

Klasifikasi Gangguan Tidur pada Individu Menggunakan Algoritma Naive Bayes Berbasis Data Gejala Klinis

Application of the Naive Bayes Algorithm for Sleep Disorder Classification Using Clinical Data

Yan Yang Thanri^{a,1}, Bob Subhan Riza^{b,2}, Juli Iriani^{c,3}, Adithya Alfahrezi Noor^{d,4}

^{a,b,c,d} Universitas Potensi utama, Medan, Indonesia

Email : ythanri@gmail.com¹

ABSTRAK

Gangguan tidur merupakan kondisi yang dapat memengaruhi kesehatan fisik dan mental seseorang serta menurunkan kualitas hidup secara keseluruhan. Deteksi dini terhadap gangguan tidur sangat penting agar intervensi medis dapat dilakukan lebih cepat dan tepat. Penelitian ini bertujuan untuk mengidentifikasi gangguan tidur pada seseorang dengan menggunakan algoritma Naïve Bayes, yaitu salah satu metode klasifikasi dalam machine learning yang berbasis probabilistik. Data yang digunakan dalam penelitian ini meliputi atribut usia, durasi tidur, kualitas tidur, tingkat stres, kategori BMI, tekanan darah, dan detak jantung, dengan total 374 data pasien. Penelitian dilakukan melalui tahapan pra-pemrosesan data, pelatihan model dengan data latih, pengujian model dengan data uji, serta evaluasi menggunakan confusion matrix dan akurasi. Hasil pengujian menunjukkan bahwa model Naïve Bayes menghasilkan akurasi sebesar 90,54%. Meskipun akurasi tergolong tinggi, model belum mampu mengklasifikasikan kelas gangguan tidur Insomnia dengan baik akibat distribusi data yang tidak seimbang. Oleh karena itu, dalam penelitian selanjutnya disarankan penerapan teknik penyeimbangan data serta pengujian dengan algoritma pembandingan lainnya. Penelitian ini diharapkan dapat menjadi landasan awal pengembangan sistem cerdas dalam mendeteksi gangguan tidur secara otomatis dan cepat.

Kata Kunci : Gangguan Tidur, Naïve Bayes, Klasifikasi, Data Mining, Deteksi Dini, Machine Learning

ABSTRACT

Sleep disorders are conditions that can significantly affect a person's physical and mental health, leading to a decrease in overall quality of life. Early detection of sleep disorders is crucial to ensure timely and accurate medical intervention. This study aims to identify sleep disorders using the Naïve Bayes algorithm, a probabilistic-based classification method in machine learning. The dataset used in this research consists of 374 patient records, including attributes such as age, sleep duration, sleep quality, stress level, BMI category, blood pressure, and heart rate. The research was conducted through stages of data preprocessing, model training using training data, model testing with test data, and performance evaluation using a confusion matrix and accuracy metric. The experimental results showed that the Naïve Bayes model achieved an accuracy of 90.54%. Although the accuracy is relatively high, the model failed to correctly classify cases of Insomnia due to imbalanced data distribution. Therefore, future research is recommended to apply data balancing techniques and compare the performance with other classification algorithms. This study is expected to serve as a foundation for the development of intelligent systems to detect sleep disorders automatically and efficiently.

Keywords : Sleep Disorder, Naïve Bayes, Classification, Data Mining, Early Detection, Machine Learning

Disubmit:

Info Artikel :
Direview:

Diterima :

1. PENDAHULUAN

Tidur merupakan bagian penting dalam kehidupan manusia untuk menjaga keseimbangan fisik, emosional, dan metabolisme tubuh. Pola hidup modern sering kali memaksa seseorang bekerja melebihi batas waktu normal, sehingga waktu istirahat, khususnya tidur, menjadi terganggu. Kurang tidur telah dikaitkan dengan peningkatan risiko berbagai gangguan kesehatan, termasuk gangguan metabolik, gangguan psikologis, serta penurunan kualitas hidup secara keseluruhan [1]. Dalam konteks pelayanan kesehatan, identifikasi gangguan tidur secara dini sangat penting agar intervensi dapat dilakukan sebelum terjadi dampak lebih serius pada pasien. Kemajuan teknologi informasi saat ini membuka peluang bagi pengembangan sistem deteksi gangguan tidur berbasis kecerdasan buatan untuk membantu tenaga kesehatan seperti dokter, perawat, atau bidan dalam pengambilan keputusan klinis [2].

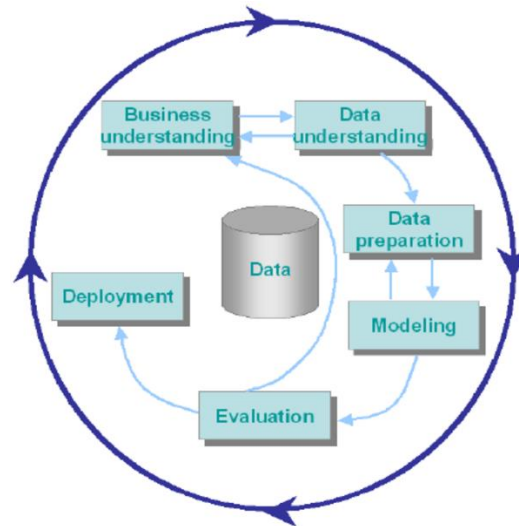
Salah satu algoritma pembelajaran mesin yang banyak digunakan untuk tugas klasifikasi adalah Naïve Bayes. Algoritma ini menggunakan pendekatan probabilitas berdasarkan Teorema Bayes dengan asumsi independensi antar fitur. Naïve Bayes dikenal sederhana, cepat, serta mampu memberikan performa yang baik meskipun data pelatihan terbatas [3]. Beberapa studi internasional menunjukkan efektivitas Naïve Bayes dalam konteks klasifikasi gangguan tidur. Ghifari dan Widasari, misalnya, menerapkan algoritma ini untuk mendeteksi *Sleep-Disordered Breathing* menggunakan sinyal ECG dan memperoleh akurasi sebesar 85% [4]. Penelitian oleh Prabowo et al. mengklasifikasikan enam jenis gangguan tidur seperti insomnia, narkolepsi, dan mimpi buruk, dengan akurasi mencapai 80% [5]. Sementara itu, Kurniawan dan Akbar mengembangkan sistem pakar berbasis Naïve Bayes untuk diagnosis gangguan tidur pada anak dan mencatat akurasi sebesar 95% [6].

