



## Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Prediksi Kualitas Tidur pada Kesehatan

Fakhrudin<sup>1\*</sup>, Sefrika Entas<sup>2</sup>

Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

Email: [fakhrudin4444@gmail.com](mailto:fakhrudin4444@gmail.com)<sup>1\*</sup>

Alamat: Jl. Gatot Subroto No.8, Cimone, Kec. Karawaci, Kota Tangerang, Banten. Kode Pos 15114

\*Penulis Korespondensi

**Abstract.** Sleep is a fundamental human need that plays a crucial role in maintaining both physical and mental health. Poor sleep quality can trigger a variety of health problems, ranging from decreased concentration to an increased risk of chronic diseases. The complexity of factors influencing sleep quality—such as stress levels, heart rate, blood pressure, physical activity, and lifestyle—makes its assessment difficult through direct observation alone. Therefore, data mining approaches are increasingly utilized to identify relevant patterns in sleep-related data. This study aims to compare the performance of the C4.5 (Decision Tree) algorithm and the Naïve Bayes algorithm in predicting sleep quality using the Sleep Health and Lifestyle dataset, which contains information from 374 respondents. The research method applied is a quantitative comparative approach employing classification techniques with 10-fold cross-validation to ensure robust evaluation. Model performance is assessed using accuracy, precision, and recall metrics to provide a comprehensive understanding of the effectiveness of each algorithm. The findings indicate that the C4.5 algorithm achieves an accuracy of 96.26% and offers advantages in terms of interpretability through its decision tree visualization, enabling easier understanding of variable relationships. In contrast, the Naïve Bayes algorithm demonstrates superior predictive performance, achieving an accuracy of 98.66% along with consistently high precision and recall across nearly all classes. These results suggest that Naïve Bayes is more effective for predictive tasks involving sleep quality, while C4.5 remains highly valuable when the goal is to interpret variable interactions and decision rules. Overall, this research highlights the potential of data mining techniques in health informatics, particularly in improving the understanding and prediction of sleep quality, which in turn can contribute to better prevention and management of sleep-related health issues.

**Keywords:** Algorithm; C4.5; Comparison; Naïve Bayes; Sleep Quality

**Abstrak.** Tidur merupakan kebutuhan mendasar manusia yang memiliki peran penting dalam menjaga kesehatan fisik maupun mental. Kualitas tidur yang buruk dapat memicu berbagai masalah kesehatan, mulai dari penurunan konsentrasi hingga peningkatan risiko penyakit kronis. Kompleksitas faktor yang memengaruhi kualitas tidur—seperti tingkat stres, detak jantung, tekanan darah, aktivitas fisik, serta gaya hidup—menjadikan proses penilaian tidak mudah bila hanya mengandalkan observasi langsung. Oleh karena itu, pendekatan berbasis data mining semakin banyak digunakan untuk mengidentifikasi pola-pola relevan dalam data terkait tidur. Penelitian ini bertujuan untuk membandingkan performa algoritma C4.5 (Decision Tree) dan Naïve Bayes dalam memprediksi kualitas tidur dengan memanfaatkan dataset Sleep Health and Lifestyle yang mencakup 374 responden. Metode penelitian yang diterapkan adalah kuantitatif komparatif dengan teknik klasifikasi serta validasi silang 10-fold untuk memastikan hasil evaluasi yang lebih andal. Kinerja model diukur menggunakan metrik akurasi, presisi, dan recall guna memberikan gambaran menyeluruh mengenai efektivitas masing-masing algoritma. Hasil penelitian menunjukkan bahwa algoritma C4.5 memperoleh akurasi sebesar 96,26% dengan keunggulan pada interpretabilitas melalui visualisasi pohon keputusan, sehingga memudahkan pemahaman hubungan antarvariabel. Sebaliknya, algoritma Naïve Bayes menunjukkan performa prediksi yang lebih unggul dengan akurasi 98,66% serta presisi dan recall tinggi pada hampir semua kelas. Temuan ini mengindikasikan bahwa Naïve Bayes lebih efektif digunakan untuk prediksi kualitas tidur, sementara C4.5 tetap bermanfaat ketika tujuan utama adalah menginterpretasikan interaksi antarvariabel dan aturan keputusan. Secara keseluruhan, penelitian ini menegaskan potensi teknik data mining dalam bidang informatika kesehatan, khususnya dalam meningkatkan pemahaman serta prediksi kualitas tidur yang pada akhirnya dapat mendukung upaya pencegahan dan penanganan masalah kesehatan terkait tidur.

