



Analisis Perbandingan Algoritma *Machine Learning* Dalam Klasifikasi Gangguan Tidur

Nabila Khansa^{1*}, Zaehol Fatah²

¹ Sistem Informasi, Universitas Ibrahimy Sukorejo Situbondo

² Sistem Informasi, Universitas Ibrahimy Sukorejo Situbondo

^{1*}nabilkhansamaudy@gmail.com, ²zaeholfatah@gmail.com

Abstrak

Gangguan tidur seperti insomnia dan *sleep apnea* merupakan masalah kesehatan global yang dapat menurunkan kualitas hidup. Deteksi dini terhadap gangguan ini penting dilakukan, khususnya dengan bantuan teknologi seperti algoritma data mining untuk meningkatkan ketepatan diagnosis. Data mining adalah bagian esensial dari analitik data dalam disiplin ilmu data science, yang memberikan berbagai manfaat luas dan aplikasi yang relevan. Penelitian ini menggunakan dataset *Sleep Health and Lifestyle* dari Kaggle untuk mengevaluasi kinerja tiga algoritma data mining, yaitu *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Neural Network*, dalam mengklasifikasi gangguan tidur. Proses pengembangan model mengikuti tahapan CRISP-DM dengan pengujian akurasi menggunakan *Cross Validation* dan evaluasi menggunakan *Confusion Matrix* dan kurva ROC. Berdasarkan hasil pengujian, algoritma *Neural Network* menunjukkan kinerja terbaik dengan akurasi 93,08% dan nilai AUC yang termasuk dalam klasifikasi "Excellent." Temuan ini menunjukkan bahwa *Neural Network* efektif dalam mengklasifikasi gangguan tidur, sehingga dapat mendukung proses diagnosa dan penanganan gangguan tidur secara lebih akurat.

Kata Kunci: Gangguan tidur, Klasifikasi, *Naïve Bayes*, *Support Vector Machine* (SVM), *Neural Network*

PENDAHULUAN

Gangguan tidur adalah masalah kesehatan yang dialami banyak orang di seluruh dunia dan berpotensi mempengaruhi kualitas hidup secara negatif (Faturrahman & Kurniasih, 2023). Beberapa faktor yang diyakini dapat menyebabkan insomnia meliputi jenis kelamin perempuan, usia, status pernikahan, tingkat pendapatan, dan jenjang pendidikan (Nisa, 2024). Dalam data ini, gangguan tidur meliputi insomnia, *sleep apnea*, dan kondisi tanpa gangguan tidur. Insomnia adalah gangguan tidur yang membuat penderitanya kesulitan untuk tidur atau tidak mendapatkan waktu tidur yang cukup (Huda, 2020). *Sleep Apnea* adalah kondisi medis di mana seseorang mengalami gangguan bernapas selama tidur, yang dapat mengakibatkan penurunan kualitas tidur (Kryger et al., 2017).

Banyak orang tidak menyadari bahwa telah mengalami gangguan tidur, meskipun terdapat gejala seperti sulit tidur atau kualitas tidur yang buruk. Bahkan, beberapa pihak sering kali baru memeriksakan diri ketika kondisi sudah parah. Di era perkembangan teknologi informasi dan komunikasi (TIK) telah mengubah dunia bisnis (Indahsari et al., 2024), maka penting untuk memiliki sistem yang mampu mendeteksi gangguan tidur lebih awal dengan menggunakan algoritma data mining, sehingga pemeriksaan dan penanganan dapat dilakukan secara cepat dan rutin. Salah satu metode yang dapat diterapkan dalam masalah ini adalah klasifikasi dengan bantuan algoritma data mining.

Klasifikasi didefinisikan sebagai proses pengelompokan data ke dalam kategori atau kelas berdasarkan karakteristik tertentu. Proses ini bertujuan untuk memprediksi label atau kelas dari data baru berdasarkan pola yang telah dipelajari dari data yang ada. Pada data Mining terdiri dari banyak algoritma yang dapat digunakan untuk proses klasifikasi yaitu *Naïve Bayes*, *Neural Network*, *Support Vector Machine* (SVM) dan masih banyak algoritma data mining lainnya (Suyanto, 2019).