Penelitian Setiawati et al. membandingkan kinerja beberapa algoritma klasifikasi seperti SVM, KNN, CNN, dan LSTM dalam memprediksi gangguan tidur. Hasilnya, Logistic Regression memiliki performa terbaik dengan nilai F1-score 84,75%, namun Naïve Bayes tetap menunjukkan hasil kompetitif [7]. Studi internasional lainnya oleh Smith et al. menggunakan Naïve Bayes dalam klasifikasi tahapan tidur dan mencatat akurasi sebesar 74,4% [8]. Chen et al. dalam jurnal *Information* menunjukkan bahwa penggunaan kombinasi Naïve Bayes dan teknik oversampling (SMOTE) untuk prediksi *Obstructive Sleep Apnea* (OSA) menghasilkan akurasi 67,7% [9]. Selain itu, Sathish et al. mengintegrasikan CNN dan teknik klasifikasi tradisional untuk memprediksi kualitas tidur berdasarkan data wearable dan aktigrafi, memperlihatkan bahwa pendekatan hibrid sangat menjanjikan [10].

Berdasarkan hasil penelitian terdahulu dan urgensi permasalahan di lapangan, penelitian ini bertujuan untuk mengembangkan model klasifikasi gangguan tidur menggunakan algoritma Naïve Bayes yang berbasis data gejala klinis. Model ini diharapkan dapat mendukung proses identifikasi dini oleh tenaga medis dalam mendeteksi gangguan tidur secara cepat, akurat, dan efisien.

2. METODE

Penelitian ini menggunakan pendekatan CRISP-DM (Cross-Industry Standard Process for Data Mining) sebagai kerangka kerja metodologi. CRISP-DM merupakan pendekatan terstruktur yang terdiri dari enam tahapan utama, yaitu: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

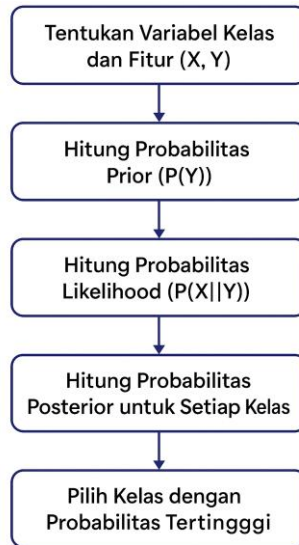
Tahapan pertama adalah *business understanding*, yang berfokus pada pemahaman tujuan utama penelitian. Dalam konteks ini, tujuan utamanya adalah membantu tenaga medis, seperti bidan dan perawat, dalam mengidentifikasi gangguan tidur secara dini agar dapat memberikan intervensi yang tepat. Setelah itu dilanjutkan dengan *data understanding*, yaitu tahap untuk memahami karakteristik data yang digunakan. Pada tahap ini, dilakukan eksplorasi terhadap data gejala klinis, riwayat kesehatan, serta data demografis seperti usia, jenis kelamin, dan pola tidur pasien.

Selanjutnya adalah tahap *data preparation*, yaitu proses pembersihan dan transformasi data agar siap digunakan dalam proses pemodelan. Proses ini meliputi penghapusan data yang tidak konsisten, penanganan data hilang, dan pengkodean variabel kategorikal menjadi bentuk numerik. Setelah data siap, dilakukan *modeling*, yaitu membangun model klasifikasi menggunakan algoritma Naïve Bayes. Pada tahap ini, data yang telah dipersiapkan dibagi menjadi data latih dan data uji untuk menguji kemampuan model dalam memprediksi gangguan tidur berdasarkan pola yang telah dipelajari.

Setelah model dibentuk, dilakukan *evaluation* untuk mengukur kinerja model menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Evaluasi ini penting untuk memastikan bahwa model memiliki performa yang baik dan dapat diandalkan dalam proses klasifikasi. Tahapan terakhir adalah *deployment*, yaitu implementasi model ke dalam sistem nyata. Dalam konteks ini, model dapat diintegrasikan ke dalam aplikasi berbasis web atau sistem pakar yang dapat digunakan oleh tenaga kesehatan untuk melakukan prediksi gangguan tidur secara langsung kepada pasien. Dengan pendekatan CRISP-DM, proses pengembangan sistem klasifikasi menjadi lebih terstruktur, sistematis, dan dapat dievaluasi di setiap tahapannya.

Tahapan Algoritma Naïve Bayes

Untuk mengklasifikasikan data berdasarkan kemungkinan suatu kejadian, algoritma Naïve Bayes digunakan karena kesederhanaannya dan efektivitasnya dalam menangani data berskala besar [11]. Berikut ini merupakan tahapan proses algoritma Naïve Bayes [12] yang digunakan dalam penelitian ini untuk mengidentifikasi gangguan tidur pada gambar 2[13-15].