**Kata kunci:** Algoritma; C4.5; Kualitas tidur; Naïve Bayes; Perbandingan

## **1. PENDAHULUAN**

Tidur merupakan kebutuhan mendasar bagi manusia yang berperan penting dalam menjaga kesehatan fisik maupun mental, karena kualitas tidur yang baik mampu meningkatkan sistem kekebalan tubuh, produktivitas, serta kestabilan emosi, sedangkan gangguan tidur dapat memicu berbagai masalah kesehatan (Sastra & Sabri, 2025). Seiring dengan berkembangnya teknologi informasi, analisis kualitas tidur kini dapat didukung oleh pendekatan data science, yakni bidang multidisiplin yang menggabungkan metode statistik, algoritma machine learning, serta manajemen data untuk menemukan pola dalam data skala besar (Anggi Trifani et al., 2022). Pentingnya penelitian ini terletak pada kontribusinya dalam memperkaya kajian data mining, khususnya melalui pemanfaatan algoritma klasifikasi seperti C4.5 dan Naïve Bayes, yang mampu membantu memahami faktor-faktor penentu kualitas tidur secara lebih objektif dan terukur. Dengan menggunakan dataset Sleep Health and Lifestyle yang memuat informasi terkait durasi tidur, stres, detak jantung, aktivitas harian, hingga gangguan tidur, penelitian ini bertujuan untuk mengetahui metode klasifikasi yang paling akurat dan efektif dalam memprediksi kualitas tidur, sehingga dapat memberikan wawasan yang bermanfaat baik secara akademis maupun praktis dalam bidang kesehatan preventif.

Penelitian terdahulu telah menunjukkan hasil yang signifikan pada penerapan kedua algoritma. Misalnya, (Imam Nawawi & Zaehol Fatah, 2024) menemukan bahwa C4.5 mampu menghasilkan akurasi tinggi pada dataset kesehatan dengan visualisasi yang mudah dipahami, meski belum dibandingkan dengan algoritma lain. Sementara itu, (Vidi et al., 2025) membuktikan bahwa Naïve Bayes efisien dalam menangani data besar dan kompleks, dengan performa stabil pada banyak fitur, namun aspek interpretabilitas model tidak menjadi fokus. Berdasarkan penelitian sebelumnya, masih terdapat celah penelitian yaitu belum adanya perbandingan langsung antara algoritma C4.5 dan Naïve Bayes pada konteks prediksi kualitas tidur. Selain itu, sebagian besar studi terdahulu hanya menilai akurasi tanpa memperhatikan metrik lain seperti presisi dan recall, padahal indikator tersebut penting untuk mengukur performa model secara menyeluruh. Oleh karena itu, penelitian ini hadir untuk mengisi gap tersebut dengan melakukan evaluasi komparatif yang lebih komprehensif, mencakup akurasi, presisi, dan recall pada kedua algoritma.

## 2. LANDASAN TEORI

### Kualitas Tidur pada Kesehatan

Kualitas tidur yang baik merupakan elemen penting dalam menjaga kesehatan fisik, mental, maupun emosional. Tidur tidak hanya berfungsi sebagai waktu istirahat, tetapi juga menjadi fase vital bagi tubuh untuk melakukan pemulihan, menjaga keseimbangan hormon, serta memperkuat sistem kekebalan. Gangguan pada kualitas tidur dapat menimbulkan berbagai masalah, seperti rasa lelah berlebihan, menurunnya konsentrasi, hingga meningkatnya risiko penyakit jantung serta gangguan psikologis berupa stres dan depresi (Cahya et al., 2025). Penelitian yang dilakukan oleh (Hardian et al., 2025) menegaskan bahwa kurang tidur berdampak negatif terhadap kinerja individu, terutama pada pekerjaan yang membutuhkan tingkat konsentrasi tinggi, misalnya di sektor kesehatan. Selain itu, lansia yang mengalami gangguan tidur cenderung menunjukkan gejala depresi lebih tinggi dibandingkan dengan mereka yang memiliki pola tidur normal. Temuan ini mempertegas bahwa kualitas tidur merupakan aspek krusial dalam menjaga kualitas hidup secara menyeluruh.

### Data Mining

Data mining merupakan proses untuk menemukan pola serta informasi penting dari data yang dipilih dengan memanfaatkan berbagai teknik, metode, maupun algoritma. Ragam metode yang digunakan dalam data mining sangat bervariasi, dan pemilihannya harus disesuaikan dengan tujuan penelitian serta keseluruhan proses *Knowledge Discovery in Databases* (Putra & Program, 2024). Dalam klasifikasi data mining yang berfungsi membangun model keputusan, biasanya didukung oleh teknik diskritisasi fitur, *resampling* data, dan seleksi fitur. Dengan memilih subset fitur yang relevan, beban komputasi dapat dikurangi, sehingga efisiensi sekaligus pemahaman terhadap model dapat meningkat.

