Penerapan Data Mining terhadap Pengaruh Ritme Circadian pada Performa Belajar Mahasiswa

U. Darmanto Soer[1)*)], Sifa Fauziah[2)], Ridho Yusuf Fitra[3)]

[1)2)3)]Teknik Informatika, Universitas Pelita Bangsa
[*)]**Correspondence author**: darmantosoer@pelitabangsa.ac.id, Cikarang, Indonesia
DOI: https://doi.org/10.37012/jtik.v10i2.2408

Abstrak

Perkembangan teknologi informasi telah menghasilkan data yang mencakup berbagai bidang kehidupan, termasuk di bidang pendidikan perguruan tinggi. Teknologi informasi telah menciptakan banyak data dalam berbagai aspek kehidupan, termasuk pendidikan. Di perguruan tinggi, data mahasiswa dan proses pembelajaran memberikan peluang untuk meningkatkan kualitas pendidikan, tetapi persaingan antar perguruan tinggi juga semakin kompetitif. Data mahasiswa di perguruan tinggi menyimpan informasi berharga mengenai profil dan hasil akademik mereka. Namun, menghadapi tuntutan zaman yang semakin produktif, manusia sering mengalami stres dan gangguan tidur yang dapat berdampak negatif pada performa dan kesehatan. Tidur dan ritme circadian memainkan peran penting dalam kesehatan dan produktivitas manusia. Gangguan tidur dan ritme circadian yang terganggu dapat mempengaruhi konsentrasi, memori, dan pemecahan masalah mahasiswa. Dalam konteks ini, menjaga ritme circadian yang sehat menjadi kunci penting dalam meningkatkan performa belajar mahasiswa. Penelitian ini bertujuan untuk mengkaji pengaruh ritme circadian terhadap performa belajar mahasiswa dengan menggunakan algoritma C4.5. Dengan sampel sebanyak 163 mahasiswa dari populasi sejumlah 295 mahasiswa. Sampel dibagi menjadi data training dan data testing menggunakan teknik linear sampling. Hasil dari perhitungan menggunakan algoritma C4.5 menunjukkan tingkat accuracy sebesar 81.81%, precision sebesar 96.29%, dan recall sebesar 83.87%. Hasil decision tree menunjukan adanya hubungan antara ritme circadian dengan performa belajar mereka, semakin sering mahasiswa bergadang akan mempengaruhi pencapaian akademik mereka.

Kata Kunci: Ritme Circadian, Algoritma C4.5, Performa Belajar, Mahasiswa

Abstract

The development of information technology has produced data covering various areas of life, including in the field of higher education. Information technology has created a lot of data in various aspects of life, including education. In higher education, student data and the learning process provide opportunities to improve the quality of education, but competition between universities is also increasingly competitive. Student data in higher education holds valuable information regarding their academic profile and results. However, facing the demands of an increasingly productive era, humans often experience stress and sleep disorders which can have a negative impact on performance and health. Sleep and circadian rhythms play an important role in human health and productivity. Sleep disorders and disrupted circadian rhythms can affect students' concentration, memory and problem solving. In this context, maintaining a healthy circadian rhythm is an important key in improving student learning performance. This research aims to examine the influence of circadian rhythms on student learning performance using the C4.5 algorithm. With a sample of 163 students from a population of 295 students. The sample is divided into training data and testing data using linear sampling techniques. The results of calculations using the C4.5 algorithm show an accuracy rate of 81.81%,

precision of 96.29%, and recall of 83.87%. The decision tree results show that there is a relationship between circadian rhythms and their learning performance, the more often students stay up late, the more it will affect their academic achievement.