Gambar 2. Bagan alur proses klasifikasi menggunakan algoritma Naïve Bayes untuk identifikasi gangguan tidur

Gambar 2 menggambarkan alur proses dalam penerapan algoritma Naïve Bayes untuk mengklasifikasikan data, dalam hal ini untuk mendeteksi gangguan tidur pada seseorang. Proses dimulai dari pengumpulan data, yang terdiri dari atribut-atribut seperti usia, durasi tidur, BMI, kebiasaan merokok, dan lain-lain. Selanjutnya, data yang dikumpulkan masuk ke tahap pra-pemrosesan, di mana dilakukan pembersihan data, penanganan nilai kosong, dan encoding variabel kategorikal. Setelah data siap, dilakukan perhitungan probabilitas prior berdasarkan jumlah kemunculan masing-masing kelas (misalnya: "Gangguan Tidur" dan "Normal"). Lalu dilanjutkan dengan perhitungan probabilitas likelihood untuk setiap fitur terhadap masing-masing kelas, baik secara frekuensi untuk data kategorikal maupun dengan distribusi Gaussian untuk data numerik. Probabilitas tersebut kemudian digunakan dalam perhitungan probabilitas posterior menggunakan Teorema Bayes. Hasil perhitungan probabilitas posterior akan menentukan kelas dengan probabilitas tertinggi, yang menjadi hasil prediksi dari sistem.

1. Tentukan Variabel Kelas dan Fitur (X, Y):
Misalnya:
 - a. Fitur (X): Usia, Durasi Tidur, BMI, Pekerjaan, Merokok
 - b. Kelas (Y): Gangguan Tidur = {Ya, Tidak}
2. Hitung Probabilitas Prior (P(Y)):
Hitung probabilitas setiap kelas berdasarkan frekuensi kemunculannya dalam data:

$$P(Y = \text{Ya}) = \frac{\text{Jumlah data dengan gangguan tidur}}{\text{Total data}} \quad (1)$$

$$P(Y = \text{Tidak}) = \frac{\text{Jumlah data tanpa gangguan tidur}}{\text{Total data}} \quad (2)$$

3. Hitung Probabilitas Likelihood (P(X|Y)):
Untuk setiap fitur dan setiap kelas, hitung probabilitas kemunculan fitur tersebut. Jika data kategorikal (misalnya pekerjaan = malam), gunakan:

$$P(\text{Pekerjaan} = \text{Malam} | Y = \text{Ya}) = \frac{\text{Jumlah orang bekerja malam dan mengalami gangguan tidur}}{\text{Jumlah orang dengan gangguan tidur}} \quad (3)$$

Jika data numerik, gunakan fungsi distribusi Gaussian:

$$P(x_i|Y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \quad (4)$$

4. Hitung Probabilitas Posterior untuk Setiap Kelas:
Gunakan rumus Teorema Bayes:

$$P(Y|X) \propto P(Y) \cdot \prod_{i=1}^n P(x_i|Y) \quad (5)$$

5. Pilih Kelas dengan Probabilitas Tertinggi:
Bandingkan nilai posterior untuk setiap kelas:
a. Jika $P(Y=Y_a|X) > P(Y=Tidak|X)$, maka prediksi = "Gangguan Tidur"
b. Jika sebaliknya, maka prediksi = "Normal"

Dataset

Untuk mendukung proses klasifikasi gangguan tidur, digunakan sebuah dataset yang berisi informasi kesehatan dan kebiasaan tidur dari beberapa individu. Dataset ini terdiri dari berbagai variabel yang merepresentasikan faktor-faktor yang diduga berpengaruh terhadap munculnya gangguan tidur, seperti usia, durasi tidur, tingkat stres, indeks massa tubuh (BMI), dan aktivitas fisik harian. Informasi ini akan digunakan sebagai input untuk membangun model klasifikasi menggunakan algoritma Naïve Bayes. Penjelasan detail mengenai isi dataset dan deskripsi setiap variabel dapat dilihat pada uraian dan tabel 1.

Tabel 1. Deskripsi Variabel dalam Dataset Gangguan Tidur

No.	Nama Variabel	Tipe Data	Deskripsi
1	Person ID	Integer	Identitas unik masing-masing individu dalam dataset.
2	Gender	Kategorikal	Jenis kelamin responden. Dalam dataset ini seluruhnya adalah "Male".
3	Age	Numerik	Usia individu dalam tahun.
4	Sleep Duration	Time/String	Lama tidur individu (format jam:menit).
5	Quality of Sleep	Numerik	Skor kualitas tidur (semakin tinggi, semakin baik).
6	Stress Level	Numerik	Tingkat stres individu dalam skala numerik.
7	BMI Category	Kategorikal	Kategori Indeks Massa Tubuh: Normal, Overweight, Obese.
8	Blood Pressure	String	Tekanan darah dalam format sistolik/diastolik (misal 125/80).
9	Daily Steps	Numerik	Jumlah langkah rata-rata yang diambil per hari.
10	Sleep Disorder	Kategorikal	Label target: jenis gangguan tidur (None, Sleep Apnea, Insomnia).

Tabel 1 memberikan gambaran mengenai variabel-variabel yang digunakan dalam dataset klasifikasi gangguan tidur. Terdapat sepuluh variabel utama, yang terdiri dari variabel identitas (Person ID), karakteristik individu seperti Gender dan Age, serta beberapa indikator kebiasaan dan kondisi kesehatan seperti Sleep Duration, Quality of Sleep, Stress Level, BMI Category, Blood Pressure, dan Daily Steps. Variabel Sleep Disorder merupakan label target yang diklasifikasikan ke dalam tiga kategori: *None*, *Sleep Apnea*, dan *Insomnia*. Sebagian besar variabel dalam dataset bersifat numerik dan kategorikal, sehingga sesuai untuk digunakan dalam model klasifikasi berbasis probabilistik seperti algoritma Naïve Bayes. Struktur dan tipe data yang seragam memudahkan proses preprocessing dan pelatihan model.

