



## *Mental Health Analysis to Prevent Mental Disorders in Students Using The K-Nearest Neighbor (K-NN) Algorithm and Random Forest Algorithm*

### **Analisis Kesehatan Mental untuk Mencegah Gangguan Mental pada Mahasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Random Forest**

**Najib Nurdiansyah<sup>1\*</sup>, Farhan Sulis Febriyan<sup>2</sup>, Zanuvar Gesit Dian Amanta<sup>3</sup>,  
Dicky Arya Saputra<sup>4</sup>, Wiga Maulana Baihaqi<sup>5</sup>**

<sup>1,2,3,4,5</sup>Department of Informatics, Faculty of Computer and Science,  
Amikom Purwokerto University, Indonesia

E-Mail: <sup>1</sup>22sa11a058@mhs.amikompurwokerto.ac.id,  
<sup>2</sup>22sa11a107@mhs.amikompurwokerto.ac.id, <sup>3</sup>22sa11a1257@mhs.amikompurwokerto.ac.id,  
<sup>4</sup>22sa11a284@mhs.amikompurwokerto.ac.id, <sup>5</sup>wiga@amikompurwokerto.ac.id

Received Jul 14th 2024; Revised Sep 26th 2024; Accepted Oct 30th 2024; Available Online Nov 24<sup>th</sup> 2024

Corresponding Author: Najib Nurdiansyah

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

#### **Abstract**

*In the modern era, mental disorders have become a global health problem. The World Health Organization (WHO) estimates that one in four people worldwide will experience a mental or neurological disorder. Disturbances often occur in students, one of whom is a student. The mental health of students, who will be the nation's next generation, is very important for their success in academic and non-academic fields and their role in society in the future. By using the K-Nearest Neighbor (K-NN) and Random Forest algorithms, this research aims to analyze mental health to prevent mental disorders in students. The dataset "Student mental health.csv" was used, which was taken from the Kaggle website which includes various variables related to student health. The process used includes exploratory data analysis, data preprocessing, data modeling using K-Nearest Neighbor (K-NN) and Random Forest algorithms, and finally evaluation. The results show that K-NN has an accuracy of 90% at an 80:20 data split, while Random Forest has an accuracy of 85% at the same data split. However, at 70:30 data splitting, K-NN performance drops to 83%, comparable to Random Forest accuracy of 83% at the same data splitting. The study concluded that, in some cases, the K-NN algorithm showed slightly better accuracy, while Random Forest showed more consistent performance across multiple data splits.*

*Keyword: K-NN, Mental Disorders, Mental Health, Random Forest*

#### **Abstrak**

Pada era modern, gangguan mental menjadi masalah kesehatan global. Organisasi Kesehatan Dunia (WHO) memperkirakan bahwa satu dari empat orang di seluruh dunia akan mengalami gangguan mental atau neurologis. Gangguan sering terjadi pada pelajar yang salah satunya adalah mahasiswa. Kesehatan mental mahasiswa, yang akan menjadi generasi penerus bangsa, sangat penting untuk keberhasilan mereka di bidang akademis ataupun non akademis dan peran mereka di masyarakat di masa depan. Dengan menggunakan algoritma K-Nearest Neighbor (K-NN) dan Random Forest, penelitian ini bertujuan untuk menganalisis kesehatan mental untuk mencegah gangguan mental pada siswa. Dataset "Student mental health.csv" digunakan, yang diambil dari situs web Kaggle yang mencakup berbagai variabel terkait kesehatan siswa. Proses yang digunakan termasuk analisis data eksploratif, preprocessing data, modeling data menggunakan algoritma K-Nearest Neighbor (K-NN) dan Random Forest, dan akhirnya evaluasi. Hasil menunjukkan bahwa K-NN memiliki akurasi sebesar 90% pada splitting data 80:20, sedangkan Random Forest memiliki akurasi sebesar 85% pada splitting data yang sama. Namun, pada splitting data 70:30, kinerja K-NN turun menjadi 83%, sebanding dengan akurasi Random Forest 83% pada splitting data yang sama. Penelitian ini menyimpulkan bahwa, dalam beberapa kasus, algoritma K-NN menunjukkan akurasi yang sedikit lebih baik, sementara Random Forest menunjukkan kinerja yang lebih konsisten dalam berbagai pembagian data.

Kata Kunci: Gangguan Mental, Kesehatan Mental, K-NN, Random Forest

## 1. PENDAHULUAN

Era Modern saat ini sering kita jumpai permasalahan pada kesehatan mental. Masalah ini menjadi tantangan kesehatan global, bahkan *World Health Organization* (WHO) memprediksi 1 dari 4 orang di dunia akan terkena gangguan mental dan neurologis [1]. Mahasiswa merupakan seseorang yang sedang menempuh pendidikan tinggi. Sebagai generasi penerus bangsa mahasiswa harus memiliki kepercayaan diri. Kesehatan mental menentukan hasil akademik serta masa depan mahasiswa untuk meneruskan estafet bangsa demi mencapai generasi emas pada tahun 2045 [2]. Mengingat pentingnya kesehatan mental bagi mahasiswa, maka pencegahan gangguan mental diperlukan sedini mungkin.

