

Implementasi Naive Bayes Untuk Klasifikasi Gangguan Tidur

Mutiara Mega Amelia¹, Bintang Maulana Fazrin², Yogi Yosua Panjaitan³, Muhamad Dicky Kurniawan⁴, Nurul Khasanah⁵

Universitas Nusa Mandiri¹²³⁴⁵

12210526@nusamandiri.ac.id¹, 12210451@nusamandiri.ac.id², 12210307@nusamandiri.ac.id³,
12210384@nusamandiri.ac.id⁴, nurul.nuk@nusamandiri.ac.id^{5*}

Diterima	Direvisi	Disetujui
(14-04-2025)	(21-04-2025)	(25-04-2025)

Abstrak - Kualitas tidur yang baik merupakan aspek penting dari gaya hidup sehat, dengan dampak signifikan terhadap kesejahteraan fisik dan mental individu. Tidur yang cukup dan berkualitas telah terbukti meningkatkan fungsi kognitif, memperkuat sistem kekebalan tubuh, dan mengurangi risiko terkena berbagai penyakit kronis. Namun, gangguan tidur seperti insomnia dan sleep apnea dapat mengganggu pola tidur dan berpotensi menyebabkan dampak negatif pada kesehatan seseorang. Dalam konteks ini, penelitian ini bertujuan untuk menyelidiki dan mengklasifikasikan gangguan tidur berdasarkan sejumlah atribut yang berkaitan dengan gaya hidup dan kesehatan tidur. Dataset yang digunakan terdiri dari 374 entri, yang mencakup beragam atribut tentang gaya hidup dan kesehatan tidur. Metode Naive Bayes digunakan untuk melakukan klasifikasi, dengan melakukan empat proporsi pembagian data training dan testing: 60:40, 70:30, 80:20, dan 90:10. Tahapan penelitian meliputi pengumpulan data, preprocessing, pemodelan menggunakan algoritma Naive Bayes, evaluasi model, dan analisis hasil. Hasil penelitian menunjukkan adanya peningkatan akurasi seiring dengan peningkatan proporsi data training, dengan akurasi tertinggi mencapai 92.11% pada pembagian data 90:10. Hasil tersebut termasuk dalam kategori excellent classification, menunjukkan keberhasilan model dalam mengidentifikasi dan mengklasifikasikan gangguan tidur berdasarkan atribut yang diberikan.

Kata Kunci : Gangguan Tidur, Klasifikasi, Naive bayes

Abstract - Good quality sleep is a crucial aspect of a healthy lifestyle, with significant impacts on an individual's physical and mental well-being. Adequate and high-quality sleep has been proven to enhance cognitive function, strengthen the immune system, and reduce the risk of various chronic diseases. However, sleep disorders such as insomnia and sleep apnea can disrupt sleep patterns and potentially cause negative health effects. In this context, this study aims to investigate and classify sleep disorders based on various attributes related to lifestyle and sleep health. The dataset used consists of 374 entries, encompassing diverse attributes regarding lifestyle and sleep health. The Naive Bayes method was used for classification, with four data training and testing proportions: 60:40, 70:30, 80:20, and 90:10. The stages of the research included data collection, preprocessing, modeling using the Naive Bayes algorithm, model evaluation, and result analysis. The results of the study showed an increase in accuracy along with an increase in the proportion of training data, with the highest accuracy reaching 92.11% in the 90:10 data split. These results fall into the category of excellent classification, indicating the success of the model in identifying and classifying sleep disorders based on the given attributes.

Keywords: Sleep Disorders, Classification, Naive Bayes

I. PENDAHULUAN

Tidur adalah proses alami bagi manusia untuk memulihkan tubuh, mengembalikan energi, memperbaiki sel-sel, dan melakukan perbaikan tubuh yang belum sempurna (Aliffia Bingga et al., 2021). Manusia membutuhkan waktu istirahat melalui tidur untuk meregenerasi sel-sel yang rusak dan mengembalikan energi yang terkuras akibat stres, kecemasan, atau ketegangan mental. Kualitas tidur yang baik umumnya berkaitan dengan tingkat stres yang lebih rendah (Dhea et al., 2023). Kebutuhan

hidup yang esensial memaksa manusia harus bekerja untuk memenuhi kebutuhan sehari-hari. Pada umumnya masyarakat bekerja pada siang hari dan istirahat pada malam hari. Kurang tidur adalah bagian penting dari gaya hidup Anda. Kualitas tidur yang baik akan menyeimbangkan metabolisme tubuh, sebaliknya jika kualitas tidur seseorang buruk maka akan mempengaruhi kondisi tubuh. Selain itu, gangguan tidur dapat mengganggu sistem metabolisme tubuh, yang berujung pada penurunan produktivitas dan kesejahteraan umum. Oleh karena itu,