3. HASIL DAN PEMBAHASAN

Dalam tahap ini, dilakukan proses klasifikasi menggunakan algoritma Naïve Bayes untuk mengidentifikasi jenis gangguan tidur pada individu. Proses dimulai dengan menghitung probabilitas awal (prior probability) dari masing-masing kelas target, yaitu *None* dan *Insomnia*. Probabilitas ini dihitung berdasarkan frekuensi kemunculan masing-masing kelas dalam dataset yang digunakan. Nilai probabilitas ini akan menjadi dasar dalam perhitungan posterior pada tahap selanjutnya. Perhitungan probabilitas kelas dapat dilihat pada Tabel 4.

Tabel 4. Probabilitas Class

Menentukan Probabilitas class		
None	297/374	0,794
Insomnia	77/374	0,206

Tabel 4 menunjukkan hasil perhitungan probabilitas masing-masing kelas dari variabel target *Sleep Disorder*, yaitu *None* dan *Insomnia*. Dari total 374 data, terdapat 297 individu yang tidak mengalami gangguan tidur (*None*) dan 77 individu yang mengalami gangguan tidur (*Insomnia*). Dengan demikian, probabilitas kelas *None* adalah 297 dibagi 374 atau sekitar 0,794, sedangkan probabilitas kelas *Insomnia* adalah 77 dibagi 374 atau sekitar 0,206. Nilai probabilitas ini akan digunakan sebagai probabilitas prior ($P(Y)$) dalam perhitungan Naïve Bayes untuk menentukan kemungkinan seseorang mengalami gangguan tidur berdasarkan fitur-fitur lainnya.

Setelah menentukan probabilitas awal untuk setiap kelas, langkah selanjutnya adalah menghitung jumlah data pada masing-masing atribut berdasarkan kelas target, yaitu *None* dan *Insomnia*. Tujuan dari langkah ini adalah untuk mengetahui distribusi nilai dari setiap atribut terhadap masing-masing kelas, yang kemudian digunakan dalam perhitungan probabilitas kondisi (likelihood) pada algoritma Naïve Bayes. Tabel 2 menyajikan jumlah kemunculan masing-masing kategori atribut berdasarkan kelas gangguan tidur.

Tabel 2. Jumlah Data pada Masing-Masing Atribut Berdasarkan Kelas

Atribut	Kategori	None	Insomnia
Total Data Kelas		297	77
Age	Adult	99	7

	Old	129	63
	Very Old	69	7
Sleep Duration	Enough	176	5
	Not Enough	121	72
Quality of Sleep	Very Good	174	6
	Good	116	66
	Bad	7	5
Stress Level	Low	177	31
	High	120	46
BMI Category	Normal	207	9
	Obese	90	68
Blood Pressure	Low	40	1
	Normal	122	3
	High	135	73
Heart Rate	65 - 72 bpm	243	67
	73 - 80 bpm	46	6
	81 - 86 bpm	8	4

Tabel 2 menyajikan distribusi jumlah data pada setiap kategori atribut berdasarkan dua kelas utama, yaitu *None* (tidak mengalami gangguan tidur) dan *Insomnia*. Terdapat total 297 data untuk kelas *None* dan 77 data untuk kelas *Insomnia*. Pada atribut **Age**, kelompok usia *Old* (dewasa tua) paling banyak mengalami insomnia dengan 63 kasus, sementara pada kelas *None*, distribusi usia relatif merata antara *Adult* dan *Old*. Untuk atribut **Sleep Duration**, terlihat perbedaan yang cukup signifikan, di mana 72 dari 77 individu yang mengalami insomnia memiliki durasi tidur yang tidak cukup (*Not Enough*), sedangkan mereka yang memiliki tidur cukup (*Enough*) lebih banyak ditemukan pada kelompok *None*. Atribut **Quality of Sleep** juga menunjukkan pola yang jelas; individu dengan kualitas tidur *Very Good* sebagian besar tidak mengalami gangguan tidur, sedangkan mereka yang mengalami insomnia mayoritas memiliki kualitas tidur *Good* atau *Bad*. **Stress Level** turut menjadi faktor penting, dengan 46 dari 77 penderita insomnia berada dalam kategori stres tinggi, sementara individu tanpa gangguan tidur lebih dominan pada stres rendah. Pada atribut **BMI Category**, kelompok dengan kategori *Obese* mencatatkan 68 dari 77 kasus insomnia, menunjukkan korelasi kuat antara obesitas dan gangguan tidur. Begitu juga dengan **Blood Pressure**, di mana tekanan darah tinggi menjadi faktor yang paling umum dalam kasus insomnia (73 kasus). Terakhir, pada atribut **Heart Rate**, sebagian besar individu baik yang mengalami maupun tidak mengalami gangguan tidur memiliki denyut jantung pada rentang 65–72 bpm, meskipun penderita insomnia juga ditemukan pada rentang yang lebih tinggi.

Setelah seluruh probabilitas atribut terhadap masing-masing kelas dihitung, langkah berikutnya adalah melakukan pengujian terhadap data baru menggunakan algoritma Naïve Bayes. Proses ini bertujuan untuk memprediksi apakah pasien mengalami gangguan tidur (*Insomnia*) atau tidak (*None*) berdasarkan nilai-nilai atribut yang dimilikinya. Data uji yang digunakan berasal dari pasien dengan ID 301, yang diambil dari dataset uji. Nilai-nilai atribut dari pasien ini akan dihitung menggunakan rumus probabilitas posterior untuk mengetahui kelas dengan probabilitas tertinggi. Data uji pasien 301 disajikan pada tabel 3.