Klasifikasi gangguan tidur menggunakan berbagai algoritma data mining menunjukkan bahwa algoritma *Naïve Bayes* mencapai akurasi 80% dalam klasifikasi gangguan tidur (Prabowo et al., 2020). Klasifikasi gangguan tidur dapat dilakukan menggunakan algoritma *Neural Network* dan *Support Vector Machine* (SVM), dengan hasil terbaik diperoleh DENGAN mencapai akurasi 91,2% (Sari, 2021).

Dalam penelitian ini, digunakan tiga algoritma data mining untuk klasifikasi gangguan tidur menggunakan *Sleep Health and Lifestyle Dataset* dari kaggle.com. Penelitian ini membandingkan ketiga algoritma tersebut untuk menentukan algoritma yang paling cocok dalam menyelesaikan kasus ini. Algoritma yang digunakan adalah *Naïve Bayes*, *Support Vector Machine* (SVM) dan *Neural Network*. *Naïve Bayes* adalah metode klasifikasi yang sederhana, yang menggunakan prinsip teorema probabilitas untuk menentukan peluang optimal dengan memperkirakan probabilitas kejadian di masa mendatang berdasarkan informasi dari data sebelumnya (Wibisono, 2023). *Neural Network* memiliki keunggulan dapat bekerja dengan baik pada data berukuran besar, karena kemampuannya untuk belajar dari pola dalam data dengan banyak variabel (Suyanto, 2019). *Support Vector Machine* (SVM) didefinisikan sebagai algoritma yang

bertujuan memisahkan data ke dalam kelas yang berbeda dengan mencari hyperplane optimal yang memberikan margin terbesar antara kelas-kelas tersebut (Suyanto, 2019).

Berdasarkan penjelasan latar belakang permasalahan, penelitian ini bertujuan untuk mencari metode algoritma yang paling akurat dalam memprediksi kemungkinan seseorang memiliki gangguan tidur.

METODE

Metode penelitian yang digunakan dalam studi ini adalah eksperimen evaluasi untuk menentukan algoritma yang paling akurat dan efektif dalam memprediksi kemungkinan seseorang terindikasi gangguan tidur. Adapun data yang digunakan pada penelitian ini adalah *Sleep Health and Lifestyle Dataset* yang diambil dari situs kaggle.com. Kaggle adalah platform yang menyediakan komunitas bagi *data scientist* dan penggemar data untuk berkolaborasi, bertukar pengetahuan, dan berkompetisi dalam tantangan analisis data yang didirikan oleh Anthony Goldbloom sebagai CEO dan Ben Hamner sebagai CTO pada tahun 2010 (Mulla, 2021). Dalam merancang model penelitian eksperimen ini menggunakan model standar yang umum dalam *Data Mining*, yaitu *Cross Industry Standard Process for Data Mining* (CRISP-DM) dan tahapan-tahapannya adalah sebagai berikut (Olson & Shi, 2007):

- Business Understanding*: Memahami tujuan bisnis dan kebutuhan pemangku kepentingan untuk menentukan masalah yang ingin diselesaikan.
- Data Understanding*: Mengumpulkan dan mengeksplorasi data untuk memahami sifat dan kualitasnya.
- Data Preparation*: Mempersiapkan data dengan pembersihan, transformasi, dan pemformatan agar siap untuk analisis.
- Modeling*: Menerapkan berbagai teknik pemodelan dan mengoptimalkan parameter untuk membangun model yang efektif.
- Evaluation*: Mengevaluasi hasil model untuk memastikan bahwa model memenuhi tujuan bisnis dan dapat diterapkan.
- Deployment*: Menerapkan model ke dalam sistem bisnis dan menyampaikan hasil serta rekomendasi kepada pemangku kepentingan.