memahami dan mengatasi gangguan tidur menjadi sangat penting untuk menjaga keseimbangan hidup yang sehat dan produktif. Gangguan tidur merupakan salah satu masalah kesehatan yang sering diabaikan, namun memiliki dampak signifikan terhadap kualitas hidup individu. Gangguan tidur adalah kondisi yang dapat menimbulkan masalah pada pola tidur, seperti sering terbangun, kesulitan untuk kembali tidur setelah terbangun, dan ketidakpuasan dengan tidur yang mengakibatkan penurunan kualitas tidur (Aliffia Bingga et al., 2021). Tidur yang berkualitas sangat penting untuk kesehatan fisik dan mental, karena proses ini memungkinkan tubuh dan otak untuk beristirahat, memulihkan energi, serta memperbaiki dan mempertahankan fungsi-fungsi vital. Prevalensi gangguan tidur semakin meningkat seiring dengan gaya hidup yang serba cepat, tekanan pekerjaan, dan penggunaan teknologi yang berlebihan. Faktor-faktor tersebut sering kali menyebabkan stres dan kecemasan yang berkontribusi pada kesulitan tidur. Selain itu, gangguan tidur juga dapat dipicu oleh kondisi medis tertentu, penggunaan obat-obatan, dan pola hidup yang tidak sehat. Dampak dari gangguan tidur sangatlah luas. Kurangnya tidur yang berkualitas dapat mengakibatkan penurunan fungsi kognitif, gangguan konsentrasi, peningkatan risiko kecelakaan, serta masalah kesehatan jangka panjang seperti obesitas, diabetes, penyakit jantung, dan depresi. Oleh karena itu, penting untuk memahami penyebab, jenis, dan cara mengatasi gangguan tidur guna meningkatkan kualitas hidup dan kesehatan masyarakat secara keseluruhan.

Dalam penelitian ini, kami memilih algoritma Naive Bayes untuk klasifikasi gangguan tidur. Algoritma ini dipilih karena kemampuannya yang cepat dan efisien dalam menangani dataset besar serta kemudahannya dalam implementasi, yang sangat penting untuk analisis data yang ekstensif. Naive Bayes dikenal karena kecepatan komputasinya yang tinggi dan efisiensinya dalam memproses data yang besar, menjadikannya pilihan yang ideal untuk penelitian dengan volume data yang signifikan. Namun, algoritma ini memiliki kelemahan inheren, yaitu asumsi independensi antar fitur, yang sering kali tidak realistis dalam konteks dunia nyata di mana fitur-fitur data cenderung saling bergantung. Meskipun demikian, berbagai penelitian telah menunjukkan bahwa Naive Bayes mampu memberikan hasil yang akurat untuk berbagai masalah klasifikasi medis, termasuk gangguan tidur. Keandalan dan ketepatannya dalam berbagai studi sebelumnya memberikan dasar yang kuat bagi kami untuk menggunakannya

dalam penelitian ini, dengan harapan dapat menghasilkan temuan yang signifikan dan dapat diandalkan dalam mengklasifikasikan gangguan tidur.