Tabel 3. Data Uji Pasien

Atribut	Nilai
Person ID	301
Gender	Female
Age	Very Old
Sleep Duration	Enough
Quality of Sleep	Very good
Stress Level	Low
BMI Category	Normal
Blood Pressure	Normal
Heart Rate	65 – 72 bpm
Sleep Disorder (Asli)	None

P(Data Uji | None) = P(sleep Disorder | None)* (Very Old | None)* (Enough | None)* (Very Good | None)* (Low | None)* (Normal | None)* (Normal | None) * (65-72 | None)

$$= 0,794 * (0,232*0,593*0,586*0,596*0,697*0,411*0,818)$$

$$= 0,0112597$$

P(Data Uji | Insomnia) = P(sleep Disorder | Insomnia)* (Very Old | Insomnia)* (Enough | Insomnia)* (Very Good | Insomnia)* (Low | Insomnia)* (Normal | Insomnia)* (Normal | Insomnia) * (65-72 | Insomnia)

$$= 0,206 * (0,091*0,065*0,078*0,403*0,117*0,093*0,870)$$

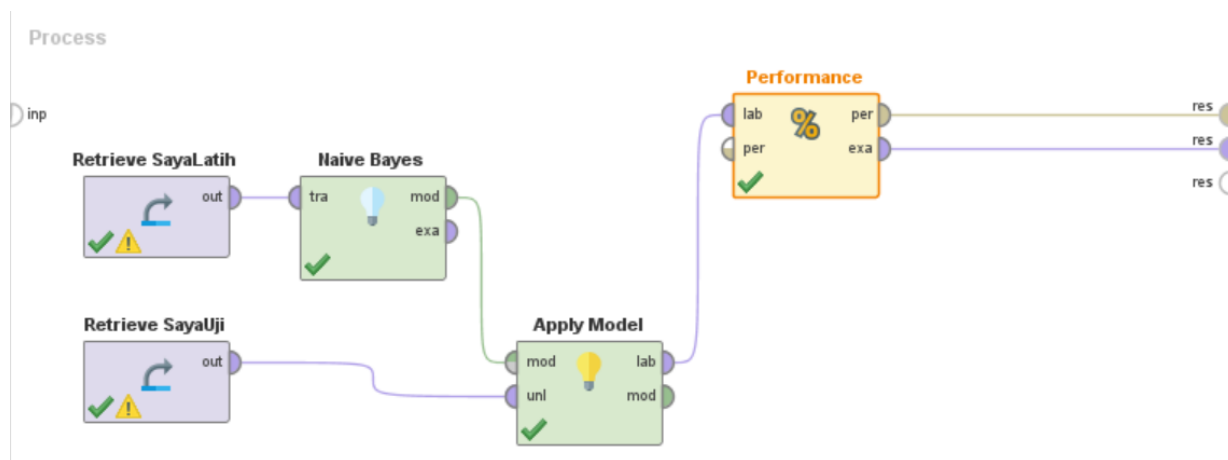
$$= 0,0000002$$

Pada hasil perhitungan diatas maka diperoleh

P(Data Uji | None) > P(Data Uji | Insomnia), Maka untuk data 301 diperoleh hasil sleep disorder :None

Berdasarkan perhitungan probabilitas menggunakan algoritma Naïve Bayes terhadap data uji dengan ID pasien 301, dilakukan perbandingan nilai *posterior probability* untuk dua kelas, yaitu *None* dan *Insomnia*. Nilai probabilitas kondisi data uji terhadap kelas *None* diperoleh sebesar **0,0112597**, sementara nilai probabilitas terhadap kelas *Insomnia* hanya sebesar **0,0000002**. Nilai-nilai ini dihitung dari hasil perkalian antara probabilitas awal (*prior probability*) masing-masing kelas dengan nilai probabilitas bersyarat dari setiap atribut yang dimiliki oleh data uji, seperti usia (*Very Old*), durasi tidur (*Enough*), kualitas tidur (*Very Good*), tingkat stres (*Low*), kategori BMI (*Normal*), tekanan darah (*Normal*), dan detak jantung (*65–72 bpm*). Karena nilai **$P(\text{Data Uji} \mid \text{None})$** lebih besar dibandingkan **$P(\text{Data Uji} \mid \text{Insomnia})$** , maka berdasarkan prinsip klasifikasi Naïve Bayes, data pasien 301 diklasifikasikan ke dalam kelas **Sleep Disorder: None**, yang berarti pasien tersebut tidak terindikasi mengalami gangguan tidur berdasarkan data yang dimiliki. Hasil ini mengonfirmasi bahwa algoritma Naïve Bayes mampu melakukan prediksi dengan dasar probabilistik secara sistematis.

Setelah dilakukan perhitungan probabilitas dan pengujian terhadap data uji, tahap berikutnya adalah membangun dan menerapkan model klasifikasi Naïve Bayes menggunakan perangkat lunak data mining. Pada gambar 3 ditunjukkan alur proses pemodelan klasifikasi mulai dari proses *retrieve* data latih dan data uji, pelatihan model Naïve Bayes, penerapan model terhadap data uji, hingga evaluasi performa model menggunakan komponen *Performance*. Alur ini memastikan bahwa data latih digunakan untuk membentuk model, sedangkan data uji digunakan untuk menilai kemampuan prediksi dari model yang dibentuk.