Keywords: *Circadian Rhythm, C4.5 algorithm, Study performance, Students*

PENDAHULUAN

Teknologi informasi telah menciptakan banyak data dalam berbagai aspek kehidupan, termasuk pendidikan. Di perguruan tinggi, data mahasiswa dan proses pembelajaran memberikan peluang untuk meningkatkan kualitas pendidikan, tetapi persaingan antar perguruan tinggi juga semakin kompetitif (Purba & Sitorus, 2022). Kemajuan zaman membawa produktivitas namun juga stres pada kesehatan fisik dan mental (Johanes, E.Manampiring, & Fatimawali., 2020). Data mahasiswa mencerminkan kinerja perguruan tinggi dalam menghasilkan lulusan berkualitas. Gangguan tidur dan ritme circadian memengaruhi produktivitas mahasiswa. Kebiasaan begadang, sering dipicu oleh tugas atau hiburan malam, dapat merusak performa belajar. Dalam lingkungan perkuliahan yang kompleks, stres dapat mengganggu mahasiswa. Pengelolaan stres dan pola tidur yang baik memiliki peran penting dalam menjaga performa akademik (Fauzi, 2016). artikel ini bertujuan menggunakan data mining untuk menguji hubungan antara ritme circadian dan performa belajar mahasiswa, dengan harapan memberikan wawasan baru tentang kaitannya.

METODE

Objek Penelitian

Objek penelitian merupakan komponen yang mengungkapkan informasi tentang objek dan subjek yang terlibat dalam penelitian. Objek penelitian mengacu pada elemen yang dipelajari, termasuk alat, data, kueri, dan proses eksekusi yang digunakan untuk mencapai hasil penelitian (Editor, 2022).

Subjek penelitian adalah mahasiswa aktif yang masih berkuliah dan yang sudah lulus menjadi fokus utama dalam penelitian ini. Objek penelitian adalah pengaruh ritme circadian pada performa belajar mahasiswa. Penelitian ini bertujuan untuk mengidentifikasi hubungan

antara ritme circadian (termasuk pola tidur, aktivitas fisik, dan faktor terkait waktu) dengan performa belajar mahasiswa tersebut.

Jenis Data

Jenis data penelitian ini berkaitan dengan cara penulis mengumpulkan informasi. Penelitian ini menggunakan data kuantitatif (dalam skala ordinal) dari jawaban-jawaban mahasiswa tentang kondisi mereka, pengalaman begadang, dan dampaknya.

Data yang Digunakan

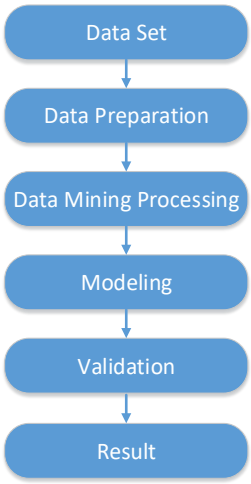
Sumber data merujuk pada asal informasi yang digunakan. Ini bisa dari individu, kelompok, wilayah, atau objek (Hans, 2022). Dalam penelitian ini, peneliti menggunakan data langsung dari mahasiswa sebagai responden melalui kuesioner. Sumber data primer terfokus pada informasi yang relevan dengan penelitian ini, dari mahasiswa aktif dan yang sudah lulus.

Metode yang Digunakan

Metode penelitian adalah cara yang digunakan untuk mengumpulkan informasi yang akurat dan bisa dipercaya. Ini membantu menjawab pertanyaan penelitian atau menguji hipotesis. Dengan metode yang cocok, hasil penelitian jadi lebih bermakna (Sugiyono, 2012). Penelitian ini adalah survei dengan menggunakan kuesioner untuk mengumpulkan data dari beberapa orang yang mewakili sekelompok orang. Survei membantu menjelaskan, memahami, atau menguji hubungan antara hal-hal tertentu dalam kelompok tersebut (Saunders, Lewis, & Thornhill, 2023). Dalam penelitian ini, peneliti menggunakan metode deskriptif analitis dengan pendekatan survei. Ini artinya peneliti secara sistematis menggambarkan objek atau fenomena yang diteliti untuk memahami fakta-fakta, sifat-sifat, dan hubungan di antara mereka (Nazir, 2005).