Kesehatan mental merupakan aspek penting dalam perkembangan setiap orang, terutama bagi mahasiswa. Hasil dari zat-zat kimiawi yang tidak seimbang pada otak menyebabkan gangguan kesehatan mental [3]. Depresi, kecemasan, dan stress merupakan contoh gangguan mental yang sering dialami mahasiswa. Ini dapat menjadikan mahasiswa kehilangan minat untuk beraktivitas, kehilangan nafsu makan, jam tidur yang tidak beraturan, penurunan energi dan kognitif. Hal ini dapat berdampak buruk pada proses pendidikan mahasiswa, serta kualitas hidup mereka secara keseluruhan. Banyak sekali faktor yang mempengaruhi kesehatan mental mahasiswa [4]. Salah satunya lingkungan pendidikan dimana kesehatan mental memainkan peran yang sangat penting. Mahasiswa akan menghadapi tekanan dari berbagai sisi seperti akademik, sosial, dan personal yang dapat memicu gangguan mental [5][6].

Perkembangan teknologi memberikan alternatif baru untuk mencegah untuk mendeteksi dan mencegah gangguan mental. Beberapa algoritma yang sesuai dapat digunakan untuk menyelesaikan permasalahan tertentu dalam *machine learning*. Teknik *machine learning* yang paling umum digunakan dalam deteksi masalah kesehatan mental diantaranya K-Nearest Neighbor (K-NN) dan Random Forest [7]. Random forest (RF) adalah metode klasifikasi yang berasal dari metode pohon regresi dan menggunakan pemisahan biner rekursif untuk mencapai titik akhir dari struktur pohon yang didasarkan pada klasifikasi dan pohon regresi [8]. K-NN adalah algoritma pembelajaran mesin sederhana yang sering digunakan untuk menyelesaikan masalah regresi dan klasifikasi [9]. Kedua algoritma tersebut merupakan algoritma populer yang populer dalam *machine learning* [10].

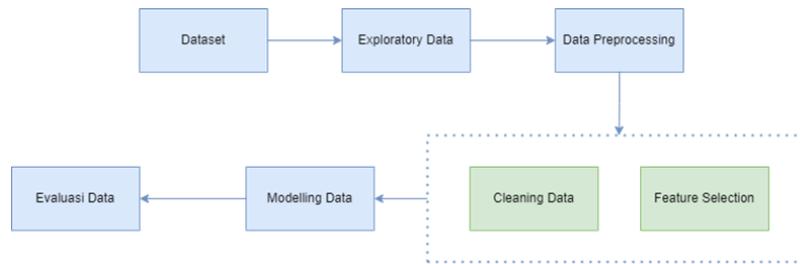
Penelitian sebelumnya yang dilakukan oleh Farhan Pratama, dkk. Yang membahas tentang perbandingan dua algoritma *machine learning*, yaitu K-NN dan Random Forest. Hasilnya menunjukkan akurasi dengan algoritma Random Forest sebesar 97% dengan pembagian data 70:30 dan 99% dengan pembagian data 80:20. Sementara dengan K-NN, menunjukkan akurasi 88% dengan pembagian data 70:20 dan 90% dengan pembagian data 80:20 [11]. Penelitian lain yang juga membahas penggunaan algoritma K-NN dan Random Forest oleh Kharisma Rahayu, dkk. Yang bertujuan untuk mengidentifikasi depresi dan kecemasan pada pengguna twitter dari klasifikasi teks. Hasil yang diperoleh dengan Random Forest menunjukkan akurasi 95,7% pada pembagian data 80:20. Kemudian hasil dengan K-NN menunjukkan akurasi sebesar 85,8% dengan pembagian yang sama yaitu 80:20 [8].

Penelitian yang telah dilakukan oleh Farhan Pratama, dkk. Serta Kharisma Rahayu, dkk. Membandingkan dua algoritma *machine learning*, yaitu, K-NN dan Random Forest, dalam mendeteksi masalah kesehatan mental. Penelitian tersebut hanya berfokus pada klasifikasi data dan mendapatkan akurasi yang cukup tinggi. Namun, pada penelitian ini memiliki perbedaan dalam hal pendekatan dan tujuan. Selain membandingkan hasil akurasi algoritma, juga berfokus pada analisis faktor-faktor risiko yang berpotensi menyebabkan gangguan mental pada mahasiswa, dengan menggunakan data yang lebih spesifik dan relevan dari dataset "Student Mental Health.csv". Analisis yang dilakukan secara mendalam ini bertujuan untuk memberikan wawasan yang lebih komprehensif mengenai kesehatan mental mahasiswa, dimana hal ini belum banyak dibahas dalam penelitian sebelumnya.

Dari permasalahan yang ada dan penelitian yang telah dipaparkan, Tujuan penelitian ini seiring dengan perkembangan teknologi *machine learning* yaitu algoritma K-NN dan Random Forest, Dimana hal ini menawarkan alternatif yang lebih efektif untuk mengklasifikasikan dan memprediksi gangguan mental berdasarkan pola data yang tidak selalu mudah diidentifikasi dengan metode konvensional [12][13]. Model-model ini ketika diterapkan dengan benar, dapat secara signifikan meningkatkan akurasi dan keandalan diagnosis kesehatan mental [14]. Ini juga mencakup adaptasi dan penerapan metode *machine learning* yang terus berkembang, memastikan bahwa model yang dihasilkan tetap relevan dan dapat diintegrasikan dengan teknologi terkini, seperti *cloud computing* dan *big data analytics*. Diharapkan penelitian ini berguna untuk memberikan kontribusi nyata dalam upaya pencegahan gangguan kesehatan mental dikalangan mahasiswa, yang merupakan salah satu kelompok paling rentan mengalami gangguan mental. Urgensi dari penelitian ini terletak pada kebutuhan mendesak untuk mengidentifikasi dan menangani faktor-faktor risiko gangguan mental mahasiswa. Hal ini sebagai bagian dari strategi jangka panjang untuk meningkatkan kesehatan mental serta kesejahteraan generasi penerus bangsa dalam rangka mencapai generasi emas tahun 2045.