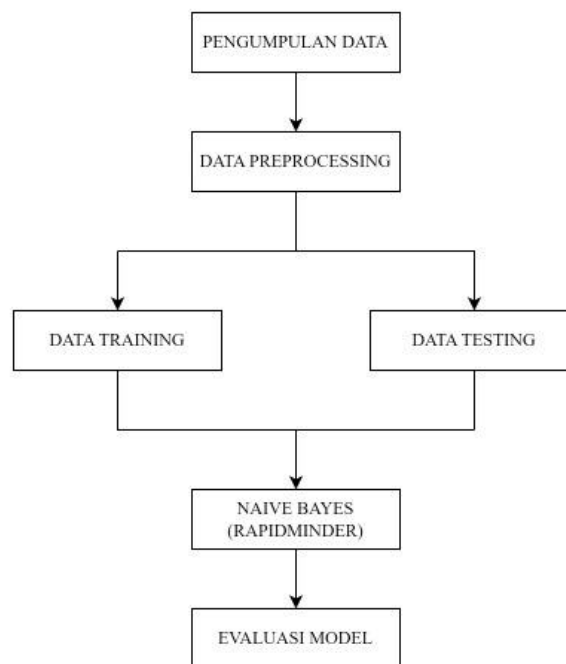
Pada penelitian sebelumnya yang dilakukan oleh Iwan Ady Prabowo, Dwi Remawati, dan Aji Pratama Wisnu Wardana dengan judul *Klasifikasi Tingkat Gangguan Tidur Menggunakan Algoritma Naïve Bayes* ditemukan bahwa metode Naïve Bayes mampu digunakan untuk klasifikasi dan menghasilkan data dengan tingkat akurasi sebesar 80% dalam pengujian sistem implementasi metode tersebut untuk diagnosis jenis gangguan tidur. Ini menunjukkan bahwa algoritma Naïve Bayes efektif dalam menjalankan tugas klasifikasi (Prabowo et al., 2020). Berdasarkan penelitian sebelumnya yang dilakukan oleh Ahmad Azhari dan Inosensia Lionetta Pricillia dengan judul *"Klasifikasi K-Nearest Neighbor pada Penderita Insomnia berdasarkan Sinyal Elektroensefalogram"*, dataset yang diperoleh menunjukkan bahwa penerapan KNN pada pengukuran EEG terhadap penderita insomnia memberikan hasil yang baik. Tingkat akurasi yang dicapai dari data uji KNN adalah 95%. Akurasi ini membuktikan bahwa klasifikasi KNN yang telah dibangun cukup efektif untuk mengidentifikasi penyakit insomnia pada penderitanya (Azhari & Lionetta Pricillia, 2022). Berdasarkan penelitian sebelumnya yang dilakukan oleh Siska Febriani dengan judul penelitian *"Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5"* "Dataset yang diperoleh pada Nilai kalsifikasi akurasi yang didapat pada penerapan algoritma C4.5 menghasilkan nilai lebih dari 60% yaitu akurasi data 72.67% (Febriani & Sulistiani, n.d.). Berdasarkan penelitian sebelumnya yang dilakukan oleh Mohamad Reyhand Faturrahman dan Aliyah Kurniasih dengan judul *"Penggunaan Metode NearMiss, SMOTE, dan Naive Bayes untuk Klasifikasi Gangguan Tidur Berdasarkan Kualitas Tidur dan Gaya Hidup"*, dataset yang digunakan menunjukkan bahwa nilai akurasi model klasifikasi Bernoulli dan Complement Naïve Bayes awalnya sebesar 76% dan 79%. Namun, setelah penerapan undersampling menggunakan NearMiss, akurasi meningkat menjadi 73% dan 81%. Peningkatan ini, bersama dengan metrik evaluasi lainnya, menunjukkan bahwa undersampling menggunakan metode NearMiss memberikan hasil yang lebih baik dalam klasifikasi (Areni et al., 2019). Berdasarkan penelitian yang dilakukan oleh Dhea Fuji Astari, Yulison Herry Chrisnanto, dan Melina dengan judul *"Klasifikasi Tingkat Stres Saat Tidur Menggunakan Algoritma Random Forest"*, hasil penelitian

menunjukkan bahwa algoritma Random Forest menghasilkan akurasi sebesar 93,65% dalam klasifikasi tingkat stres saat tidur. Hal ini menyimpulkan bahwa algoritma Random Forest merupakan pendekatan yang efektif untuk mengklasifikasikan tingkat stres saat tidur (Dhea et al., 2023).

Tujuan dari penelitian ini adalah untuk mengevaluasi kinerja algoritma Naive Bayes dalam mengklasifikasikan gangguan tidur dengan memanfaatkan dataset yang mencakup berbagai atribut gaya hidup dan kesehatan tidur. Penelitian ini berusaha untuk mengatasi permasalahan gangguan tidur yang sering diabaikan namun berdampak signifikan terhadap kualitas hidup individu, termasuk dampak negatif pada produktivitas, kesehatan mental, dan kesejahteraan keseluruhan. Dengan membandingkan kinerja Naive Bayes dengan metode klasifikasi lainnya yang telah dibahas dalam penelitian sebelumnya, seperti K-Nearest Neighbor (KNN), C4.5, dan Random Forest, penelitian ini bertujuan untuk menentukan efektivitas Naive Bayes dalam mengidentifikasi gangguan tidur secara akurat dan efisien. Selain itu, penelitian ini juga mengeksplorasi bagaimana jumlah data pelatihan mempengaruhi akurasi model, mengingat bahwa volume data yang memadai sering kali menjadi faktor kunci dalam meningkatkan performa model pembelajaran mesin. Melalui analisis komprehensif ini, penelitian diharapkan dapat memberikan kontribusi signifikan dalam bidang diagnostik gangguan tidur, membantu tenaga medis dalam memberikan perawatan yang lebih baik kepada pasien, serta menawarkan wawasan baru yang dapat digunakan untuk pengembangan algoritma klasifikasi yang lebih canggih di masa mendatang. Dengan demikian, hasil penelitian ini tidak hanya bermanfaat bagi komunitas medis tetapi juga bagi penelitian ilmiah yang lebih luas dalam upaya meningkatkan kualitas hidup individu yang menderita gangguan tidur.