Data Mining

Penelitian ini menggunakan Python untuk tahap preprocessing data dan metode algoritma C4.5 untuk analisa dengan RapidMiner. Data yang ada di hitung dengan algoritma sesuai dengan metodenya, lalu dicari hasil akurasinya. Adapun tahapan pengujian data seperti berikut:



Gambar 1. Tahapan Pengujian

Algoritma C4.5 adalah pengembangan dari algoritma ID3 oleh Quinlan. Dalam C4.5, pohon keputusan digambarkan sebagai struktur pohon dengan node internal mewakili atribut keputusan, cabang-cabang mewakili hasil atribut yang diuji, dan daun mewakili kelas. Pohon ini mudah diubah menjadi aturan klasifikasi. Algoritma semacam C4.5 umumnya akurat, tetapi hasilnya dipengaruhi oleh karakteristik data yang diproses (Evicienna. & Amalia, 2013). C4.5 dapat menangani atribut kategorikal dan numerik. Untuk kedua jenis ini, C4.5 menggunakan informasi (*information gain*) yang diperoleh dan rasio keuntungan (*gain ratiom*) untuk pilih atribut terbaik. Adapun langkah-langkah algoritma C4.5 adalah sebagai berikut:

1. Persiapkan data training.
2. Hitung entropy. Entropy adalah ukuran ketidakpastian dalam keputusan atribut. Semakin tinggi entropy, semakin tinggi ketidakpastian. Entropy dihitung dengan rumus

$$Entropy\ (S) = \sum_{i=1}^{n} -p_i \times \log_2 p_i \quad \text{..... (1)}$$

Dimana :

S = Himpunan Kasus

n = Jumlah Partisi S

pi = Probabilitas (Jumlah "ya" dibagi total kasus)

3. Hitung gain. Gain mengukur pengaruh atribut dalam mengklasifikasikan data. Gain dihitung dengan rumus:

$$Gain = S - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times (S_I) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Dimana :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi S

|*Si*| = Jumlah Kasus pada Partisi ke-i

|*S*| = Jumlah Kasus dalam S.

Nilai gain digunakan untuk memilih variabel yang menjadi node dalam pohon keputusan. Variabel dengan gain tertinggi menjadi node dalam decision tree.

Data Cleaning

Langkah ini melibatkan menghapus data yang tidak dibutuhkan, memperbaiki data yang rusak atau hilang, dan menghilangkan data yang tidak valid. Dalam penelitian ini, peneliti akan menghapus data yang tidak memiliki nilai (*missing value*) dan *value* pada kolom IPK, mengubah tipe data pada kolom IPK menjadi float, serta menghapus data mahasiswa yang tidak aktif.

Tabel 1. Dataset setelah Proses Cleaning

Jumlah Mahasiswa

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Sudah lulus	17	6.2	6.2	6.2
	Aktif	257	93.8	93.8	100.0
	Total	274	100.0	100.0	

Data Selection

Dalam proses ini, peneliti akan memilih atribut yang relevan untuk digunakan dalam proses pemodelan dan menentukan atribut yang akan dijadikan sebagai label. Pemilihan atribut ini didasarkan pada tujuan penelitian. Atribut yang akan digunakan meliputi:

Tabel 2. Pemilihan Atribut

No.	Atribut	Indikator
1.	Aktivitas bergadang	✓
2.	Kecukupan tidur	✓
3.	IPK Mahasiswa	✓
4.	Dampak negatif buruknya pola tidur	✓

Data Transformation

Tahap ini mengubah data ke format yang berbeda agar sesuai dengan proses data processing. Tujuannya adalah mencegah nilai atribut memiliki jangkauan yang sangat besar, yang bisa membuat pengenalan pola data dan pembentukan decision tree menjadi lambat. Pada penelitian ini atribut dengan value yang akan diubah adalah:

Tabel 3. Transformasi Atribut IPK

Nilai	Transformasi
3.50 – 4.00	A
3.00 – 3.49	B
2.76 – 2.99	C
2.00 – 2.75	D

Tabel 4. Transformasi Atribut Berdampak

Nilai	Transformasi
“Agak Setuju”, “Setuju”, “Sangat Setuju”	Ya
“Agak Tidak Setuju”, “Tidak Setuju”, “Sangat Tidak Setuju	Tidak