Penelitian ini berfokus pada pembangunan model prediksi menggunakan data mining dengan metode algoritma C4.5, yang menghasilkan pohon keputusan (*decision tree*). Pengujian dilakukan melalui aplikasi RapidMiner untuk memprediksi penyakit monkeypox, sehingga penanganannya dapat dilakukan secara lebih cepat dan tepat (Ekadewi & Louisa Prayitno, 2024)

### Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*) yang diperkenalkan oleh J. Ross Quinlan pada tahun 1996. Dibandingkan dengan pendahulunya, C4.5 mampu menghasilkan pohon keputusan yang lebih baik dan lebih andal (Sukri & Handrianto, 2024). Penelitian terdahulu juga memanfaatkan algoritma ini dalam membangun model prediksi melalui pendekatan data mining. Misalnya, (Ekadewi & Louisa

Prayitno, 2024) menggunakan C4.5 yang diimplementasikan pada aplikasi RapidMiner untuk memprediksi penyakit monkeypox, sehingga dapat ditangani secara lebih cepat dan tepat.

Secara umum, proses kerja algoritma C4.5 dimulai dengan membentuk struktur pohon keputusan melalui pemilihan atribut yang memiliki nilai *gain ratio* tertinggi sebagai akar pohon. *Gain ratio* berfungsi untuk mengukur seberapa baik suatu atribut dalam memisahkan data ke dalam kelas-kelas target. Pemilihan atribut dilakukan secara rekursif, di mana setiap simpul pada pohon merepresentasikan kondisi atau aturan tertentu berdasarkan atribut yang dipilih. Proses ini terus berlanjut hingga terbentuk cabang-cabang yang berakhir pada keputusan akhir berupa kelas target.

### ***Algoritma Naïve Bayes***

Algoritma Naive Bayes adalah salah satu algoritma dalam teknik klasifikasi. Naive Bayes merupakan metode pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Meskipun konsepnya sederhana, algoritma ini sering menunjukkan performa tinggi pada dataset berskala besar dan variabel dengan kecenderungan statistik yang kuat. Selain itu, Naive Bayes dikenal karena kemampuannya dalam mengklasifikasikan teks dengan baik (Pratmanto et al., 2024)

### **Evaluasi Model**

Evaluasi model dilakukan untuk mengetahui seberapa baik algoritma C4.5 dan Naïve Bayes dalam memprediksi kualitas tidur pada kesehatan. Beberapa ukuran yang digunakan dalam evaluasi ini adalah akurasi, presisi, recall,

TP : True Positif (Model memprediksi positif atau 1, ternyata benar prediksi)

TN : True Negatif (Model memprediksi negatif atau 0, ternyata benar prediksi negatif)

FP : False Positif (Memprediksi positif atau 1, ternyata salah prediksi nya)

FN : False Negatif (Memprediksi Negatif atau 0, ternyata salah prediksi)

### ***Accurasi***

Merupakan rasio prediksi benar (positif dan negative) dengan keseluruhan data. Akurasi = Jumlah data yang benar/jumlah semua data uji\*100%

### ***Precision***

Merupakan rasio presik benar positif disbanding dengan keseluruhan hasil yang diprediksi. Precision :  $TP/(TP+FP)$

**Recall**

Merupakan rasio prediksi benar positif dibanding dengan keseluruhan data yang benar positif. Recall :  $TP/(TP+FN)$

**3. METODE PENELITIAN**

Penelitian ini bertujuan untuk melakukan perbandingan kinerja dua algoritma klasifikasi, yaitu C4.5 dan Naïve Bayes, dalam memprediksi kualitas tidur berdasarkan data kesehatan dan gaya hidup. Pendekatan yang digunakan adalah kuantitatif dengan memanfaatkan data sekunder yang diperoleh dari platform Kaggle sebagai sumber dataset penelitian.