Gambar 3. Proses Pemodelan dan Evaluasi Klasifikasi Menggunakan Algoritma Naïve Bayes

Gambar 3 menggambarkan alur proses klasifikasi menggunakan algoritma Naïve Bayes dalam sebuah platform data mining (seperti RapidMiner). Proses dimulai dari dua komponen input, yaitu Retrieve SayaLatih dan Retrieve SayaUji, yang masing-masing digunakan untuk memanggil dataset pelatihan dan dataset pengujian. Dataset pelatihan kemudian disalurkan ke komponen Naïve Bayes, di mana model klasifikasi dilatih berdasarkan atribut-atribut yang tersedia. Setelah model terbentuk, komponen Apply Model digunakan untuk menerapkan model yang telah dilatih terhadap data uji. Komponen ini membutuhkan dua input, yaitu model hasil pelatihan dan data uji (unlabeled data). Hasil klasifikasi dari Apply Model selanjutnya dikirim ke komponen Performance, yang bertugas mengevaluasi performa model terhadap label aktual dari data uji. Komponen ini akan menghitung metrik evaluasi seperti akurasi, presisi, recall, F1-score, dan lainnya untuk mengetahui seberapa baik model Naïve Bayes dalam mengklasifikasikan data. Diagram ini menunjukkan hubungan antar komponen secara sistematis, yang

mencerminkan proses utama dalam penerapan machine learning supervised, yaitu pelatihan, penerapan, dan evaluasi model.

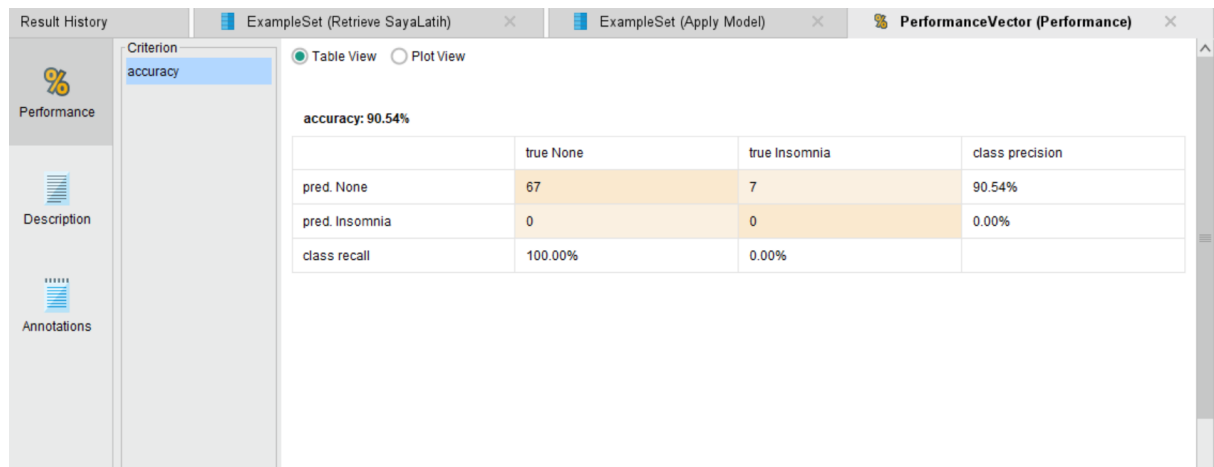
Setelah proses pelatihan dan penerapan model dilakukan, langkah berikutnya adalah menganalisis hasil prediksi dari algoritma Naïve Bayes terhadap data uji. Output dari proses ini ditampilkan dalam tampilan tabular seperti pada Gambar 4, yang memperlihatkan perbandingan antara label aktual dan hasil prediksi model. Selain itu, ditampilkan pula tingkat kepercayaan (confidence) dari setiap prediksi terhadap masing-masing kelas. Informasi ini penting untuk mengevaluasi kinerja model dalam mengklasifikasikan gangguan tidur secara otomatis berdasarkan atribut-atribut yang tersedia.

Row No.	Person ID	Sleep Disorder	prediction(Sleep Disorder)	confidence(None)	confidence(Insomnia)	Gender	Age	Sleep Duration	Quality of Sleep
1	301	None	None	1.000	0.000	Female	Very Old	Enough	Very good
2	302	None	None	1.000	0.000	Female	Very Old	Enough	Very good
3	303	None	None	1.000	0.000	Female	Very Old	Not Enough	Good
4	304	None	None	0.998	0.002	Female	Very Old	Enough	Good
5	305	None	None	0.973	0.027	Female	Very Old	Not Enough	Good
6	306	None	None	0.973	0.027	Female	Very Old	Not Enough	Good
7	307	Insomnia	None	0.939	0.061	Female	Very Old	Not Enough	Good
8	308	Insomnia	None	0.939	0.061	Female	Very Old	Not Enough	Good
9	309	Insomnia	None	0.939	0.061	Female	Very Old	Not Enough	Good
10	310	Insomnia	None	0.939	0.061	Female	Very Old	Not Enough	Good
11	311	Insomnia	None	0.939	0.061	Female	Very Old	Not Enough	Good
12	312	Insomnia	None	0.939	0.061	Female	Very Old	Not Enough	Good

Gambar 4. Hasil Prediksi Klasifikasi Gangguan Tidur Menggunakan Model Naïve Bayes di RapidMiner