Split Validation

Split Validation adalah teknik validasi yang membagi data menjadi data pelatihan dan data pengujian secara acak. Dalam penelitian klasifikasi data kuesioner, data disiapkan dan dibagi menjadi dua untuk analisis dan evaluasi. Jumlah sampel

ditentukan menggunakan rumus slovin dengan *error sampling* 5%, kemudian sampel dipilih secara acak untuk memastikan setiap anggota populasi memiliki kesempatan yang sama untuk dipilih sebagai sampel. Perhitungannya adalah sebagai berikut:

$$n = \frac{274}{1 + (274)(0{,}05^2)} \approx 162{,}61128 = 163$$

Dari 163 sampel yang akan digunakan, dibagi dengan menggunakan linear sampling. Perhitungan pembagian sampel diambil dari atribut "dampak negatif buruknya pola tidur" sebagai berikut:

$$data\ training = \frac{800}{1000}163 = 130$$

$$data\ testing = \frac{200}{1000}163 = 33$$

HASIL DAN PEMBAHASAN

Perhitungan Algoritma C4.5

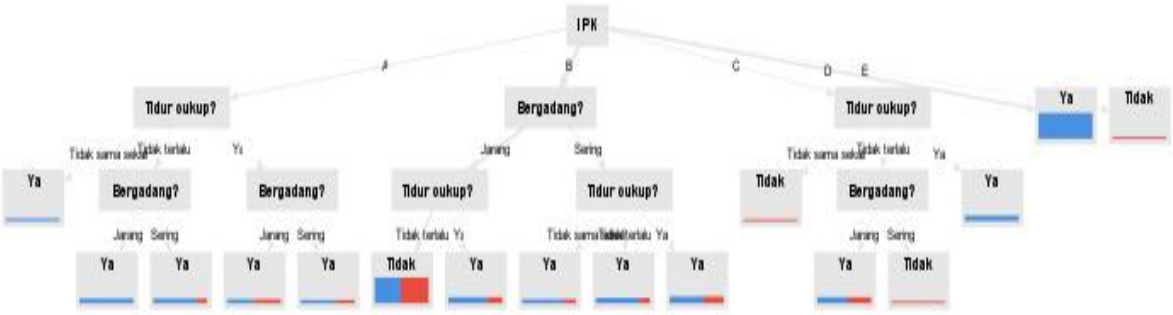
Pohon keputusan dibuat dari perhitungan *Entropy* dan *Gain*. Setelah pohon terbentuk, langkah berikutnya adalah mencari aturan dari cabang-cabangnya. Berikut adalah hasil perhitungan dengan Algoritma C4.5:

Tabel 5. Perhitungan C4.5

Node			Jumlah Kasus (S)	Tidak (S1)	Ya (S2)	Entropy	Gain
1	Total		130	95	35	0.840359	
	Bergadang?						0.067235
		Jarang	62	25	37	0.972806	
		Sering	68	10	58	0.602430	
	Tidur cukup?						0. 089264
		Tidak sama sekali	34	2	32	0.322757	
		Tidak terlalu	63	25	38	0.969063	
		ya	33	8	25	0.799049	
	IPK						0.178727
		A	19	4	15	0.742488	
		B	60	25	35	0.979869	
		C	15	5	10	0.918296	
		D	35	0	35	0	
		E	1	1	0	0	

Tabel di atas hasil data training dengan algoritma C4.5. Data dikelompokkan berdasarkan atribut dengan informasi gain tertinggi, yaitu nilai IPK. Ini menunjukkan bahwa nilai IPK memiliki informasi gain tertinggi untuk pembentukan pohon keputusan.