**Sumber Data**

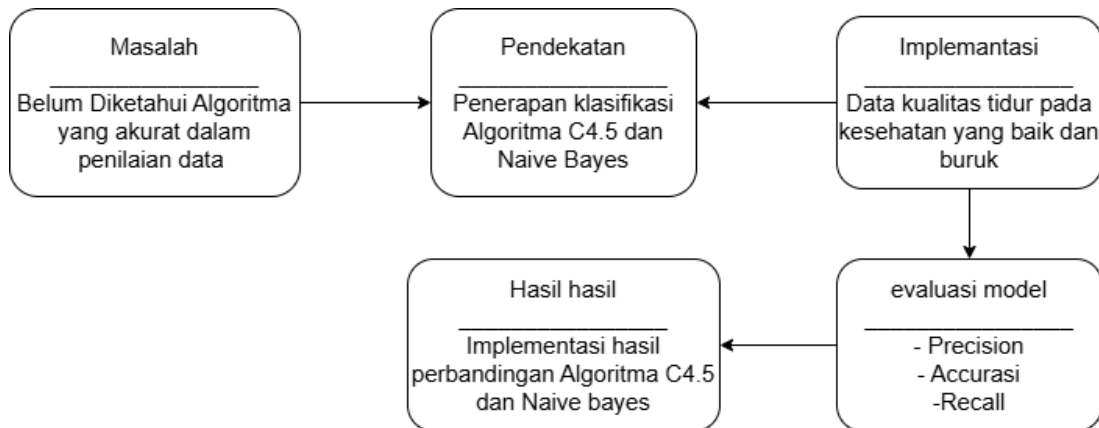
Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari platform Kaggle, dengan nama dataset "Sleep Health and Lifestyle Dataset" yang melibatkan 374 responden. Dataset ini dipilih karena menyediakan berbagai fitur relevan yang mencakup aspek kesehatan dan gaya hidup yang berpotensi memengaruhi kualitas tidur, seperti usia, jenis kelamin, pekerjaan, durasi tidur, tingkat aktivitas fisik, tingkat stres, kategori BMI, tekanan darah, detak jantung, langkah harian, dan gangguan tidur. Variabel target dalam penelitian ini adalah "Quality of Sleep", yang merepresentasikan kualitas tidur individu. Berikut sampel himpunan datanya:

**Tabel 1.** Sampel Himpunan Data.

No	Nama Atribut	Tipe data	Deskripsi
1	Gender	Kategorikal	Jenis kelamin responden (contoh: Male, Female)
2	Age	Numeric	Usia responden dalam tahun
3	Sleep duration	Numeric	Lama tidur rata-rata dalam satu hari (dalam jam)
4	Quality of sleep	Kategorikal	Kualitas tidur (contoh: Buruk, Cukup, Baik)
5	Pyhsical activity level	Numeric	Tingkat aktivitas fisik harian (semakin tinggi semakin aktif)
6	Stress level	Numeric	Tingkat stres (skala 1–10, makin tinggi makin stres)
7	BMI kategori	Kategorikal	Kategori indeks massa tubuh (Underweight, Normal, Overweight)
8	Blood pressure	Kategorikal	Tekanan darah dalam format string (misalnya: 126/83)
9	Heart rate	Numeric	Detak jantung per menit
10	Daily step	Kategorikal	Jumlah langkah harian (format teks, contoh: "4,200", "10,000") — perlu konversi numerik untuk analisis lanjut

11	Sleep disorder	kategorikal	Jenis gangguan tidur (contoh: None, Sleep Apnea, Insomnia)
----	----------------	-------------	--

