8754>
- Brodley, C. E., Lane, T., & Stough, T. M. (2020). Knowledge discovery and data mining. *American Scientist*, 87(1), 54–61. <https://doi.org/10.1511/1999.16.807>
- Febriani, S., & Sulistiani, H. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. *89Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(4), 89–95.
- Furqan, M., Kurniawan, R., & HP, K. I. (2020). Evaluasi Performa Support Vector Machine Classifier Terhadap Penyakit Mental. *Jurnal Sistem Informasi Bisnis*, 10(2), 203–210. <https://doi.org/10.21456/vol10iss2pp203-210>
- Gaol, N. Y. L. (2020). Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5. *Jurnal Informasi & Teknologi*, 2, 23–29. <https://doi.org/10.37034/jidt.v2i1.22>
- Gunawan, I., & Yelmi, Y. (2021). Rancang Bangun Robot Pengawas Dokumen Berbasis Raspberry Pi2 dengan Pemrograman Python. *Jurnal Ilmu Komputer Dan Bisnis*, 12(1), 144–149. <https://doi.org/10.47927/jikb.v12i1.99>
- Hendra, H., Muhaemin, M., & Santosa, S. (2023). Klasifikasi Pasien Gangguan Jiwa Menggunakan Algoritma C4. 5 Sebagai Dasar Pengambilan Keputusan Kesehatan Jiwa. *Prosiding Seminar Nasional* .... <https://jurnal.umj.ac.id/index.php/semnaslit/article/view/19400%0Ahttps://jurnal.umj.ac.id/index.php/semnaslit/article/download/19400/9465>
- Hikmatulloh, H., Rahmawati, A., Wintana, D., & Ambarsari, D. A. (2019). Penerapan Algoritma Iterative Dichotomiser Three (Id3) Dalam Mendiagnosa Kesehatan Kehamilan. *Klik - Kumpulan Jurnal Ilmu Komputer*, 6(2), 116. <https://doi.org/10.20527/klik.v6i2.189>
- Kukuh Wahyudi, A., Azizah, N., & Saputro, H. (2022). Data Mining Klasifikasi Kepribadian Siswa Smp Negeri 5 Jepara Menggunakan Metode Decision Tree Algoritma C4.5. *Journal of Information System and Computer*, 2(2), 8–13. <https://doi.org/10.34001/jjister.v2i2.392>
- Manullang, N., Sembiring, R. W., Gunawan, I., Parlina, I., & Irawan, I. (2021). Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5. *Jurnal Ilmu Komputer Dan Teknologi*, 2(2), 1–5. <https://doi.org/10.35960/ikomti.v2i2.700>
- Marlina, D., & Bakri, M. (2021). Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(1), 23–28.
- Nanda, Y. A. I., & Sari, B. W. (2020). Naive Bayes Algorithm Implementation To Detect Human Personality Disorders. *Jurnal Techno Nusa Mandiri*, 17(1), 9–16. <https://doi.org/10.33480/techno.v17i1.1239>
- Nisa, K., Informatika, P. S., & Bangsa, U. H. (2024). Klasifikasi Penyakit Gangguan Mental dengan Algoritma LightGBM. 9, 1086–1094.
- Novika, T., Poningsih, P., Okprana, H., Windarto, A. P., & Siahaan, H. (2021). Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika. *Jurnal Media Informatika Budidarma*, 5(1), 9.

- <https://doi.org/10.30865/mib.v5i1.2498>
- Nugraha, F. F., Sunandar, I., & Julian, C. (2022). Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan Algoritma C4.5. *Teknologi*, 7(March), 10–20.
- Nurjayadi, R., & Kristiana, T. (2019). Penerapan Association Rule Menggunakan Algoritma Apriori Untuk Analisa Penjualan AuFa Baby Shop. *Indonesia Journal on Computer and Information Technology (IJCIT)*, 4(September), 205–214.
- Onie, S., Usman, Y., Widyastuti, R., Lusiana, M., Angkasawati, T. J., Musadad, D. A., Nilam, J., Vina, A., Kamsurya, R., Batterham, P., Arya, V., Pirkis, J., & Larsen, M. (2024). Indonesia's first suicide statistics profile: an analysis of suicide and attempt rates, underreporting, geographic distribution, gender, method, and rurality. *The Lancet Regional Health - Southeast Asia*, 22, 100368. <https://doi.org/10.1016/j.lansea.2024.100368>
- Prihatmono, M. W., & Watratan, A. F. (2019). Implementasi Algoritma C4.5 Menggunakan Python Untuk Klasifikasi Kepuasan Konsumen. *Progres*, 49–55. <https://jurnal.stmikprofesional.ac.id/index.php/Progress/article/view/146/22>
- Putri, N. B., & Wijayanto, A. W. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, 11(1), 59–66. <https://doi.org/10.34010/komputika.v11i1.4350>
- Ridlo, I. A. (2020). Pandemi COVID-19 dan Tantangan Kebijakan Kesehatan Mental di Indonesia. *INSAN Jurnal Psikologi Dan Kesehatan Mental*, 5(2), 162. <https://doi.org/10.20473/jpkm.v5i220.162-171>
- Sagala, N., & Tampubolon, H. (2018). Komparasi Kinerja Algoritma Data Mining pada Dataset Konsumsi Alkohol Siswa. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, 4(2), 98–103. <https://doi.org/10.23917/khif.v4i2.7061>
- Sepharni, A., Hendrawan, I. E., & Rozikin, C. (2022). Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 7(2), 117. <https://doi.org/10.30998/string.v7i2.12012>
- Setyaningrum, W., & Yanuarita, H. A. (2020). Pengaruh Covid-19 Terhadap Kesehatan Mental Masyarakat Di Kota Malang. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, 4(4). <https://doi.org/10.58258/jisip.v4i4.1580>
- Suci, E. P. M. E., & Milkhatun. (2020). Analisis Rekam Medis Pasien Risiko Perilaku Kekerasan dengan Menggunakan Algoritma C4.5 di Rumah Sakit Jiwa Atma Husada Mahakam Samarinda. *Borneo Student Research*, 2(1), 16–24. <https://journals.umkt.ac.id/index.php/bsr/article/download/1423/636/>
- Supini, P., Gandakusumah, A. R. P., Asyifa, N., Auliya, Z. N., & Ismail, D. R. (2024). Faktor-Faktor yang Mempengaruhi Kesehatan Mental pada Remaja. *JERUMI: Journal of Education Religion Humanities and Multidiciplinary*, 2(1), 166–172. <https://doi.org/10.57235/jerumi.v2i1.1760>
- Treise, C., Brown, R. J., & Perez, J. (2019). Towards a Multi-Level Phenomenology of Delusional Disorder: The Dissociative Thought-Script. *Psychopathology*, 52(1), 50–58. <https://doi.org/10.1159/000499596>
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>