Decision Tree



Gambar 2. Pembentukan Pohon Keputusan

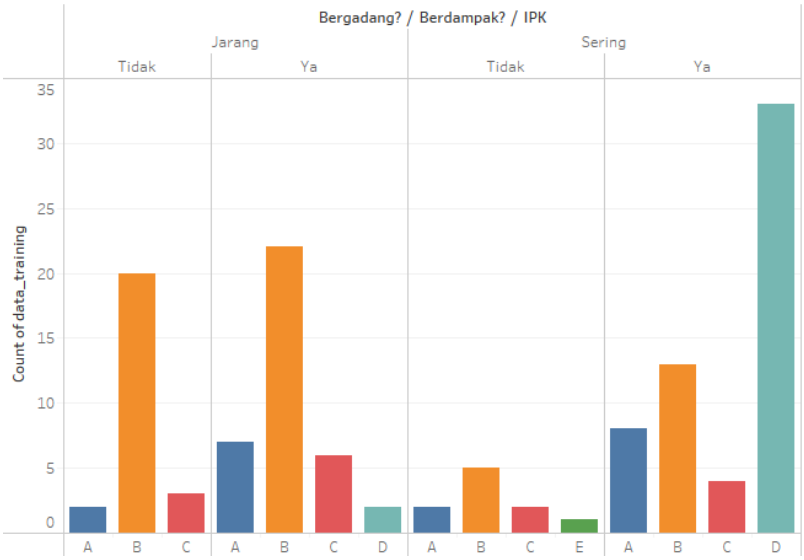
Di atas adalah *decision tree* dari data training mahasiswa. Dari pohon keputusan ini, ditemukan aturan yang menghubungkan mahasiswa dengan performa belajar dan ritme circadian, seperti berikut:

Tree

```
IPK = A
|    Tidur cukup? = Tidak sama sekali: Ya (Ya=2, Tidak=0)
|    Tidur cukup? = Tidak terlalu
|    |    Bergadang? = Jarang: Ya (Ya=5, Tidak=0)
|    |    Bergadang? = Sering: Ya (Ya=4, Tidak=1)
|    Tidur cukup? = Ya
|    |    Bergadang? = Jarang: Ya (Ya=2, Tidak=2)
|    |    Bergadang? = Sering: Ya (Ya=2, Tidak=1)
IPK = B
|    Bergadang? = Jarang
|    |    Tidur cukup? = Tidak terlalu: Tidak (Ya=17, Tidak=18)
|    |    Tidur cukup? = Ya: Ya (Ya=5, Tidak=2)
|    Bergadang? = Sering
|    |    Tidur cukup? = Tidak sama sekali: Ya (Ya=3, Tidak=1)
|    |    Tidur cukup? = Tidak terlalu: Ya (Ya=5, Tidak=1)
|    |    Tidur cukup? = Ya: Ya (Ya=5, Tidak=3)
IPK = C
|    Tidur cukup? = Tidak sama sekali: Tidak (Ya=0, Tidak=1)
|    Tidur cukup? = Tidak terlalu
|    |    Bergadang? = Jarang: Ya (Ya=4, Tidak=3)
|    |    Bergadang? = Sering: Tidak (Ya=0, Tidak=1)
|    Tidur cukup? = Ya: Ya (Ya=6, Tidak=0)
IPK = D: Ya (Ya=35, Tidak=0)
IPK = E: Tidak (Ya=0, Tidak=1)
```

Gambar 3. Struktur Pohon Keputusan

Hasil tersebut menunjukkan bahwa pola tidur dan kebiasaan begadang memengaruhi performa belajar mahasiswa. Tidur baik dan begadang sehat meningkatkan performa, sementara tidur buruk dan begadang tidak sehat merugikan. Namun, beberapa mahasiswa dengan IPK A merasa tidur cukup walaupun bergadang, dan beberapa mahasiswa IPK B merasa kurang tidur tanpa memengaruhi performa.



Gambar 4. Diagram Faktor Pengaruh Bergadang

Diagram di atas menunjukkan bahwa seringnya bergadang cenderung mempengaruhi performa belajar mahasiswa dengan nilai IPK indikator D, sedangkan jarang bergadang cenderung mempengaruhi performa belajar mahasiswa dengan nilai IPK indikator B. Ini menggambarkan bahwa sering bergadang memiliki pengaruh besar pada performa belajar mahasiswa.