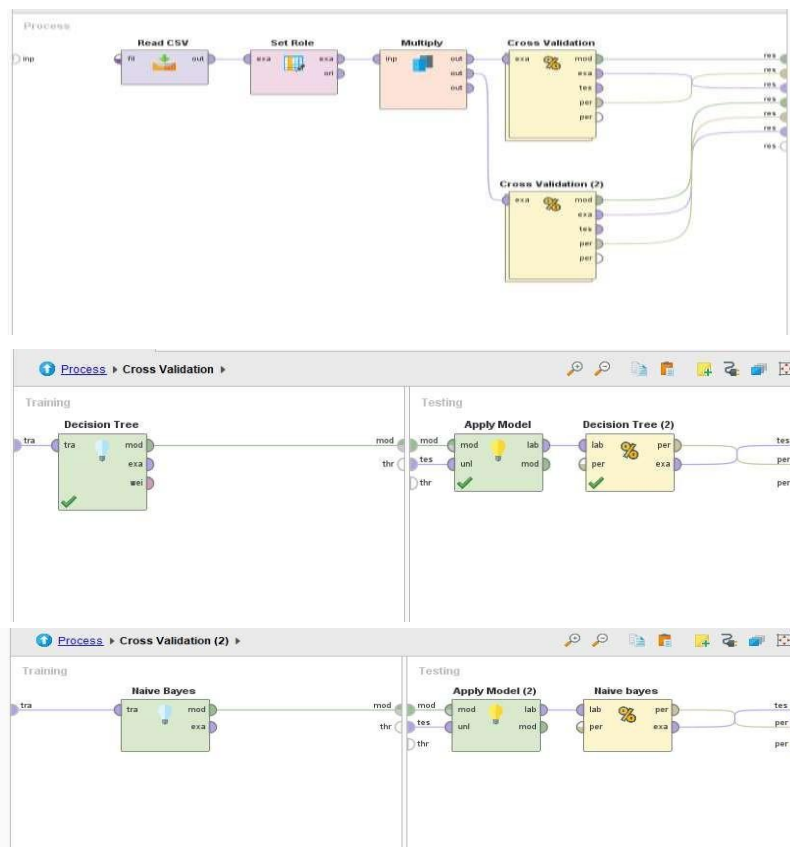
**Kerangka berfikir**



**Gambar 1.** Kerangka berfikir.

**4. HASIL DAN PEMBAHASAN**

**Pengujian Model**



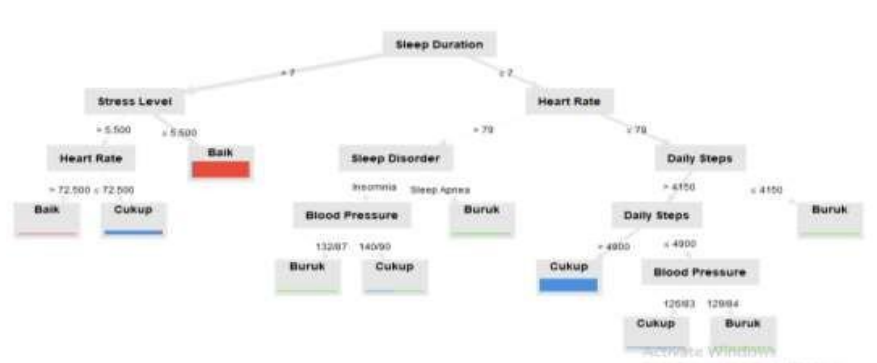
**Gambar 2.** Pengujian model.

Dalam proses yang ditampilkan pada gambar, peneliti ini menggunakan dua algoritma klasifikasi, yaitu C4.5 (Decision Tree) dan Naive Bayes, untuk membandingkan performa dalam memprediksi kualitas tidur. Dataset terlebih dahulu dibaca melalui operator Read CSV,

lalu diberi peran target klasifikasi menggunakan Set Role. Selanjutnya, data diduplikasi menggunakan Multiply, sehingga dapat digunakan oleh dua proses evaluasi yang berbeda secara paralel.

Kedua algoritma dievaluasi melalui teknik Cross Validation sebanyak 10 lipatan (10-fold cross validation), yang bertujuan untuk mengukur akurasi model secara objektif dan menghindari overfitting. Pada sisi training, model dibentuk berdasarkan algoritma yang dipilih (Decision Tree atau Naive Bayes), kemudian pada fase testing, model diterapkan pada data uji dan hasil klasifikasinya dievaluasi menggunakan metrik performa.

### Algoritma C4.5



Gambar 3. Pohon Keputusan.

### Entropy Awal (1.173)

Nilai ini dihitung dari distribusi kelas pada kolom target **Quality of Sleep**. Karena kelasnya ada tiga (*Baik*, *Cukup*, *Buruk*), maka entropy dihitung dengan rumus:

$$Entropy = \sum_{i=1}^c -P_i * \log_2(P_i)$$

### Information Gain

Information Gain mengukur seberapa besar penurunan ketidakpastian setelah data dibagi berdasarkan suatu atribut.

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

Baik = 180

Cukup = 182

Buruk = 12

Total = 374

- Hitung Probabilitas Tiap Kelas

PBaik = 180/374 = 0,481 PCukup =

182/374 = 0,487 PBuruk = 12/374

= 0,032

- Rumus Entropy

Maka :  $Entropy(S) = -(0.481 \cdot \log_2(0.481) + 0.487 \cdot \log_2(0.487) + 0.032 \cdot \log_2(0.032))$

- Perhitungan

Baik =  $0.481 \cdot \log_2(0.481) = 0.481 \times -1.057 = -0.509$  Cukup =

$0.487 \cdot \log_2(0.487) = 0.487 \times -1.041 = -0.507$  Buruk :

$0.032 \cdot \log_2(0.032) = 0.032 \times -4.971 = -0.159$

Jumlah:

$Entropy(S) = -(-0.509 - 0.507 - 0.159) = 1.175$

Berikut nilai Baik, Cukup, Buruk pada setiap atribut:

**Tabel 1. Gender.**

Count of Quality of Sleep Row Labels	Column Labels			Grand Total
	Baik	Buruk	Cukup	
Female	105	6	74	185
Male	75	6	108	189
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 2. Age.**

Count of Quality Sleep of Row Labels	Column Labels			Grand Total
	Baik	Buruk	Cukup	
27			1	1
28		3	2	5
29		2	11	13
30		2	11	13
31	1		17	18
32	3		14	17
33			13	13
34		2		2
35	9		3	12
36	10	2		12
37	19		1	20
38	19		1	20
39	13	1	1	15
40	4			4
41	10		2	12
42	6		3	9

43	17	17	34	
44		30	30	
45		14	14	
48		3	3	
49	2	9	11	
50	1	19	20	
51	4	4	8	
52	3	6	9	
53	17		17	
54	7		7	
55	2		2	
56	2		2	
57	9		9	
58	6		6	
59	16		16	
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 3.** Occupation.