II. METODE PENELITIAN

Pada penelitian ini terdapat beberapa tahapan yang harus dilalui, seperti yang ditunjukkan pada gambar 1. Penelitian ini menggunakan metode Naive Bayes, di mana data akan diproses melalui serangkaian tahapan yang sistematis untuk memastikan keakuratan hasil. beberapa tahapan tersebut sebagai berikut:



Sumber: Hasil Penelitian (2025)

Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Pengumpulan data adalah langkah yang sangat penting dalam metode ilmiah, yang melibatkan pengumpulan informasi relevan dari tinjauan pustaka dan observasi lapangan. Peneliti mengekstrak informasi esensial dari sumber-sumber tersebut dan mencatatnya pada kartu informasi. Proses ini, yang mencakup baik penelitian berbasis pustaka maupun observasi lapangan langsung, disebut sebagai teknik pengumpulan data (Rumina, n.d.). Pengumpulan data dilakukan untuk mengidentifikasi dan memperoleh informasi yang penting dan akurat mengenai klien (Gultom, n.d.). Data yang kami gunakan dalam penelitian ini berasal dari sumber publik yang tersedia di Kaggle, sebuah platform terkenal yang menyediakan berbagai dataset untuk keperluan analisis dan pengembangan model. Dengan menggunakan dataset yang terdiri dari 374 data, kami berupaya untuk menguji kemampuan model dalam mengklasifikasikan dan memprediksi dengan akurasi tinggi. Pemilihan dataset ini didasarkan pada ketersediaannya yang luas dan relevansi terhadap objek penelitian, sehingga memungkinkan kami untuk memperoleh hasil yang valid dan dapat diandalkan.

2. Data Preprocessing

Data Preprocessing adalah langkah penting dalam proses pengolahan data sebelum menerapkan algoritma pengklasifikasian seperti Naive Bayes. Data Preprocessing mencakup

serangkaian teknik dan metode yang digunakan untuk membersihkan, mengubah, dan menyiapkan data mentah menjadi format yang sesuai untuk analisis dan pemodelan. Untuk data yang digunakan dalam penelitian ini hanya melalui proses pemilihan label dan penentuan tipe data.

Sebelum data diproses menggunakan algoritma Naive Bayes, dilakukan pembagian data terlebih dahulu menjadi data training dan data testing. Pembagian data ini bertujuan untuk memastikan bahwa model yang dibangun dapat diuji dan divalidasi dengan data yang tidak terlihat sebelumnya, sehingga hasil evaluasi model dapat lebih dipercaya dan mencerminkan kinerja model yang sebenarnya pada data baru.

Pada penelitian ini, dilakukan pembagian data melalui empat proses, yaitu:

- a. 60:40: Pada perbandingan ini, data dibagi dengan 60% data untuk pelatihan dan 40% data untuk pengujian.
- b. 70:30: Pada perbandingan ini, data dibagi dengan 70% data untuk pelatihan dan 30% data untuk pengujian.
- c. 80:20: Pada perbandingan ini, data dibagi dengan 80% data untuk pelatihan dan 20% data untuk pengujian.
- d. 90:10: Pada perbandingan ini, data dibagi dengan 90% data untuk pelatihan dan 10% data untuk pengujian.

Setelah data dibagi sesuai dengan proporsi yang telah ditentukan, data training digunakan untuk melatih model Naive Bayes. Model ini kemudian diuji menggunakan data testing untuk mengevaluasi kinerjanya.

3. Pembangunan Model dengan Naive Bayes

Secara umum algoritma Naive Bayes untuk membangun sebuah hasil kesimpulan adalah sebagai berikut Naive Bayes Classifier merupakan salah satu algoritma dalam teknik data mining yang menerapkan teori bayes dalam klasifikasi. Teorema keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (Pattern Recognition), Naive bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Naive bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidak tergantungan yang kuat. Selain itu Naive Bayes juga dapat menganalisis variabel-variabel yang paling mempengaruhinya dalam bentuk peluang (Retnosari et al., 2021).

4. Evaluasi Model

Evaluasi Model adalah proses menikahi kinerja model klasifikasi yang telah dibangun

untuk menentukan seberapa baik model tersebut dalam mengklasifikasi. Tahap evaluasi bertujuan untuk mengukur seberapa efektif model yang telah dikembangkan pada tahap sebelumnya (Giovani et al., 2020). Proses evaluasi ini melibatkan penggunaan berbagai metrik dan teknik untuk mengukur akurasi, presisi, recall, dan kemampuan model dalam menangani data yang tidak terlihat selama pelatihan.