## 2. METODE PENELITIAN

Proses penelitian dan metode diilustrasikan dengan dengan bagan sebagai berikut.



**Gambar 1.** Methodologi Penelitian

### 2.1. Dataset

Dataset adalah gabungan dari beberapa informasi atau data mentah yang terdiri dari penjelasan yang berhubungan dan dapat diolah, dikelola, dan dianalisis. Jenis data yang termasuk ke dalam kumpulan data dan format yang terstruktur, seperti tabel atau file, dan berisi informasi dari berbagai sumber seperti angka dan gambar.

### 2.2. *Exploratory Data Analysis*

Proses *Exploratory Data Analysis* (EDA) merupakan sebuah cara atau pendekatan untuk memahami pola, karakteristik dan hubungan yang ada pada data secara visualisasi dan deskriptif. Tujuan dari EDA adalah untuk membantu menganalisis data, memahami struktur data, menemukan pola tertentu, dan mengumpulkan informasi yang diperlukan untuk proses dalam mengambil keputusan [15]. Pada eksplorasi data ini, peneliti dapat membuat sebuah hipotesis yang dapat diuji di masa yang akan datang. Visualisasi data di dalam EDA dapat menggunakan grafik, scatter plot, histogram, dan statistik deskriptif dengan menggunakan perhitungan mean, median, dan standar deviasi sangat umum [16].

### 2.3. *Data Preprocessing*

Proses pemrosesan awal dimulai setelah data dikumpulkan. Tujuannya untuk memilih fitur-fitur yang diperlukan dan menghilangkan nilai yang kosong atau hilang di dalam data [17]. Untuk menghindari kesalahan yang terjadi, variabel dependen yang tidak digunakan dalam *training* dan pengujian dapat dipilih untuk dibuang atau hapus. Tahapan-tahapan di atas termasuk ke dalam data *preprocessing*.

#### 2.3.1. *Cleaning Data*

Setelah *Data Preprocessing* yang akan dilakukan dalam penelitian ini adalah mengumpulkan data; penulisan ulang nilai yang kosong atau tidak memiliki nilai. Data ini perlu diteliti untuk menemukan informasi yang tidak akurat. Hal ini dilakukan karena data yang hilang atau kosong dapat memengaruhi akurasi. Akurasi dapat berkurang karena terdapat nilai yang hilang atau kosong di dalam data tersebut [18]. Demi memastikan bahwa dataset tidak terdapat nilai yang hilang atau kosong, setiap baris dataset pada setiap variabel mempunyai nilai panjang.

#### 2.3.2. *Feature Selection*

Feature Selection merujuk kepada fitur pemilihan setiap variabel yang tersedia dari dataset untuk digunakan dalam membuat sebuah model. Tujuan memilih fitur yang sesuai adalah meningkatkan produktivitas model dengan mengubah variabel yang tidak relevan. Data akan menjadi lebih mudah untuk diproses jika data yang dimiliki relevan [19]. Dalam penelitian kali ini, pemilihan fitur menggunakan tiga puluh variabel.

### 2.4. *Modelling Data*

Metode yang digunakan adalah algoritma K-Nearest Neighbor dan Random Forest, metode tersebut dapat memberikan prediksi berdasarkan data individu atau kelompok yang sama dengan persoalan yang diamati. Oleh karena itu, metode ini dapat digunakan untuk menghasilkan tingkat personalisasi pada model, yang penting untuk dilakukan dalam memprediksi tingkat kesehatan mental individu, tiap-tiap individu memiliki hasil yang berbeda dengan lainnya.

Random Forest sering digunakan dalam menyelesaikan sebuah persoalan. Dalam kasus kali ini, Algoritma Random Forest (RF) adalah sebuah metode pembelajaran mesin yang berisi kumpulan pengklasifikasi yang menggabungkan berbagai jenis pohon keputusan untuk mencapai hasil yang diinginkan [20]. Metode ini membantu mengasosiasikan hasil dan memahami faktor-faktor yang paling memengaruhi kesehatan mental.

Konsep penelitian dengan algoritma K-Nearest Neighbor telah banyak dilakukan oleh peneliti-peneliti sebelumnya diantaranya yaitu, algoritma K-Nearest Neighbor digunakan untuk pengenalan pola, pengenalan teks, pengolahan objek dan lain-lain. Algoritma K-Nearest Neighbor dianggap mempunyai kesederhanaan dalam pengolahan data training dan data testing dalam jumlah yang sangat besar.