Count of Quality of Sleep	Column Labels			Grand Total
	Baik	Buruk	Cukup	
Accountant	31		6	37
Doctor	4		67	71
Engineer	60	1	2	63
Lawyer	42		5	47
Manager			1	1
Nurse	35	4	34	73
Sales Representative		2		2
Salesperson			32	32
Scientist		2	2	4
Software Engineer	2	1	1	4
Teacher	6	2	32	40
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 4.** Sleep Duration.

Count of Quality of Sleep  Row Labels	Column Labels			Grand Total
	Baik	Buruk	Cukup	
5.8		2		2
5.9		3	1	4
6			31	31
6.1			25	25
6.2			12	12
6.3			13	13
6.4		2	7	9
6.5		3	23	26
6.6		2	18	20
6.7			5	5
6.8			5	5
6.9			3	3
7.1	16		3	19
7.2	36			36
7.3	12		2	14
7.4	3		2	5
7.5	5			5
7.6	5		5	10
7.7	10		14	24
7.8	19		9	28
7.9	3		4	7
8	13			13
8.1	15			15
8.2	11			11
8.3	5			5
8.4	14			14
8.5	13			13
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 5.** Physical Aktivitiy Level.

Count of Quality of Sleep Row Labels	Column Labels			Grand Total
	Baik	Buruk	Cukup	
30	32	3	33	68
32		2		2
35		4		4
40		3	3	6
42			2	2
45	2		66	68
47			1	1
50	1		3	4
55	2		4	6
60	66		4	70
65			2	2
70	3			3
75	35		32	67
80	2			2
85	2			2
90	35		32	67
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 6.** Stress Level.

Count of Quality Sleep of Row Labels	Labels			Grand Total
	Baik	Buruk	Cukup	
3	71			71
4	45		25	70
5	60		7	67
6	4		42	46
7		7	43	50
8		5	65	70
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 7.** BMI kategorikal.

Count of Quality of Sleep Row Labels	Column Labels			Grand Total
	Baik	Buruk	Cukup	
Normal	130		65	195
Normal Weight	14	4	3	21
Obese	2	3	5	10
Overweight	34	5	109	148
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 8.** Blood Pressure.

Count of Quality of Sleep	Column Labels			Grand
	Row Labels	Baik	Buruk	Cukup
115/75	32			32
115/78	2			2
117/76	1		1	2
118/75	2			2
118/76	1			1
119/77	2			2
120/80	7		38	45
121/79	1			1
122/80	1			1
125/80	38		27	65
125/82			4	4
126/83			2	2
128/84			2	2
128/85			3	3
129/84		2		2
130/85	58		41	99
130/86		2		2
131/86		2		2
132/87		3		3
135/88			2	2
135/90			27	27
139/91	2			2
140/90		3	1	4
140/95	33		32	65
142/92			2	2
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 9.** Blood Pressure.

Count of Quality of Sleep	Column Labels			Grand
	Row Labels	Baik	Buruk	Cukup
65	40		27	67
67	2			2
68	94			94
69	1		1	2
70	36		40	76
72	3		66	69

73	2			2
74		2		2
75			36	36
76			2	2
77			2	2
78		2	3	5
80		3		3
81		2		2
82			1	1
83			2	2
84			2	2
85		3		3
86	2			2
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 10.** Daily steps.

<b>Count of Quality of Sleep</b>	<b>Column Labels</b>			<b>Grand</b>
<b>Row Labels</b>	<b>Baik</b>	<b>Buruk</b>	<b>Cukup</b>	<b>Total</b>
3000		3		3
3300			2	2
3500			3	3
3700	2			2
4000		3		3
4100		2		2
4200			2	2
4800		2		2
5000	40		28	68
5200		2		2
5500			4	4
5600			2	2
6000			68	68
6200	1			1
6800	2		1	3
7000	66			66
7300	2			2
7500	2			2
8000	63		38	101
10000	2		34	36
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

**Tabel 11.** Sleep Disorder.

Count of Quality of Sleep	Column Labels			Grand
Row Labels	Baik	Buruk	Cukup	Total
Insomnia	6	5	66	77
None	139		80	219
Sleep Apnea	35	7	36	78
<b>Grand Total</b>	<b>180</b>	<b>12</b>	<b>182</b>	<b>374</b>

Berikut hasil nilai entropy dan gain pada setiap atributnya

**Tabel 12.** Entropy dan Gain.

Peringkat	Atribut	Entropy Awal	Information Gain
1	Sleep Duration	1.173	0.842
2	Age	1.173	0.753
3	Heart Rate	1.173	0.730
4	Stress Level	1.173	0.710
5	Daily Steps	1.173	0.700
6	Occupation	1.173	0.569
7	Blood Pressure	1.173	0.481
8	Physical Activity Level	1.173	0.460
9	BMI Category	1.173	0.197
10	Sleep Disorder	1.173	0.186
11	Gender	1.173	0.022

accuracy: 96.25% +/- 2.28% (micro average: 96.26%)

	true Cukup	true Buruk	true Baik	class precision
pred. Cukup	175	4	3	96.15%
pred. Buruk	2	8	0	80.00%
pred. Baik	5	0	177	97.25%
class recall	96.15%	66.67%	98.33%	