Pengetahuan yang diperoleh pada tahap awal proses CRISP-DM akan membantu menemukan metode algoritma terbaik yang dapat digunakan untuk berbagai kebutuhan, termasuk memprediksi dan mengidentifikasi situasi penting.

HASIL DAN PEMBAHASAN

Dalam penelitian ini, dilakukan pengumpulan data yang didapat dari *website* Kaggle.com, yaitu *Sleep Health and Lifestyle Dataset* (*Sleep Health and Lifestyle Dataset*, n.d.). Dataset ini memiliki 374 data dengan 13 atribut dimana 12 atribut merupakan atribut biasa sedangkan 1 atribut sebagai *class* atau label. Deskripsi atribut berada pada tabel 1.

Tabel 1. Deskripsi Atribut

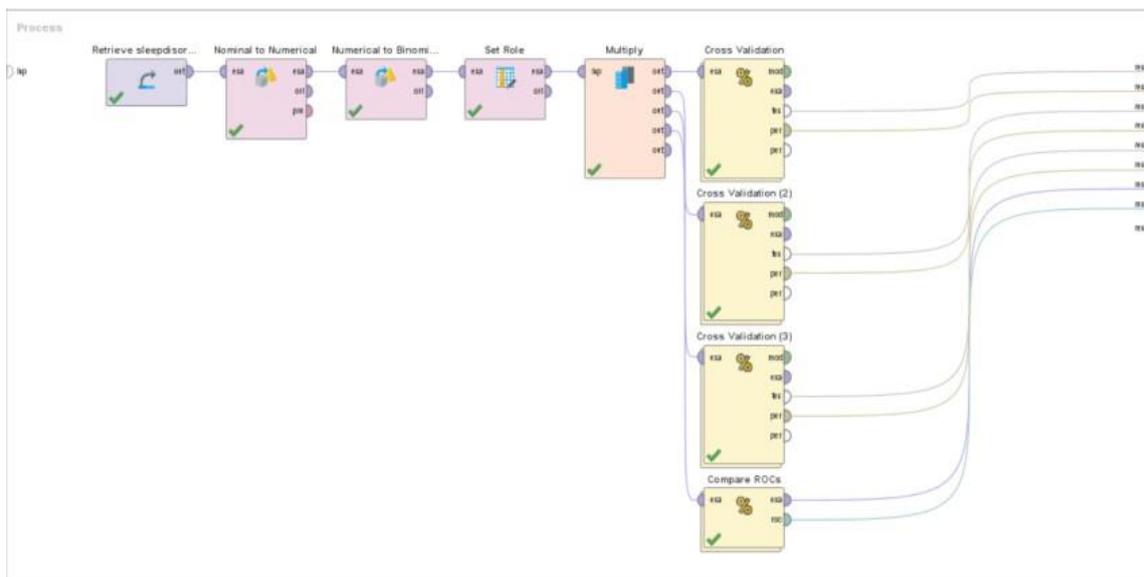
| Atribut | Keterangan |
|--|--|
| <i>Person ID</i> | Sebuah pengenal untuk setiap individu. |
| <i>Gender</i> | Jenis kelamin orang tersebut (Laki-laki/Perempuan). |
| <i>Age</i> | Usia orang tersebut dalam tahun. |
| <i>Occupation</i> | Pekerjaan atau profesi orang tersebut. |
| <i>Sleep Duration (hour)</i> | Jumlah jam tidur orang tersebut per hari. |
| <i>Quality of Sleep (scale: 1-10)</i> | Penilaian subjektif tentang kualitas tidur, dengan skala 1 hingga 10. |
| <i>Physical Activity Level (minutes/day)</i> | Jumlah menit aktivitas fisik yang dilakukan orang tersebut setiap hari. |
| <i>Stress Level (scale: 1-10)</i> | Penilaian subjektif tentang tingkat stres yang dialami, dengan skala 1 hingga 10. |
| <i>BMI Category</i> | Kategori BMI orang tersebut (misalnya, Kurus, Normal, Obesitas). |
| <i>Blood Pressure</i> | Pengukuran tekanan darah orang tersebut, ditunjukkan sebagai tekanan sistolik di atas tekanan diastolik. |
| <i>Heart Rate (bpm)</i> | Denyut jantung istirahat orang tersebut dalam detak per menit. |
| <i>Daily Steps</i> | Jumlah langkah yang diambil orang tersebut per hari. |