## 2.5. Evaluasi Model

Salah satu langkah penting dalam menilai kinerja model adalah melakukan evaluasi model klasifikasi menggunakan *Matrix*. Tujuan dari proses ini adalah untuk mengetahui seberapa baik model mampu melakukan klasifikasi dengan benar. *Accuracy*, *Precision*, *Recall*, *F1 Score*, dukungan, dan konsistensi adalah lima komponen utama evaluasi matriks klasifikasi. Untuk mengevaluasi kinerja algoritma pembelajaran yang diawasi, tabel *Confusion Matrix*, juga dikenal sebagai *Matrix* kebingungan, digunakan. Tabel ini memiliki dua baris yang menampilkan contoh kelas yang sebenarnya, dan satu kolom menampilkan *instance* kelas yang diprediksi oleh model. Pengukuran dapat di hitung menggunakan rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP + TN}{TP + FN} \quad (3)$$

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

*True Positives* (TP) merupakan Jumlah prediksi benar untuk kelas tertentu, *True Negatives* (TN) merupakan Jumlah prediksi benar bukan untuk kelas tertentu (tidak termasuk TP kelas itu), *False Positives* (FP) merupakan Jumlah prediksi salah yang menunjukkan anggota kelas tertentu (tidak termasuk TP kelas itu), dan *False Negatives* (FN) merupakan Jumlah prediksi salah yang menunjukkan bukan anggota kelas tertentu (tidak termasuk baris dan kolom dari kelas tersebut).

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Data

Pada fase ini, dataset diperoleh dari situs web Kaggle, yang merupakan situs web yang berfokus pada ilmu data [3]. Yang berisi data mahasiswa, dengan berjumlah asli 101 data, yang terdiri dari 10 variabel dalam satu table.

**Tabel 1.** Dataset Student Mental Health

No	Date	Gender	Age	Major	Year of Study	CGPA	Depression	Anxiety	PA	Treatment
1	8/7/2020 12:02	Female	18	Engineering	1	3.00-3.49	No	Yes	No	Yes
2	8/7/2020 12:04	Male	21	Islamic education	2	3.00-3.49	No	No	Yes	No
3	8/7/2020 12:05	Male	19	BIT	1	3.00-3.49	No	Yes	Yes	Yes
4	8/7/2020 12:06	Female	22	Laws	3	3.00-3.49	Yes	Yes	No	No
5	8/7/2020 12:13	Male	23	Mathemathics	4	3.00-3.49	No	No	No	No
...	...	...	...	...	...	...	...	...	...	...
96	13/07/2020 19:56	Female	21	BCS	1	3.50-4.00	No	No	Yes	No
98	13/07/2020 21:21	Male	18	Engineering	2	3.00-3.49	No	Yes	Yes	No
99	13/07/2020 21:22	Female	19	Nursing	3	3.50-4.00	Yes	Yes	No	Yes
100	13/07/2020 21:23	Female	23	Islamic Education	4	3.50-4.00	No	No	No	No
101	18/07/2020 20:16	Male	20	Biomedical science	2	3.00-3.49	No	No	No	No

### 3.2. Preprocessing Data

Preprocessing data adalah tahap penting dalam analisis data menggunakan algoritma Random Forest atau algoritma machine learning lainnya. Tujuannya adalah untuk mempersiapkan data mentah yang akan diubah menjadi data yang relevan agar sesuai dengan format yang dibutuhkan oleh model pembelajaran mesin.

#### 3.2.1. Cleaning Data

Pada dataset Student Mental Health dilakukan manajemen data dari nama kolom, data kosong atau null, penghapusan variabel yang tidak relevan pada tahap ini. Terdapat satu nilai null pada dataset Student Mental Health dan fitur time stamp yang tidak dibutuhkan dalam analisis. Kemudian data kosong dan fitur yang tidak dibutuhkan dihapus agar memudahkan dalam menganalisis data. Setelah itu, nama kolom diubah agar dapat dipahami pada proses selanjutnya.

#### 3.2.2. Feature Selection

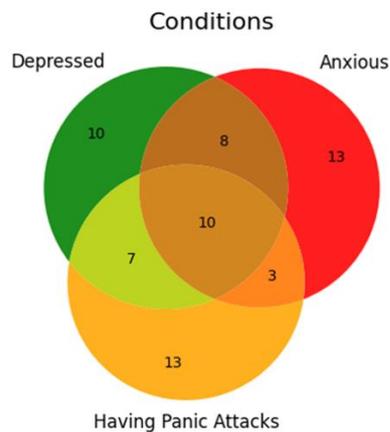
Pada tahapan ini kita akan melakukan seleksi variable data yang relevan dalam penelitian analisis data yang akan digunakan, setelah dilakukan analisis seleksi data pada setiap variable hasil menunjukkan ditemukan data yang relevan yaitu:

1. Gender = Jenis Kelamin mahasiswa
2. Age = Usia mahasiswa
3. Major = Jurusan mahasiswa
4. Year Study = Masa kuliah mahasiswa
5. CGPA = IPK para mahasiswa
6. Material Status = Apakah mahasiswa memiliki hubungan atau tidak
7. Depression = Apakah mahasiswa mengalami depresi atau tidak
8. Panic Attack = Apakah mahasiswa mengalami Serangan Panik atau tidak
9. Treatment = apakah mahasiswa pernah ke psikolog atau tidak

Dari hasil analisis seleksi variable data yang kami seleksi, ditemukan hampir semua variable data kami akan gunakan karena relevan dengan penelitian analisis data yang kami lakukan.