Pada Gambar 4, proses pengujian model klasifikasi Naïve Bayes dilakukan menggunakan perangkat lunak RapidMiner. Dalam tahap ini, data latih dan data uji telah berhasil diimpor ke dalam sistem melalui operator *Retrieve SayaLatih* dan *Retrieve SayaUji*. Selanjutnya, pengguna menyeret (*drag*) semua operator yang diperlukan ke dalam canvas proses, termasuk operator Naïve Bayes, Apply Model, dan Performance. Setelah model dilatih menggunakan data latih, hasil prediksi terhadap data uji ditampilkan dalam bentuk tabel yang mencakup informasi seperti *Person ID*, *Sleep Disorder (aktual)*, *prediction(Sleep Disorder)*, dan tingkat *confidence* (kepercayaan model terhadap prediksi yang dihasilkan). Nilai confidence ditunjukkan dalam dua kolom: satu untuk kelas “None” dan satu lagi untuk “Insomnia”, dengan rentang nilai antara 0 hingga 1. Prediksi dengan confidence mendekati 1 menunjukkan bahwa model sangat yakin terhadap klasifikasi tersebut. Berdasarkan tampilan ini, dapat dilihat bahwa sebagian besar prediksi memiliki tingkat akurasi dan keyakinan yang tinggi, yang menunjukkan bahwa model Naïve Bayes bekerja dengan baik terhadap data uji yang digunakan.

Langkah terakhir dalam proses klasifikasi menggunakan algoritma Naïve Bayes adalah mengevaluasi performa model terhadap data uji. Evaluasi ini dilakukan dengan menggunakan metrik *confusion matrix*, yang menunjukkan perbandingan antara nilai prediksi model dengan nilai sebenarnya. Gambar 5 memperlihatkan *confusion matrix* beserta nilai akurasi yang dihasilkan oleh model. Informasi ini sangat penting untuk mengetahui seberapa baik model dapat mengklasifikasikan data dengan benar, terutama dalam membedakan antara pasien yang mengalami gangguan tidur (*Insomnia*) dan yang tidak (*None*).



Gambar 5. Confusion Matrix dan Akurasi Model Naïve Bayes terhadap Data Uji

Dari Gambar 5, diketahui bahwa model menghasilkan *confusion matrix* pada tabel 4.

Tabel 4. Perhitungan Evaluasi Kinerja Model Naïve Bayes Berdasarkan Confusion Matrix

	Prediksi: None	Prediksi: Insomnia
True: None	67 (True Negative)	0 (False Positive)
True: Insomnia	7 (False Negative)	0 (True Positive)

Tabel 4 menunjukkan hasil *confusion matrix* dari pengujian model Naïve Bayes. Model berhasil memprediksi 67 data dengan label *None* secara benar (**True Negative**) dan tidak ada kesalahan klasifikasi sebagai *Insomnia* (**False Positive** = 0). Namun, dari 7 data berlabel *Insomnia*, semuanya salah diklasifikasikan sebagai *None* (**False Negative** = 7), dan tidak ada yang terklasifikasi dengan benar (**True Positive** = 0). Hal ini menandakan bahwa model sangat baik mengenali kelas *None*, tetapi gagal total dalam mengenali kelas

Setelah proses klasifikasi dilakukan, langkah penting berikutnya adalah mengevaluasi kinerja model menggunakan metrik evaluasi. Salah satu metrik yang paling umum digunakan dalam evaluasi model klasifikasi adalah akurasi (*accuracy*). Akurasi memberikan gambaran seberapa besar proporsi prediksi yang dilakukan model sesuai dengan label sebenarnya dari data uji. Nilai akurasi yang tinggi menunjukkan bahwa model memiliki kemampuan prediktif yang baik secara keseluruhan. Berikut adalah perhitungan akurasi berdasarkan confusion matrix yang telah diperoleh:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Akurasi} = \frac{0 + 67}{0 + 67 + 0 + 7} = \frac{67}{74} \approx 0.9054 = 90.54\%$$

Hasil perhitungan akurasi dari model Naïve Bayes menunjukkan nilai sebesar 90,54%, yang diperoleh dari rasio jumlah prediksi yang benar terhadap total data uji. Nilai ini berarti bahwa model berhasil mengklasifikasikan 67 dari 74 data uji secara tepat. Namun, meskipun akurasi tergolong tinggi, perlu diperhatikan bahwa semua prediksi yang benar berasal dari kelas *None*, sementara seluruh data berlabel *Insomnia* (sebanyak 7 data) gagal diklasifikasikan dengan benar.

Dengan kata lain, akurasi tinggi ini dipengaruhi oleh dominasi jumlah data pada kelas *None*, sehingga model tampak "baik" secara keseluruhan, padahal kinerjanya buruk dalam mengenali kelas minoritas (*Insomnia*). Oleh karena itu, meskipun akurasi berguna sebagai metrik umum, evaluasi kinerja model sebaiknya juga mempertimbangkan metrik lain seperti recall, precision, dan F1-score, terutama dalam kasus klasifikasi dengan distribusi kelas yang tidak seimbang.

Kegagalan model Naïve Bayes dalam mengenali kelas *Insomnia* ini merupakan salah satu keterbatasan utama yang perlu dicermati. Hal ini terjadi karena model sangat bergantung pada frekuensi data dalam proses pelatihan, sehingga kelas minoritas yang jumlahnya sedikit akan memiliki probabilitas prediksi yang sangat rendah. Untuk mengatasi hal ini, strategi jangka panjang yang dapat diterapkan meliputi penggunaan teknik penyeimbangan data seperti SMOTE atau *undersampling*, serta penerapan pembobotan kelas (*class weight adjustment*) agar model tidak bias terhadap kelas mayoritas. Alternatif lainnya adalah membandingkan performa dengan algoritma lain yang lebih toleran terhadap ketidakseimbangan data seperti Random Forest atau Gradient Boosting, serta memperluas dataset dengan jumlah sampel yang seimbang antar kelas. Dengan pendekatan ini, diharapkan performa klasifikasi dapat menjadi lebih adil, akurat, dan representatif terhadap semua kategori gangguan tidur.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Naïve Bayes untuk mengidentifikasi gangguan tidur berdasarkan sejumlah atribut seperti usia, durasi tidur, kualitas tidur, tingkat stres, kategori BMI, tekanan darah, dan detak jantung. Model yang dibangun menghasilkan akurasi sebesar **90,54%** pada data uji, yang menunjukkan bahwa algoritma ini mampu memberikan performa klasifikasi yang baik secara umum. Kelebihan utama dari algoritma Naïve Bayes terletak pada kesederhanaan model, efisiensi komputasi, serta kemampuannya dalam menangani dataset berskala kecil hingga menengah.