Data yang dikumpulkan kemudian dianalisis dan diolah, dengan atribut-atribut yang tidak relevan dibersihkan melalui tahap *preprocessing*.

Selanjutnya, dilakukan eksperimen permodelan data dan pengujian performa dengan menggunakan *Cross Validation* pada algoritma *Naïve Bayes*, *Support Vector Machine (SVM)* dan *Neural Network*. *Cross Validation* digunakan untuk mengevaluasi performa model yang lebih baik di berbagai subset data, meningkatkan keakuratan prediksi model pada data yang belum dilihat (Witten et al., 2016). Kemudian, performa model dianalisis menggunakan *Confusion matrix*, tabel yang digunakan untuk menggambarkan kinerja model klasifikasi dengan cara menghitung jumlah prediksi yang benar dan salah untuk masing-masing kelas (Witten et al., 2016).

Algoritma *Naïve Bayes* mendukung klasifikasi data dengan akurasi baik ketika tipe datanya nominal atau berbentuk huruf. Sementara itu, *Neural Network* dan *Support Vector Machine (SVM)* mampu menghasilkan akurasi tinggi dalam klasifikasi data numerik atau angka. Ini menunjukkan bahwa tipe data memiliki pengaruh besar dalam menyelesaikan masalah klasifikasi menggunakan algoritma data *mining* (Bishop, 2006).

Proses implementasi klasifikasi gangguan tidur menggunakan tiga algoritma, yaitu *Naïve Bayes*, *Support Vector Machine (SVM)* dan *Neural Network* ditampilkan dalam bentuk skema pada *RapidMiner Studio 9.1* seperti yang terlihat pada gambar 1.



Gambar 1. Skema Klasifikasi Gangguan Tidur Menggunakan *RapidMider Studio 9.1*

Implementasi klasifikasi gangguan tidur menggunakan 3 algoritma data *mining* di *RapidMiner Studio 9.1* memanfaatkan tujuh operator yang tersedia di aplikasi tersebut. Berikut adalah operator-operator tersebut beserta penjelasannya:

- Retreive Sleepdisorder*, operator ini mengandung file CSV yang telah diimpor. Data *Sleep Health and Lifestyle Dataset* diunduh dari situs kaggle.com dan sudah tersedia dalam format CSV.
- Nominal to Numerical*, operator ini digunakan untuk mengubah tipe data nominal menjadi numerik. Dalam *Sleep Health and Lifestyle Dataset*, atribut seperti *Blood Pressure*, *BMI Category*, *Gender*, *Occupation*, dan *Sleep Disorder* memiliki tipe nominal dan harus diubah menjadi numerik karena algoritma *Neural Network* tidak dapat memproses data nominal.
- Numerical to Binominal*, operator ini digunakan untuk mengubah tipe data numerik menjadi binomial. Dalam *Sleep Health and Lifestyle Dataset*, label *Sleep Disorder* yang sebelumnya telah diubah menjadi numerik, harus diubah ke data binomial karena operator *Compare ROSs* tidak dapat memproses label yang memiliki 3 kelas. Sedangkan label *Sleep Disorder* memiliki 3 kelas yaitu *None*, *Insomnia* dan *Sleep Apnea*.
- Set Role*, operator ini merupakan operator yang digunakan untuk menetapkan atribut kelas atau label dari data.
- Multiply*, operator yang digunakan untuk menghubungkan sejumlah operator agar dapat dijalankan secara bersamaan
- Cross Validation*, operator yang digunakan untuk menampilkan seberapa akurat kinerja dari algoritma. Dalam implementasi klasifikasi gangguan tidur, terdapat tiga operator *cross validation* karena hanya menggunakan tiga algoritma data *Mining*, sehingga jumlah *cross validation* tergantung pada jumlah algoritma yang digunakan. Operator ini mencakup operator *apply model*, operator untuk menerapkan model pada data training ke data testing, dan operator *performance* untuk melakukan evaluasi kinerja algoritma. Hasil evaluasi mencakup *accuracy*, *precision*, *recall*, serta *confusion matrix*.
- Compate ROCs*, operator yang digunakan untuk memvisualisasikan kinerja kurva ROC dari setiap algoritma. Dalam perbandingan ROC, terdapat operator untuk tiga algoritma yang digunakan dalam penelitian ini.