### 3.3. Exploratory Data

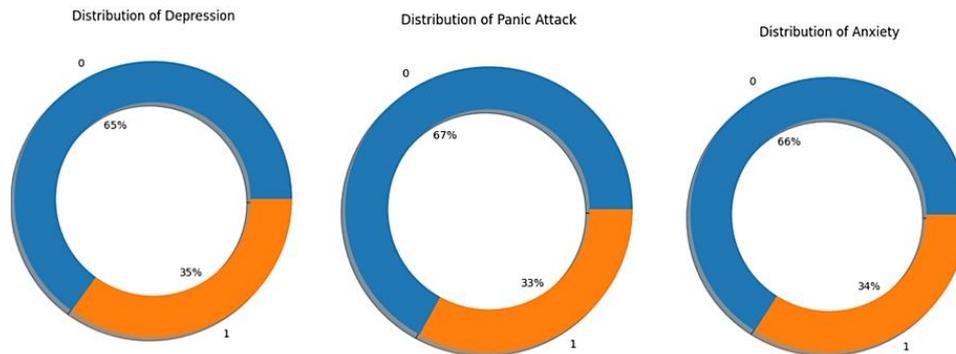
Gambar dibawah merupakan visualisasi data dari hasil Eksplorasi data yang telah dilakukan.



**Gambar 2.** Visualisasi hasil ekplorasi data

Gambar 2 setelah dianalisis terdapat lebih banyak mahasiswa yang mengalami gangguan mental health karena dari banyaknya sample variable data seperti depresi 10 data, *Anxious* 13 data, dan *Panic Attack* 13 data, mahasiswa yang mengalami tiga gangguan tersebut yang kita dapat simpulkan bahwa mahasiswa tersebut memiliki gangguan Kesehatan mental ternyata mendapatkan 10 dari banyaknya sample data yang kami lakukan kombinasi.

Grafik diagram lingkaran pada gambar 3 menunjukkan 3 jenis kondisi kesehatan mental para mahasiswa. Dari grafik diatas menunjukkan bahwa dominan banyak mahasiswa yang mengalami banyak kondisi kesehatan depresi, kecemasan, serangat panik dibandingkan yang tidak. Hampir sekitar lebih 50% mahasiswa mengalami gangguan kondisi kesehatan mental sekitar 65% mahasiswa mengalami depresi, 66% mahasiswa mengalami kecemasan. Dan 67% mahasiswa mengalami serangan kepanikan. Dari tiga kondisi tersebut kita dapat melihat bahwa banyak mahasiswa yang mengalami kondisi gangguan mental dilihat dari tingkat persen setiap kondisi tanpa harus melakukan analisis penggabungan antara 3 kondisi tersebut.



**Gambar 3.** Diagram distribusi kesehatan mental mahasiswa

### 3.4. *Splitting Data*

Dataset dibagi menjadi data *training* untuk pelatihan model dan data *test* untuk pengujian serta evaluasi [21]. *Splitting data* dilakukan 2 kali dalam penelitian ini yakni 70:30 serta 80:20. Berarti pada *splitting data* 70:30 artinya 70% digunakan untuk data latih dan 30% untuk data uji. Sementara pada *splitting data* 80:20, 80% digunakan untuk data latih, kemudian 20% untuk data uji. Dilakukannya pembagian data ini bertujuan untuk mengevaluasi performa dari model dengan dua skenario berbeda.

### 3.5. K-Nearest Neighbor

**Table 2.** Hasil *splitting data* 80:20 K-NN

	Precision	Recall	F1-score	Support
0	0,88	1,00	0,94	15
1	1,00	0,60	0,75	5
Accuracy	-	-	0,90	20
Macro avg	0,94	0,80	0,84	20
Weighted avg	0,91	0,90	0,89	20

Dari tabel 2 menunjukkan bahwa akurasi yang didapat yaitu sebesar 90% menggunakan metode *Confusion Matrix* pada *splitting data* 80:20 menggunakan algoritma K-NN. Tabel ini menampilkan metrik evaluasi seperti akurasi, *recall*, *F1-score*, dan *support* untuk dua kelas, 0 dan 1. Untuk kelas 0, akurasi model adalah 0,88, *recall* 1,00, dan *F1-score* 0,94 dengan dukungan 15 data; untuk kelas 1, akurasi model adalah 1,00, *recall* 0,60, dan *F1-score* 0,75 dengan dukungan 5 data. Secara keseluruhan, akurasi model adalah 0,90, dengan nilai rata-rata akurasi 0,94, *recall* 0,80, dan *F1-score* 0,84.

**Table 3.** Hasil *splitting data* 70:30 K-NN

	Precision	Recall	F1-score	Support
0	0,85	0,96	0,90	23
1	1,00	0,43	0,55	7
Accuracy	-	-	0,83	30
Macro avg	0,80	0,69	0,72	30
Weighted avg	0,82	0,83	0,82	30

Tabel 3 menunjukkan hasil evaluasi model K-Nearest Neighbors (K-NN) yang dilakukan dengan metode *splitting data* 70:30. Tabel ini menunjukkan metrik evaluasi untuk kelas 0 dan 1, termasuk *precision*, *recall*, *F1-score*, dan *support*. Untuk kelas 0, *precision*-nya adalah 0,85, *recall* 0,96, dan *F1-score* 0,90, dengan 23 data didukung. Untuk kelas 1, *precision*-nya adalah 1,00, *recall* 0,43, dan *F1-score* 0,55, dengan 7 data didukung. Secara keseluruhan, model memiliki akurasi 0,83. Nilai rata-rata makro (makro) untuk ketepatan, pengenalan, dan nilai F1 masing-masing adalah 0,80, 0,69, dan 0,72. Nilai rata-rata berbobot (berat rata-rata) masing-masing adalah 0,82, 0,83, dan 0,82.