Namun demikian, hasil evaluasi juga mengungkap kelemahan penting dari model ini, yaitu ketidakmampuannya dalam mengenali kelas minoritas, yakni *Insomnia*. Hal ini dibuktikan dengan nilai *recall* dan *precision* sebesar 0% untuk kelas tersebut, serta seluruh kasus *Insomnia* dalam data uji yang salah diklasifikasikan sebagai *None*. Keadaan ini mengindikasikan bahwa model sangat terpengaruh oleh ketidakseimbangan distribusi kelas dalam dataset.

Untuk pengembangan selanjutnya, penelitian ini menyarankan penerapan teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Over-sampling Technique*) atau *undersampling*, guna meningkatkan sensitivitas model terhadap kelas minoritas. Selain itu, perbandingan performa dengan algoritma lain seperti Random Forest, Support Vector Machine (SVM), atau Gradient Boosting juga dapat memberikan alternatif solusi yang lebih akurat.

Sebagai rencana implementasi sistem nyata, model ini dapat dikembangkan menjadi sebuah sistem pendukung keputusan (Decision Support System) yang terintegrasi dengan aplikasi kesehatan berbasis web atau mobile. Sistem ini dapat digunakan oleh tenaga medis seperti dokter atau bidan dalam melakukan skrining awal terhadap pasien, khususnya saat kunjungan rumah atau pemeriksaan kesehatan berkala. Dengan hanya memasukkan data dasar pasien seperti usia, kualitas tidur, tingkat stres, dan detak jantung, sistem akan secara otomatis memberikan prediksi jenis gangguan tidur yang mungkin terjadi. Hal ini akan sangat membantu tenaga medis dalam mengambil langkah intervensi dini sebelum pasien mengalami gangguan tidur yang lebih parah. Dengan dukungan sistem yang cepat, efisien, dan berbasis data, proses deteksi gangguan tidur dapat dilakukan secara lebih tepat waktu, akurat, dan terstandarisasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Universitas Potensi Utama atas segala dukungan yang telah diberikan selama proses penyusunan dan pelaksanaan penelitian ini. Dukungan fasilitas, bimbingan akademik, serta lingkungan yang kondusif telah memberikan kontribusi yang signifikan terhadap kelancaran dan keberhasilan penelitian ini.

REFERENSI

- [1] G. H. Carskadon and W. C. Dement, *Principles and Practice of Sleep Medicine*, 5th ed., Elsevier Saunders, 2011.
- [2] D. J. Hand and K. Yu, "Idiot's Bayes—not so stupid after all," *International Statistical Review*, vol. 69, no. 3, pp. 385–398, 2001.
- [3] "Naive Bayes classifier," *Wikipedia*. [Online]. Available: https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- [4] A. Ghifari and E. R. Widasari, "Sistem Pendeteksi Sleep-Disordered Breathing menggunakan Naïve Bayes," *J. Teknol. Inform. Ilmu Komput.*, vol. 10, no. 4, pp. 815–822, 2023.
- [5] I. A. Prabowo, D. Remawati, and A. P. W. Wardana, "Klasifikasi Tingkat Gangguan Tidur Menggunakan Algoritma Naïve Bayes," *J. TIKomSiN*, vol. 8, no. 2, 2018.
- [6] B. D. Kurniawan and M. Akbar, "Sistem Pakar Diagnosa Gangguan Tidur pada Anak Menggunakan Naïve Bayes," *J. Comput. Syst. Inform.*, vol. 5, no. 4, 2023.
- [7] M. Setiawati, D. Aldianto, and S. Sandiwarno, "Prediction Analysis of Sleep Disorders Using ML-Based Techniques," *J. Sistem Informasi Bisnis*, vol. 15, no. 1, pp. 89–101, 2025.
- [8] A. Smith et al., "Application of Machine Learning to Sleep Stage Classification," *arXiv preprint arXiv:2111.10577*, 2021.
- [9] N. S. Chen et al., "Utilizing Multi-Class Classification Methods for Automated Sleep Disorder Prediction," *Information*, vol. 15, no. 8, pp. 295–308, 2024.
- [10] V. R. K. Sathish et al., "Predicting Sleeping Quality using CNN," *arXiv preprint arXiv:2204.00516*, 2022.
- [11] Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Variable selection for Naïve Bayes classification. *Computers & Operations Research*, 135, 105456.
- [12] Jefriyanto, J., Ainun, N., & Al Ardha, M. A. (2023). Application of Naïve Bayes Classification to Analyze Performance Using Stopwords. *Journal of Information System, Technology and Engineering*, 1(2), 49-53.
- [13] Kim, T., & Lee, J. S. (2023). Maximizing AUC to learn weighted naive Bayes for imbalanced data classification. *Expert Systems with Applications*, 217, 119564.
- [14] Xiong, Y., Ye, M., & Wu, C. (2021). Cancer Classification with a Cost-Sensitive Naive Bayes Stacking Ensemble. *Computational and Mathematical Methods in Medicine*, 2021(1), 5556992.
- [15] Berrar, D. (2025). Bayes' theorem and naive Bayes classifier.