Model evaluasi merupakan desain atau rancangan evaluasi yang dikembangkan ahli evaluasi ataupun evaluator dalam melaksanakan evaluasi suatu program. Dalam ilmu evaluasi program pendidikan ada banyak model yang bisa digunakan untuk mengevaluasi suatu program. Meskipun antarasatudengan lainnya berbeda, namun maksudnya sama yaitu melakukan kegiatanpengumpulan data atau informasi yang berkenaan dengan objek yang dievaluasi, yang tujuannya untuk menyediakan bahan bagi pengambil keputusan dalam menentukan tindak lanjut suatu program.

III. HASIL DAN PEMBAHASAN

Pada bagian ini, akan dibahas hasil-hasil yang diperoleh dari penelitian yang telah dilakukan, serta analisis terhadap hasil tersebut. Berikut ini adalah hasil dan pembahasan dari setiap tahap penelitian:

1. Pengumpulan Data

Dalam penelitian ini dataset yang digunakan "Sleep Health and Lifestyle" didapatkan dari Kaggle ([Data Yang Digunakan](#)) yang berisi 374 entri yang mencakup berbagai atribut terkait kesehatan tidur dan gaya hidup individu. Atribut-atribut tersebut meliputi *person ID* (ID Orang), *gender* (Jenis Kelamin), *age* (Usia), *occupation* (Pekerjaan), *sleep duration* (Durasi Tidur), *quality of sleep* (Kualitas Tidur), *physical activity level* (Tingkat Aktivitas Fisik), *stress level* (Tingkat Stres), *BMI category* (Kategori BMI), *blood pressure* (Tekanan Darah), *heart rate* (Detak Jantung), *daily steps* (Langkah Harian), dan *sleep disorder* (Gangguan Tidur). "Sleep Disorder" merupakan label yang diklasifikasikan menjadi tiga kategori: none (individu tidak menunjukkan gangguan tidur), insomnia (individu mengalami kesulitan tidur atau tidur yang tidak berkualitas), dan *sleep apnea* (individu mengalami jeda dalam bernapas saat tidur).

2. Data Preprocessing

Dalam tahap data preprocessing untuk dataset "Sleep Health and Lifestyle", kami memastikan bahwa dataset yang digunakan sudah bersih dan tidak mengandung data yang

hilang (missing data), sehingga tidak diperlukan imputasi atau penghapusan data (Zulfiqri et al., 2024). Langkah pertama dalam preprocessing adalah pemilihan label, di mana atribut Sleep Disorder ditetapkan sebagai target klasifikasi, mengkategorikan gangguan tidur individu menjadi none, insomnia, dan sleep apnea. Selanjutnya, kami memastikan tipe data setiap atribut sesuai dengan kebutuhan analisis. Proses ini memastikan bahwa dataset siap untuk analisis lebih lanjut dan pengembangan model machine learning.

3. Pembangunan Model dengan Naïve Bayes

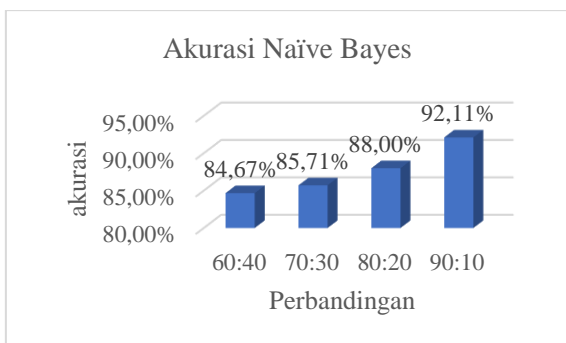
Dalam tahap pemodelan, kami menggunakan algoritma Naïve Bayes untuk mengklasifikasikan gangguan tidur berdasarkan atribut-atribut dalam dataset "Sleep Health and Lifestyle". Untuk mengevaluasi kinerja model, kami membandingkan hasil dengan empat proporsi pembagian data untuk training dan testing: 60:40, 70:30, 80:20, dan 90:10. Dalam setiap pembagian, proporsi tertentu dari data digunakan untuk melatih model (data training) dan sisanya untuk menguji kinerja model (data testing). Berikut hasil akurasi yang didapat menggunakan split data:

Tabel 1. Hasil Akurasi Naïve Bayes

Naïve bayes	
Perbandingan	Akurasi Naïve Bayes
60:40	84,67%
70:30	85,71%
80:20	88,00%
90:10	92,11%

Sumber: Hasil Penelitian (2025)

Pada Tabel 1. dapat dilihat bahwa hasil akurasi dari algoritma Naive Bayes menunjukkan peningkatan yang signifikan seiring dengan peningkatan proporsi data latih terhadap data uji. Dengan perbandingan 60:40, akurasi yang dicapai adalah 84,67%. dan pada perbandingan 90:10, akurasi tertinggi dicapai dengan nilai 92,11%. Grafik akurasi dapat dilihat pada Gambar 2.