### 3.6. Random Forest

**Table 4.** Hasil *splitting data* 80:20 Random Forest

	Precision	Recall	F1-score	Support
0	0,83	1,00	0,91	15
1	1,00	0,40	0,57	5
Accuracy	-	-	0,85	20

	Precision	Recall	F1-score	Support
Macro avg	0,92	0,70	0,74	20
Weighted avg	0,88	0,85	0,82	20

Dari tabel 4 menunjukkan bahwa akurasi yang didapat yaitu sebesar 85% menggunakan metode *confusion matrix* pada *splitting data* 80:20 menggunakan algoritma Random Forest. Akurasi model adalah 0,85 secara keseluruhan, dengan nilai rata-rata *precision* 0,92, *recall* 0,70, dan *F1-score* 0,74. Untuk kelas 0, akurasi adalah 0,83, *recall* 1,00, dan *F1-score* 0,91 dengan dukungan 15 data.

**Table 5.** Hasil *splitting data* 70:30 Random Forest

	Precision	Recall	F1-score	Support
0	0,82	1,00	0,90	23
1	1,00	0,29	0,44	7
Accuracy	-	-	0,83	30
Macro avg	0,91	0,70	0,67	30
Weighted avg	0,86	0,83	0,80	30

Dari tabel diatas menunjukkan bahwa akurasi yang didapat yaitu sebesar 90% menggunakan metode *Confusion Matrix* pada *splitting data* 70:30 menggunakan algoritma Random Forest. Hasil evaluasi model Random Forest menggunakan metode *splitting data* 70:30 ditunjukkan dalam tabel ini. Akurasi model adalah 0,83 secara keseluruhan, dengan nilai *precision* rata-rata 0,91, *recall* 0,70, dan *F1-score* 0,67. Untuk kelas 0, *precision*-nya adalah 0,82, *recall* 1,00, dan *F1-score* 0,90, dengan 23 data yang didukung.

Tabel-tabel di atas menunjukkan *Matrix* evaluasi untuk model klasifikasi biner dengan dua kelas, yaitu 0 dan 1. Untuk kelas 0, model memiliki *Precision* 0,83, *Recall* 1,00, dan *F1-score* 0,91 dengan *support* sebanyak 15 contoh. Sementara itu, untuk kelas 1, model memiliki *precision* 1,00, *recall* 0,40, dan *F1-score* 0,57 dengan *support* sebanyak 5 contoh. Secara keseluruhan, akurasi model adalah 0,85, yang berarti model berhasil mengklasifikasikan 85% dari semua contoh dengan benar. Nilai rata-rata makro (*macro average*) untuk *Precision*, *Recall*, dan *F1-score* masing-masing adalah 0,92, 0,70, dan 0,74, yang memperlakukan setiap kelas dengan bobot yang sama. Sedangkan nilai rata-rata tertimbang (*weighted average*) untuk *Precision*, *Recall*, dan *F1-score* masing-masing adalah 0,88, 0,85, dan 0,82, yang mempertimbangkan proporsi tiap kelas dalam dataset. Hasil ini menunjukkan bahwa model memiliki performa yang lebih baik dalam mengklasifikasikan kelas yang lebih dominan (kelas 0) dibandingkan kelas yang lebih sedikit (kelas 1), terlihat dari perbedaan signifikan antara *Recall* dan *F1-score* kedua kelas.

### 3.7. Analisis dan Evaluasi Model

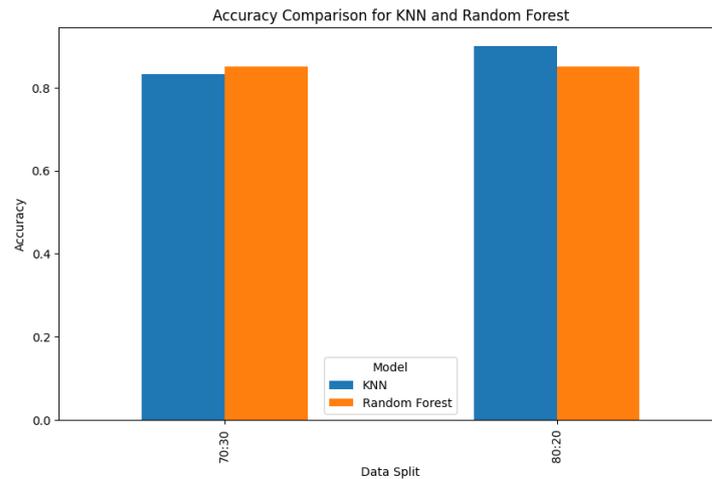
Evaluasi performa model menggunakan *Matrix* evaluasi dari data pengujian [22]. Berdasarkan pada hasil yang didapat dari menggunakan algoritma K-Nearest Neighbor dan Random Forest dengan *splitting data* 80:20 dan 70:30, bisa dilihat bahwa algoritma K-NN lebih unggul pada tingkat akurasi dalam penelitian ini. Seperti pada tabel 6.