**Gambar 4.** Gambar Confusion Matrix Decision Tree.

**Precision : TP/(TP+FP)**

**Tabel 13.** precision decision Tree.

Baik	97,25%
Cukup	96,15%
Buruk	80,00%

$$\text{Cukup} = 175 / (175 + 4 + 3) = 0,9615 = 96,15\%$$

$$\text{Buruk} = 8 / (2 + 8 + 0) = 0,8 = 80,00\% \quad \text{Baik} = 177 / (5 + 0 + 177) = 0,9725 = 97,25\%$$

Recall

**Recall : TP/(TP+FN)**

**Tabel 14.** Recall Decision Tree.

Baik	98,33%
Cukup	96,15%
Buruk	66,67%

$$\text{Cukup} = 175 / (175 + 2 + 5) = 0,9670 = 96,70\% \quad \text{Buruk} = 8 / (4 + 8 + 0) = 0,6666 = 66,67\%$$

$$\text{Baik} = 177 / (3 + 0 + 177) = 0,9833 = 98,33\%$$

**Accurat**

Accurasi Akurasi = Jumlah data yang benar/jumlah semua data uji\*100%

$$360 / 374 * 100\% = 0,96256 = 96,26\%$$

**Algoritma Naïve Bayes**

```

SimpleDistribution
Distribution model for label attribute Quality of Sleep

Class Cukup (0.487)
11 distributions

Class Buruk (0.032)
11 distributions

Class Baik (0.481)
11 distributions
    
```

**Gambar 5.** Simpledistribution Naive Bayes.

Baik = 180

Cukup = 182

Buruk = 12

Total = 374

- Hitung Probabilitas Tiap Kelas

$$P_{\text{Baik}} = 180/374 = 0,481$$

$$P_{\text{Cukup}} = 182/374 = 0,487 \quad P_{\text{Buruk}}$$

$$= 12/374 = 0,032$$

accuracy: 98.66% +/- 1.90% (micro average: 98.66%)

	true Cukup	true Buruk	true Baik	class precision
pred. Cukup	179	0	2	98.90%
pred. Buruk	2	12	0	85.71%
pred. Baik	1	0	178	99.44%
class recall	98.35%	100.00%	98.89%	

**Gambar 6.** Confusion Matrix Naive Bayes.

### Precision

**Precision :  $TP/(TP+FP)$**

**Tabel 15.** Precison naive Bayes.

Baik	99,44%
Cukup	98,90%
Buruk	85,71%

$$\text{Cukup} = 179 / (179 + 0 + 2) = 0,9889 = 98,90\%$$

$$\text{Buruk} = 12 / (2 + 12 + 0) = 0,8571 = 85,71\%$$

$$\text{Baik} = 178 / (1 + 0 + 178) = 0,9944 = 99,44\%$$

### Recall

**Recall :  $TP/(TP+FN)$**

**Tabel 16.** Recall Naive Bayes.

Baik	98,89%
Cukup	98,35%
Buruk	100%

$$\text{Cukup} = 179 / (179 + 0 + 2) = 0,9889 = 98,89\%$$

$$\text{Buruk} = 12 / (0 + 12 + 0) = 1 = 100\%$$

$$\text{Baik} = 178 / (2 + 0 + 178) = 0,9888 = 98,89\%$$

**Accurat**

**Accurasi Akurasi = Jumlah data yang benar/jumlah semua data uji\*100%**

$369/374*100%=0,9866=98,66\%$

**5. KESIMPULAN DAN SARAN****Kesimpulan**

Berdasarkan hasil penelitian mengenai perbandingan algoritma C4.5 dan Naïve Bayes dalam memprediksi kualitas tidur, dapat ditarik beberapa poin penting. Faktor yang paling berpengaruh terhadap kualitas tidur adalah tingkat stres, yang terbukti menjadi penentu utama dalam klasifikasi data. Algoritma C4.5 mampu memberikan gambaran yang cukup jelas melalui visualisasi pohon keputusan dengan tingkat akurasi sebesar 96,26%. Keunggulan ini menjadikan C4.5 lebih mudah dipahami oleh pengguna non-teknis karena mampu menunjukkan hubungan antarvariabel secara terstruktur. Sementara itu, algoritma Naïve Bayes menunjukkan performa yang lebih unggul dengan tingkat akurasi mencapai 98,66%, disertai nilai presisi dan recall yang tinggi pada semua kelas. Hal ini membuktikan bahwa Naïve Bayes lebih efektif dalam melakukan klasifikasi kualitas tidur secara keseluruhan, terutama dalam mendeteksi kualitas tidur yang buruk. Dengan demikian, dapat disimpulkan bahwa Naïve Bayes layak dipilih sebagai metode yang lebih optimal untuk prediksi kualitas tidur, sedangkan C4.5 lebih sesuai digunakan ketika dibutuhkan interpretasi hasil yang mudah dipahami.