Sumber: Hasil Penelitian (2025)

Gambar 2. Grafik Perbandingan Akurasi Naïve Bayes

Hasil tersebut menunjukkan akurasi algoritma Naïve Bayes berdasarkan perbandingan pembagian data yang berbeda. Berikut penjelasannya:

- 60:40: Pada perbandingan ini, data dibagi dengan 60% data untuk pelatihan dan 40% data untuk pengujian. Akurasi yang dicapai adalah 84.67%.
- 70:30: Pada perbandingan ini, data dibagi dengan 70% data untuk pelatihan dan 30% data untuk pengujian. Akurasi yang dicapai adalah 85.71%.
- 80:20: Pada perbandingan ini, data dibagi dengan 80% data untuk pelatihan dan 20% data untuk pengujian. Akurasi yang dicapai adalah 88.00%.
- 90:10: Pada perbandingan ini, data dibagi dengan 90% data untuk pelatihan dan 10% data untuk pengujian. Akurasi yang dicapai adalah 92.11%.
- kita bisa melihat bahwa semakin besar proporsi data yang digunakan untuk pelatihan (training), semakin tinggi akurasi yang dicapai oleh algoritma Naïve Bayes. Pembagian 90:10 memberikan akurasi tertinggi, yaitu 92.11%, menunjukkan bahwa model mendapatkan kinerja yang lebih baik dengan lebih banyak data pelatihan. Sebaliknya, pembagian 60:40 memberikan akurasi terendah, yaitu 84.67%.

Dari hasil tersebut dapat disimpulkan bahwa ada korelasi positif antara proporsi data pelatihan dan akurasi model Naïve Bayes. Semakin besar proporsi data pelatihan, semakin tinggi akurasi yang dicapai. Namun, penting untuk dicatat bahwa peningkatan akurasi ini juga harus dipertimbangkan dengan potensi overfitting, di mana model mungkin sangat baik dalam memprediksi data pelatihan tetapi tidak *generalisasi* dengan baik pada data yang benar-benar baru. Pembagian data yang lebih besar untuk pelatihan (misalnya, 90:10) memberikan lebih banyak informasi kepada model, meningkatkan kemampuannya untuk mengenali pola yang lebih *kompleks* dalam data. Namun, sangat penting untuk memastikan bahwa data uji yang cukup tetap tersedia untuk benar-benar mengevaluasi kinerja model secara *objektif*.

4. Evaluasi Model

Berikut adalah perhitungan evaluasi hasil yang di dapat dari data yang telah diolah menggunakan RapidMiner:

- Confusion matrix*

Berikut adalah hasil confusion matrix yang dihasilkan beserta penjelasan lebih detail untuk setiap baris dalam confusion matrix

	true None	true Sleep Apnea	true Insomnia	class precision
pred None	22	0	3	88.00%
pred Sleep Apnea	0	8	0	100.00%
pred Insomnia	0	0	5	100.00%
class recall	100.00%	100.00%	62.50%	

Sumber: Hasil Penelitian (2025)

Gambar 3. Confusion Matrix

Pada Gambar 3. Dapat dilihat bahwa perhitungan yang dapat dijelaskan sebagai berikut:

Prediksi None (Baris Pertama):

- 1) 22 benar-positif (*true positive*) untuk kategori *None*, artinya dari semua *individu* yang sebenarnya tidak memiliki gangguan tidur (*None*), model berhasil mengidentifikasi 22 *individu* dengan benar.
- 2) 0 false-positif (*false positive*) untuk kategori *Sleep Apnea*, artinya Tidak ada *individu* yang sebenarnya menderita *sleep apnea* namun salah diprediksi sebagai tidak memiliki gangguan tidur (*None*) oleh model.
- 3) 3 false-positif (*false positive*) untuk kategori *Insomnia*, artinya ada 3 *individu* yang sebenarnya menderita *insomnia*, tetapi salah diprediksi oleh model sebagai tidak memiliki gangguan tidur (*None*).

Prediksi *Sleep Apnea* (Baris Kedua):

- 1) 0 false-negatif (*false negative*) untuk kategori *None*, artinya tidak ada *individu* yang sebenarnya tidak memiliki gangguan tidur (*None*), tetapi salah diprediksi sebagai *sleep apnea* oleh model.
- 2) 8 benar-positif (*true positive*) untuk kategori *Sleep Apnea*, artinya model berhasil mengidentifikasi 8 *individu* dengan *sleep apnea* dengan benar.
- 3) 0 false-positif (*false positive*) untuk kategori *Insomnia*, artinya tidak ada *individu* yang sebenarnya menderita *insomnia* namun salah diprediksi sebagai *sleep apnea* oleh model.