**Table 6.** Perbandingan akurasi algoritma K-NN dan Random Forest

	K-NN				Random Forest					
	Precision	Recall	F1 Score	Support	Precision	Recall	F1 Score	Support		
80:20	0	0,88	1,00	0,94	15	0	0,83	1,00	0,91	15
	1	1,00	0,6	0,75	5	1	1,00	0,40	0,57	5
	Accuracy	-	-	0,90	20	Accuracy	-	-	0,85	20
	Macro avg	0,94	0,80	0,84	20	Macro avg	0,92	0,70	0,74	20
	Weighted avg	0,91	0,90	0,89	20	Weighted avg	0,88	0,85	0,82	20
		Precision	Recall	F1 Score	Support	Precision	Recall	F1 Score	Support	
70:30	0	0,85	0,96	0,90	23	0	0,82	1,00	0,90	23
	1	1,00	0,43	0,55	7	1	1,00	0,29	0,44	7
	Accuracy	-	-	0,83	30	Accuracy	-	-	0,83	30
	Macro avg	0,80	0,69	0,72	30	Macro avg	0,91	0,70	0,67	30
	Weighted avg	0,82	0,83	0,82	30	Weighted avg	0,86	0,83	0,80	30

Pada kedua *split data* (80:20 dan 70:30), K-NN menunjukkan performa yang lebih baik dalam hal *Recall* dan *F1 Score* untuk kelas '1', meskipun akurasi keseluruhan untuk kedua model relatif sama pada *split data*

70:30. Random Forest cenderung memiliki *Recall* yang lebih rendah untuk kelas '1', yang berarti model ini kurang baik dalam mendeteksi kasus '1' dibandingkan K-NN.



**Gambar 4.** Diagram batang perbandingan akurasi antara K-Nearest Neighbors (K-NN) dan Random Forest

Dengan menggunakan diagram batang, kami dapat melihat perbandingan akurasi antara K-Nearest Neighbors (K-NN) dan Random Forest pada dua pembagian data, yaitu 70:30 dan 80:20. Hasil menunjukkan bahwa pada pembagian data 70:30, kedua model memiliki akurasi yang hampir sama, dengan Random Forest sedikit lebih baik. Namun, pada pembagian data 80:20, K-NN menunjukkan akurasi yang lebih tinggi. Meskipun kedua model memiliki akurasi keseluruhan yang sebanding pada pembagian data 70:30, K-NN tampaknya memiliki skor Recall dan F1 yang lebih baik untuk kelas '1', sementara Random Forest memiliki skor Recall yang lebih rendah untuk kelas tersebut, menunjukkan bahwa K-NN lebih baik dalam mendeteksi kasus '1'.

#### 4. KESIMPULAN

Penelitian ini menyimpulkan bahwa penggunaan algoritma K-Nearest Neighbor dan Random Forest dapat memberikan wawasan yang berharga untuk menganalisis kesehatan mental mahasiswa dan mengidentifikasi faktor-faktor risiko utama yang dapat menyebabkan gangguan mental. Hasil dari penelitian ini dapat digunakan sebagai dasar untuk mengembangkan strategi pencegahan yang lebih efektif dan personal dalam menangani masalah kesehatan mental mahasiswa. Algoritma Random Forest dan K-Nearest Neighbor (K-NN) menunjukkan performa yang bervariasi dalam klasifikasi data kesehatan mental mahasiswa. Random Forest menunjukkan akurasi yang lebih tinggi dibandingkan K-NN dalam beberapa kasus, dengan akurasi mencapai 85% pada pembagian data 80:20 dan 83% pada pembagian data 70:30. K-NN menunjukkan akurasi yang sedikit lebih tinggi pada pembagian data 80:20 dengan akurasi 90%, namun performa menurun pada pembagian data 70:30 dengan akurasi 83%. Perbedaan performa ini menunjukkan bahwa Random Forest lebih konsisten dalam klasifikasi data dibandingkan K-NN. Penggunaan teknologi dan pembelajaran mesin dalam analisis kesehatan mental dapat memberikan kontribusi signifikan dalam menciptakan lingkungan pendidikan yang lebih sehat dan mendukung bagi mahasiswa.

#### UCAPAN TERIMAKASIH

Kami mengucapkan terima kasih kepada Universitas Amikom Purwokerto yang telah memberikan fasilitas dan sarana yang diperlukan untuk penelitian ini. Dukungan dan bimbingan dari Fakultas Ilmu Komputer sangat berharga dalam proses penelitian ini. Kami juga menyampaikan terima kasih kepada Pusat Studi Sistem Cerdas dan Computer Vision yang telah memberikan bimbingan serta masukan yang berharga selama proses penelitian. Tanpa dukungan dari lembaga dan institusi ini, penelitian ini tidak akan dapat terwujud.

#### REFERENSI

- [1] M. Sumathi and D. B., "Prediction of Mental Health Problems Among Children Using Machine Learning Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 552–557, 2016, doi: 10.14569/ijacsa.2016.070176.
- [2] Z. Tanjung and S. Amelia, "Menumbuhkan Kepercayaan Diri Siswa," *JRTI (Jurnal Ris. Tindakan Indones.*, vol. 2, no. 2, pp. 2–6, 2017, doi: 10.29210/3003205000.
- [3] R. Maringga and K. Kusnawi, "Exploratory Data Analysis Faktor Pengaruh Kesehatan Mental di Tempat Kerja," *CogITo Smart J.*, vol. 7, no. 2, pp. 215–226, 2021, doi: 10.31154/cogito.v7i2.312.215-