**DAFTAR PUSTAKA**

- Anggi Trifani, Agus Perdana Windarto, & Hendry Qurniawan. (2022). Penerapan Data Mining Klasifikasi C4.5 dalam Menentukan Tingkat Stres Mahasiswa Akhir. *Jurnal Riset Rumpun Ilmu Teknik*, 1(2), 91–105. <https://doi.org/10.55606/jurritek.v1i2.414>
- Cahya, D., Buani, P., Rahmawati, A., Informatika, P. S., Mandiri, U. N., Studi, P., Informasi, S., Mandiri, U. N., Foundation, N. S., & Searchev, R. (2025). PENERAPAN DECISION TREE UNTUK KLASIFIKASI GANGGUAN TIDUR. 10(1), 66–74.
- Daewana, F. (2025). Perbandingan Algoritma Naïve Bayes Dan C4.5 Dalam Prediksi Keputusan Karyawan Untuk Meninggalkan Perusahaan. *Innovative: Journal Of Social Science Research*, 5(1), 1712–1726. <https://doi.org/10.31004/innovative.v5i1.17664>
- Ekadewi, I., & Louisa Prayitno, L. (2024). Simpatik: Jurnal Sistem Informasi dan Informatika Penerapan Metode Algoritma C4.5 Pada Diagnosis Penyakit MonkeyPox. 4(1), 1–8. <http://jurnal.bsi.ac.id/index.php/simpatik>

- F. Narti, "JTIM : Jurnal Teknologi Informasi dan Multimedia Perbandingan Algoritma C4 . 5 dan Naive Bayes dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," vol. 4, no. 1, 2022.
- Hardian, A., Sitepu, E., Mulyapradana, A., Sitopu, J. W., Wardono, B. H., Bina, U., Informatika, S., Agung, U. D., & Simalungun, U. (2025). Indonesian Research Journal on Education. 5(2022), 1079–1085.
- Imam Nawawi, & Zaehol Fatah. (2024). Penerapan Decision Trees dalam Mendeteksi Pola Tidur Sehat Berdasarkan Kebiasaan Gaya Hidup. *Jurnal Ilmiah Sains Teknologi Dan Informasi*, 2(4), 34–41. <https://doi.org/10.59024/jiti.v2i4.969>
- Pratmanto, D., Widayanto, A., Kristania, Y. M., & Wijianto, R. (2024). Analisis Perbandingan Algoritma Naive Bayes Dan KNN Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Vidio Di Google Play Store. 4(2), 119–124.
- Putra, A. R., & Program. (2024). Penerapan Teknik Data Mining Terhadap Prediksi Masa Studi Mahasiswa. 1, 33–37.
- Sastra, A., & Sabri, K. (2025). Analisis Data Science terhadap Kualitas Tidur dan Faktor-Faktor yang Mempengaruhinya : Studi Kasus Sleep Health and Lifestyle Dataset. 3(2), 92–99.
- Schneider, K.-M. (2005). Techniques for Improving the Performance of Naive Bayes for Text Classification. *Computational Linguistics and Intelligent Text Processing*, (i), 682–693. [https://doi.org/10.1007/978-3-540-30586-6\\_76](https://doi.org/10.1007/978-3-540-30586-6_76)
- Song, J. H., & Woo, H. Y. (2015). A study on AQ (adversity quotient), job satisfaction and turnover intention according to work units of clinical nursing staffs in Korea. *Indian Journal of Science and Technology*, 8(November), 74–78. <https://doi.org/10.17485/ijst/2015/v8i>
- Sukri, M. H., & Handrianto, Y. (2024). Penerapan Algoritma C4. 5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta. *Informatics and Computer Engineering Journal*, 4(1), 11–24.
- Suyadi, S., Setyanto, A., & Fattah, H. Al. (2017). Analisis Perbandingan Algoritma Decision Tree (C4.5) Dan K-Naive Bayes Untuk
- Vidi, M. A., Ridha, A., Kansha, W. M., Studi, P., Rekayasa, T., Lunak, P., Vokasi, S., Studi, P., Komputer, I., Data, S. S., Informasi, S. T., Vidi, M. A., National, P., Of, C., & Indonesia, S. S. (2025). *Proceeding National Conference of Research and Community Service Sisi Indonesia*.