Prediksi *Insomnia* (Baris Ketiga):

- 1) 0 false-negatif (*false negative*) untuk kategori *None*, artinya tidak ada *individu* yang sebenarnya tidak memiliki gangguan tidur (*None*), tetapi salah diprediksi sebagai *insomnia* oleh model.
- 2) 0 false-negatif (*false negative*) untuk kategori *Sleep Apnea*, artinya tidak ada *individu* yang sebenarnya menderita *sleep apnea*, tetapi salah diprediksi sebagai *insomnia* oleh model.
- 3) 5 benar-positif (*true positive*) untuk kategori *Insomnia*, artinya model berhasil mengidentifikasi 5 *individu* dengan *insomnia* dengan benar.

b. Akurasi

Akurasi adalah seberapa mendekati suatu hasil pengukuran dengan nilai atau data sebenarnya (Setiawan & Afdhalludin, n.d.). Akurasi dapat dihitung sebagai:

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Jumlah Prediksi}} = \frac{22+8+5}{22+3+8+0+5} = \frac{35}{38} \approx 92.11\% \quad (1)$$

c. Recall

Recall didefinisikan sebagai rasio antara jumlah item relevan yang berhasil dipilih terhadap total jumlah item relevan yang ada (Rinestu et al., 2022).

$$\text{Recall None} = \frac{\text{True Positive None}}{\text{True Positive None} + \text{False Negative None}} = \frac{22}{22+0} = \frac{22}{22} = 100\% \quad (2)$$

$$\text{Recall Sleep Apnea} = \frac{\text{True Positive Sleep Apnea}}{\text{True Positive Sleep Apnea} + \text{False Negative Sleep Apnea}}$$

$$\text{Recall Sleep Apnea} = \frac{8}{8+0} = \frac{8}{8} = 100\% \quad (3)$$

$$\text{Recall Insomnia} = \frac{\text{True Positive Insomnia}}{\text{True Positive Insomnia} + \text{False Negative Insomnia}}$$

$$\text{Recall Insomnia} = \frac{5}{5+3} = \frac{5}{8} \approx 62.50\% \quad (4)$$

d. Precision

Precision adalah sebuah ukuran yang menilai proporsi dokumen yang ditemukan oleh suatu proses pencarian dan dianggap relevan untuk kebutuhan informasi. Ini adalah rasio antara jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen yang ditemukan (Laila Safitry - et al., n.d.).

$$\text{Precision None} = \frac{\text{True Positive None}}{\text{True Positive None} + \text{False Positive None}} = \frac{22}{22+3} = \frac{22}{25} = 88\% \quad (5)$$

$$\text{Precision Sleep Apnea} = \frac{\text{True Positive Sleep Apnea}}{\text{True Positive Sleep Apnea} + \text{False Positive Sleep Apnea}}$$

Precision Sleep Apnea =

$$\frac{8}{8+0} = \frac{8}{8} = 100\%$$

(6)

Precision Insomnia =

True Positive Insomnia

True Positive Insomnia + False Positive Insomnia

$$\text{Precision Insomnia} = \frac{5}{5+0} = \frac{5}{5} \approx 100\%$$

(7)

Model Naïve Bayes memiliki performa yang sangat baik dalam mengklasifikasikan kategori *None* dan *Sleep Apnea* dengan *recall* dan *precision* 100%. Namun, untuk kategori *Insomnia*, model menunjukkan performa yang lebih rendah dengan *recall* sebesar 62.50%, meskipun *precision* tetap 100%. Hal ini mengindikasikan bahwa model kadang-kadang gagal mengidentifikasi semua kasus *Insomnia* dengan benar, meskipun prediksi yang dilakukan untuk *Insomnia* selalu benar. Model ini secara keseluruhan memiliki akurasi tinggi sebesar 92.11%.