- 226.
- [4] N. S. Mohd Shafiee and S. Mutalib, "Prediction of Mental Health Problems among Higher Education Student Using Machine Learning," *Int. J. Educ. Manag. Eng.*, vol. 10, no. 6, pp. 1–9, 2020, doi: 10.5815/ijeme.2020.06.01.
- [5] D. O. Aluh, A. Abba, and A. B. Afosi, "Prevalence and correlates of depression, anxiety and stress among undergraduate pharmacy students in Nigeria," *Pharm. Educ.*, vol. 20, pp. 236–248, 2020, doi: 10.46542/pe.2020.201.236248.
- [6] M. Atas, W. Urban, D. A. N. Rural, and K. Jember, "Kesehatan mental pada remaja di lingkungan sekolah menengah atas wilayah urban dan rural kabupaten jember," vol. 11, no. 3, pp. 537–544, 2023.
- [7] R. A. Rahman, K. Omar, S. A. M. Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, "Application of machine learning methods in mental health detection: A systematic review," *IEEE Access*, vol. 8, pp. 183952–183964, 2020, doi: 10.1109/ACCESS.2020.3029154.
- [8] K. Rahayu, V. Fitriya, D. Septhya, R. Rahmadden, and L. Efrizoni, "Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 108–114, 2023, doi: 10.57152/malcom.v3i2.780.
- [9] A. Oktaviana, D. P. Wijaya, A. Pramuntadi, and D. Heksaputra, "Prediction of Type 2 Diabetes Mellitus Using The K-Nearest Neighbor ( K-NN ) Algorithm Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor ( K-NN )," vol. 4, no. July, pp. 812–818, 2024.
- [10] P. Josso, A. Hall, C. Williams, T. Le Bas, P. Lusty, and B. Murton, "Application of random-forest machine learning algorithm for mineral predictive mapping of Fe-Mn crusts in the World Ocean," *Ore Geol. Rev.*, vol. 162, no. June, p. 105671, 2023, doi: 10.1016/j.oregeorev.2023.105671.
- [11] F. Pratama, Z. Hadryan Nst, Z. Khairi, and L. Efrizoni, "Perbandingan Algoritma Random Forest Dan K-Nearest Neighbor Dalam Klasifikasi Kesehatan Mental Mahasiswa," *J. Ilm. Betrik*, vol. 15, no. 1, pp. 31–37, 2024.
- [12] A. Samad, S. TAZE, and M. Kürsad UÇAR, "Enhancing Milk Quality Detection with Machine Learning: A Comparative Analysis of K-NN and Distance-Weighted K-NN Algorithms," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 3, pp. 2021–2029, 2024, doi: 10.38124/ijisrt/ijisrt24mar2123.
- [13] J. Fisher, S. Allen, G. Yetman, and L. Pistolesi, "Assessing the influence of landscape conservation and protected areas on social wellbeing using random forest machine learning," *Sci. Rep.*, vol. 14, no. 1, pp. 1–16, 2024, doi: 10.1038/s41598-024-61924-4.
- [14] J. Ehiabhi and H. Wang, "A Systematic Review of Machine Learning Models in Mental Health Analysis Based on Multi-Channel Multi-Modal Biometric Signals," *BioMedInformatics*, vol. 3, no. 1, pp. 193–219, 2023, doi: 10.3390/biomedinformatics3010014.
- [15] N. Fathirachman Mahing, A. Lazuardi Gunawan, A. Foresta Azhar Zen, F. Abdurrachman Bachtiar, and S. Agung Wicaksono, "Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 7, pp. 1527–1536, 2023, doi: 10.25126/jtiik.1078010.
- [16] M. Dwidiyanti, D. R. Sawitri, B. Munif, A. Fitri, D. Yuli, and A. Information, "ORIGINAL ARTICLE THE HAPPY SPIRITUAL INTERVENTION FOR OVERCOMING INTOLERANCE OF UNCERTAINTY AMONG FAMILIES OF PATIENTS WITH MENTAL DISORDERS," vol. 19, no. 2, pp. 72–77, 2024, doi: 10.20884/1.jks.2024.19.2.11032.
- [17] T. Solang and A. Nugroho, "Analisis Kesehatan Mental Mahasiswa Universitas Kristen Satya Wacana Menggunakan Metode Clustering Algoritma K-Means," *J. TEKINKOM*, vol. 6, no. 1, pp. 8–15, 2023, doi: 10.37600/tekinkom.v6i1.641.
- [18] Sopiatalul Ulum, R. F. Alifa, P. Rizkika, and C. Rozikin, "Perbandingan Performa Algoritma K-NN dan SVM dalam Klasifikasi Kelayakan Air Minum," *Gener. J.*, vol. 7, no. 2, pp. 141–146, 2023, doi: 10.29407/gj.v7i2.20270.
- [19] K. B. Simarmata, K. D. Hartomo, and K. D. Hartomo, "Analisa Rekomendasi Fitur Persetujuan Pinjaman Perusahaan Financial Technology Menggunakan Metode Random Forest," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2055–2070, 2022, doi: 10.35957/jatisi.v9i3.2258.
- [20] L. Savitri and R. Nursalim, "Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning," *Diophantine J. Math. Its Appl.*, vol. 2, no. 01, pp. 30–36, 2023, doi: 10.33369/diophantine.v2i01.28260.
- [21] M. Fadli and R. A. Saputra, "Klasifikasi Dan Evaluasi Performa Model Random Forest Untuk Prediksi Stroke," *JT J. Tek.*, vol. 12, no. 2, pp. 72–80, 2023, [Online]. Available: <http://jurnal.umt.ac.id/index.php/jt/index>
- [22] M. Rijal, F. Aziz, and S. Abasa, "Prediksi Depresi: Inovasi Terkini Dalam Kesehatan Mental Melalui Metode Machine Learning," *J. Pharm. Appl. Comput. Sci.*, vol. 2, no. 1, pp. 9–14, 2024, doi: 10.59823/jopacs.v2i1.47.