Setelah skema di *RapidMiner* selesai dieksekusi, hasilnya akan menampilkan akurasi dari ketiga algoritma yang digunakan untuk klasifikasi gangguan tidur. Adapun hasil dari pengujian dari masing masing algoritma yaitu:

Algoritma Naïve Bayes

Algoritma *Naïve Bayes* dijelaskan sebagai metode klasifikasi yang menggunakan pendekatan probabilistik. Algoritma ini memanfaatkan teorema Bayes dengan asumsi bahwa semua fitur yang ada bersifat independen satu sama lain. Hal ini membuat *Naïve Bayes* sangat efisien dalam memproses data dengan banyak atribut, dan sering digunakan dalam berbagai aplikasi (Wulandari, 2017). Adapun hasil *accuracy*, *recall*, *precision* serta *confussion matrix* algoritma *Naïve Bayes* yaitu:

accuracy: 93.08% +/- 5.29% (micro average: 93.05%)

| | true false | true true | class precision |
|--------------|------------|-----------|-----------------|
| pred. false | 206 | 13 | 94.06% |
| pred. true | 13 | 142 | 91.61% |
| class recall | 94.06% | 91.61% | |

recall: 91.63% +/- 7.34% (micro average: 91.61%) (positive class: true)

| | true false | true true | class precision |
|--------------|------------|-----------|-----------------|
| pred. false | 206 | 13 | 94.06% |
| pred. true | 13 | 142 | 91.61% |
| class recall | 94.06% | 91.61% | |

precision: 91.87% +/- 7.07% (micro average: 91.61%) (positive class: true)

| | true false | true true | class precision |
|--------------|------------|-----------|-----------------|
| pred. false | 206 | 13 | 94.06% |
| pred. true | 13 | 142 | 91.61% |
| class recall | 94.06% | 91.61% | |

Gambar 2. Hasil *accuracy*, *recall*, *precision* serta *confussion matrix* algoritma *Naïve Bayes*

Algoritma Support Vector Machine (SVM)

Algoritma *Support Vector Machine* (SVM) dijelaskan sebagai metode klasifikasi yang sangat efektif, terutama dalam menghadapi masalah klasifikasi dengan data yang memiliki dimensi tinggi. SVM bekerja dengan cara mencari *hyperplane* optimal yang memisahkan dua kelas dalam ruang fitur, di mana jarak antara *hyperplane* dan titik-titik data terdekat dari masing-masing kelas (disebut sebagai *support vectors*) dimaksimalkan (Wulandari, 2017). Adapun hasil *accuracy*, *recall*, *precision* serta *confussion matrix* algoritma *Support Vector Machine* (SVM) yaitu:

accuracy: 90.92% +/- 3.37% (micro average: 90.91%)

| | true false | true true | class precision |
|--------------|------------|-----------|-----------------|
| pred. false | 202 | 17 | 92.24% |
| pred. true | 17 | 138 | 89.03% |
| class recall | 92.24% | 89.03% | |

recall: 89.00% +/- 9.32% (micro average: 89.03%) (positive class: true)

| | true false | true true | class precision |
|--------------|------------|-----------|-----------------|
| pred. false | 202 | 17 | 92.24% |
| pred. true | 17 | 138 | 89.03% |
| class recall | 92.24% | 89.03% | |

precision: 89.83% +/- 6.98% (micro average: 89.03%) (positive class: true)

| | true false | true true | class precision |
|--------------|------------|-----------|-----------------|
| pred. false | 202 | 17 | 92.24% |
| pred. true | 17 | 138 | 89.03% |
| class recall | 92.24% | 89.03% | |