IV. KESIMPULAN

Penelitian ini mengkaji klasifikasi gangguan tidur dengan menggunakan algoritma Naive Bayes, memanfaatkan berbagai atribut yang terkait dengan gaya hidup dan kesehatan tidur. Evaluasi model dilakukan dengan membagi data menjadi empat proporsi berbeda untuk pelatihan dan pengujian (60:40, 70:30, 80:20, dan 90:10). Hasil penelitian menunjukkan bahwa akurasi model meningkat seiring dengan bertambahnya data yang digunakan untuk pelatihan, dengan akurasi tertinggi sebesar 92.11% pada pembagian 90:10. Model Naive Bayes berhasil mengklasifikasikan kategori "None" dan "Sleep Apnea" dengan *recall* dan *precision* mencapai 100%. Namun, untuk kategori "Insomnia," model menunjukkan *recall* yang lebih rendah sebesar 62.50%, meskipun *precision* tetap 100%. Ini menunjukkan bahwa meskipun semua prediksi *Insomnia* yang dibuat oleh model benar, model tidak selalu dapat mendeteksi semua kasus *Insomnia*. Secara keseluruhan, model Naive Bayes menunjukkan performa yang kuat dengan akurasi total sebesar 92.11%, membuktikan kemampuannya dalam mengklasifikasikan gangguan tidur berdasarkan atribut yang dianalisis.

V. REFERENSI

Aliffia Bingga, I., Bingga, I. A., & Dokter, P. (2021). *KAITAN KUALITAS TIDUR DENGAN DIABETES MELITUS TIPE 2*.

- Areni, I. S., Amirullah, I., & Arifin, N. (2019). Klasifikasi Kematangan Stroberi Berbasis Segmentasi Warna dengan Metode HSV. *Jurnal Penelitian Enjiniring*, 23(2), 113–116.
<https://doi.org/10.25042/jpe.112019.03>
- Azhari, A., & Lionetta Pricillia, I. (2022). *Klasifikasi K-Nearest Neighbor pada Penderita Insomnia berdasarkan Sinyal Elektroensefalogram* (Vol. 4, Issue 2).
- Dhea, F., Astari, Y. H., Chrisanto, M. T., Informatika, J., Achmad, Y., Cimahi, J., Terusan, J., Sudirman, K., Cimahi, S., Cimahi, K., & Barat, J. (2023). *KLASIFIKASI TINGKAT STRES SAAT TIDUR MENGGUNAKAN ALGORITMA RANDOM FOREST*. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 5).
- Febriani, S., & Sulistiani, H. (n.d.). *ANALISIS DATA HASIL DIAGNOSA UNTUK KLASIFIKASI GANGGUAN KEPERIBADIAN MENGGUNAKAN ALGORITMA C4.5*.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). *ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI*. *Jurnal Teknoinfo*, 14(2), 115.
<https://doi.org/10.33365/jti.v14i2.679>
- Gultom, H. (n.d.). *DATA-DATA YANG DIGUNAKAN DALAM PROSES ASUHAN KEPERAWATAN DAN METODE PENGUMPULAN DATA*.
- Laila Safitry -, D., Al Harani -, A., Divyaning -, E., Hadya Hanifa -, F., Fauzia Chema -, D., & Paryudi -, I. (n.d.). *Perbandingan Metode Decision Tree, Naive Bayes, dan Induction Rule untuk Klasifikasi Penyakit Diabetes*. In *Journal of Informatics and Advanced Computing* (Vol. 4, Issue 1).
- Prabowo, I. A., Remawati, D., & Wardana, A. P. W. (2020). *Klasifikasi Tingkat Gangguan Tidur Menggunakan Algoritma Naive Bayes*. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 8(2).
<https://doi.org/10.30646/tikomsin.v8i2.519>
- Retnosari, R., Studi, P., Informasi, S., Nusa, S., & Jakarta, M. (2021). *ANALISIS KELAYAKAN KREDIT USAHA MIKRO BERJALAN PADA PERBANKAN DENGAN METODE NAIVE BAYES*.
- Rinestu, M., Made Indra, I. P., Marsanto, B., & Trisakti, S. (2022). *Classification Of Investment Decisions During Covid-19 Pandemic Using Naive Bayes Klasifikasi Keputusan Investasi Di Masa Pandemi Covid-19 Dengan Menggunakan Naive Bayes*. In *Management Studies and*

Entrepreneurship Journal (Vol. 3, Issue 4).
<http://journal.yrpiiku.com/index.php/msej>
Rumina. (n.d.). *TEHNIK PENGUMPULAN DATA
DALAM PENELITIAN PENDIDIKAN*.
Setiawan, A. E., & Afdhalludin, A. (n.d.). *Analisa
Metode Fuzzy Mamdani Dan Sugeno
Untuk Deteksi Daerah Rentan Banjir*.
Zulfiqri, R., Sari, B. N., & Padilah, T. N. (2024).
ANALISIS SENTIMEN ULASAN

PENGGUNA APLIKASI MEDIA SOSIAL
INSTAGRAM PADA SITUS GOOGLE
PLAY STORE MENGGUNAKAN NAÏVE
BAYES CLASSIFIER. *Jurnal Informatika
Dan Teknik Elektro Terapan*, 12(3).
<https://doi.org/10.23960/jitet.v12i3.4995>