Proses *Confusion Matrix*

Proses pembahasan confusion matrix dilakukan untuk mendapatkan nilai accuracy, precision, dan recall pada Algoritma C4.5. Berikut adalah hasil dari proses yang telah dilakukan:

Tabel 6. Confusion Matrix

accuracy: 81.82%

	true Ya	true Tidak	class precision
pred. Ya	26	1	96.30%
pred. Tidak	5	1	16.67%
class recall	83.87%	50.00%	

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$
$$= \frac{27}{33} \times 100\%$$
$$= 81.81\%$$
$$Precision = \frac{TP}{TP + FP} \times 100\%$$
$$= \frac{26}{26 + 1} \times 100\%$$
$$= 96.29\%$$
$$recall = \frac{TP}{TP + FN} \times 100\%$$
$$= \frac{26}{26 + 5} \times 100\%$$
$$= 83.87\%$$

Tabel 7. Hasil Perhitungan Confusion Matrix

Accuracy	Precision	Recall
81.81%	96.29%	83.87%

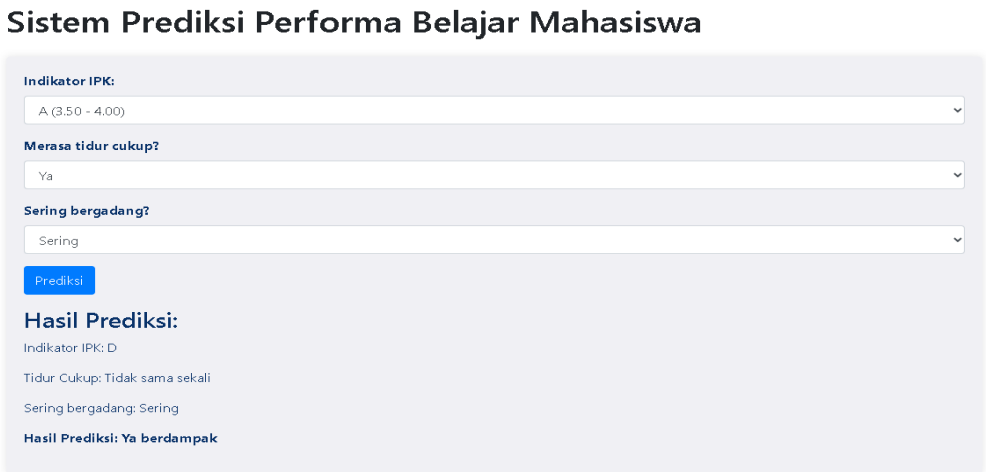
Setelah perhitungan, dapat ditarik kesimpulan dari sampel data 163 mahasiswa. Melalui algoritma C4.5 di RapidMiner, ditemukan hasil: akurasi 81.81%, ketepatan 96.29%, dan penemuan kembali 83.87%.

Hasil Implementasi Sistem

Hasil implementasi berupa sistem prediksi performa belajar mahasiswa dengan menentukan apakah dampak bergadang berpengaruh pada performa belajar mahasiswa dapat dilihat dari gambar berikut:

Gambar 5. Tampilan Input Sistem

Gambar 5 adalah tampilan awal sistem prediksi. Prosesnya yaitu mahasiswa mengisi informasi tentang pengalaman mereka, termasuk nilai IPK, perasaan tidur setelah bergadang, dan frekuensi begadang.



Gambar 6. Output Sistem Prediksi

KESIMPULAN DAN REKOMENDASI

Dari penelitian data mining mengenai pengaruh ritme circadian terhadap performa belajar mahasiswa menggunakan algoritma C4.5, disimpulkan bahwa:

1. Pola tidur berkaitan dengan performa belajar mahasiswa. Tidur buruk umumnya mengakibatkan performa belajar rendah.
2. Algoritma C4.5 efektif dalam menganalisis hubungan tersebut. Data training menunjukkan akurasi 81.81%, ketepatan 96.29%, dan penemuan kembali 83.87%.
3. Sistem prediksi ini bisa memberikan hasil prediksi performa belajar mahasiswa.Ucapan Terima Kasih

Adapun beberapa saran yang bisa diberikan oleh peneliti untuk pengembangan pada penelitian selanjutnya adalah:

1. Perlu meningkatkan keterkaitan ritme circadian dan performa belajar, misalnya dengan atribut lebih spesifik seperti durasi tidur saat bergadang, jam tidur rata-rata, dan lainnya.
2. Lakukan pengujian dengan metode data mining lain seperti ID3, CART, Naïve Bayes, untuk perbandingan yang lebih baik.

REFERENSI

Editor, S. (2022). *Mengenal Lebih Dalam Apa Itu Objek Penelitian*. (sampoernauniversity.ac.id)

Evicienna., & Amalia, H. (2013). Algoritma C4.5 Untuk Prediksi Hasil pemilihan Legislatif DPRD DKI Jakarta. *Techno Nusa Mandiri, IX*.

Fauzi, V. (2016). *Hubungan Antara Pola Tidur dengan Tingkat Konsentrasi pada Mahasiswa Angkatan 2013 Program Studi S1 Keprawatan Fakultas Kesehatan Universitas Muhamadiyah malang.*

Hans, R. (2022). *Sumber Data Sekunder dan Primer? Intip Versi DQLab Yuk*. (DQLab)

Johanes, R., E.Manampiring, A., & Fatimawali. (2020). Hubungan Antara Aktivitas Olahraga dengan Ritme Sirkadiandan dan Stres. *J. Biomedik, 12*(2), 102-109.

Khoeri, I, & Mulyana, DI (2021). Implementasi machine learning dengan decision tree algoritma C4. 5 dalam penerimaan karyawan baru pada PT. Gitareksa dinamika jakarta. *Jurnal Sosial Teknologi*, sostech.greenvest.co.id, https://sostech.greenvest.co.id/index.php/sostech/article/view/126

Nazir. (2005). *Metode Penelitian.*

Nurfitrayani, N, Islamiyah, I, & Masa, APA (2023). Penerapan Klasifikasi Algoritma C4. 5 Dan Algoritma C5. 0 Untuk Mengetahui Tingkat Kepuasan Mahasiswa Terhadap Website Sistem Informasi Terpadu …. *JURNAL MEDIA INFORMATIKA BUDIDARMA*

Purba, D., & Sitorus, C. (2022). Analisis Asosiasi Latar Belakang Mahasiswa Terhadap Capaian Nilai Akademik Menggunakan Algoritma Apriori di Fakultas Ilmu Komputer Universitas Katolik Santo Thomas. *KAKIFIKOM (Kumpulan Artik. Karya Ilm. Fak. Ilmu Komput.,), 04*.

Saunders, M. N., Lewis, P., & Thornhill, A. (2023). *Research Methods for Business Students : 9th Edition. 2023.*

Sugiyono. (2012). *Metode Penelitian Bisnis.*

Solehuddin, M, Syafei, WA, & Gernowo, R (2022). Metode Decision Tree untuk Meningkatkan Kualitas Rencana Pelaksanaan Pembelajaran dengan Algoritma C4. 5. *Jurnal Penelitian Dan Pengembangan …*

Syahputra, F, Hartono, H, & Rosnelly, R (2021). Penerapan Algoritma C4. 5 Dalam Memprediksi Ketersediaan Uang Pada Mesin ATM. *Jurnal Media Informatika Budidarma*

Triawan, B, Lubis, I, & Kadim, LAN (2024). Penerapan Data Mining Untuk Prediksi Penjualan Spanduk Menggunakan Algoritma C4. 5. *Journal of Mathematics …*, journal.binainternusa.org, http://journal.binainternusa.org/index.php/matech/article/view/172

Zer, PFIRH, Hayadi, BH, & ... (2022). Pendekatan Machine Learning Menggunakan Algoritma C4. 5 Berbasis Pso Dalam Analisa Pemahaman Pemrograman Website.

Jurnal Informatika dan ..., journal.eng.unila.ac.id, http://journal.eng.unila.ac.id/index.php/jitet/article/view/2700