Gambar 3. Hasil *accuracy*, *recall*, *precision* serta *confussion matrix* algoritma *Support Vector Machine* (SVM)

Algoritma Neural Network

Algoritma *Neural Network* dijelaskan sebagai model komputasi yang terinspirasi oleh cara kerja otak manusia. *Neural Network* terdiri dari sejumlah neuron yang saling terhubung, di mana setiap neuron dapat menerima input, memprosesnya, dan menghasilkan output. Proses pembelajaran pada *Neural Network* dilakukan melalui metode yang dikenal sebagai *backpropagation*, di mana jaringan belajar dari kesalahan yang dihasilkan untuk memperbaiki bobot koneksi antar neuron (Wulandari, 2017). Adapun hasil *accuracy*, *recall*, *precision* serta *confussion matrix* algoritma *Neural Network* yaitu:

accuracy: 93.08% +/- 5.29% (micro average: 93.05%)

| | | | |
|--------------|------------|-----------|-----------------|
| | true false | true true | class precision |
| pred. false | 206 | 13 | 94.06% |
| pred. true | 13 | 142 | 91.61% |
| class recall | 94.06% | 91.61% | |

recall: 91.63% +/- 7.34% (micro average: 91.61%) (positive class: true)

| | | | |
|--------------|------------|-----------|-----------------|
| | true false | true true | class precision |
| pred. false | 206 | 13 | 94.06% |
| pred. true | 13 | 142 | 91.61% |
| class recall | 94.06% | 91.61% | |

precision: 91.87% +/- 7.07% (micro average: 91.61%) (positive class: true)

| | | | |
|--------------|------------|-----------|-----------------|
| | true false | true true | class precision |
| pred. false | 206 | 13 | 94.06% |
| pred. true | 13 | 142 | 91.61% |
| class recall | 94.06% | 91.61% | |

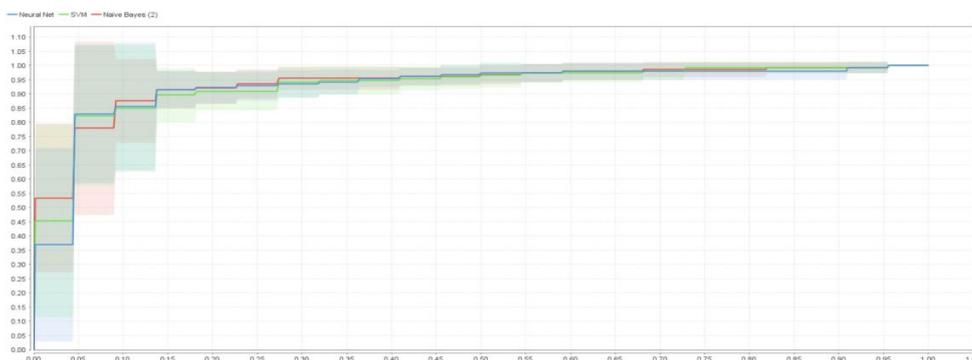
Gambar 4. Hasil accuracy, recall, precision serta confusion matrix algoritma Neural Network

Berikut adalah bentuk tabel dari hasil pengukuran hasil accuracy, recall, precision dari ketiga algoritma tersebut:

Tabel 2. Hasil Pengukuran Algoritma

| Algoritma | Accuracy | Recall | Precision |
|------------------------|----------|--------|-----------|
| Naïve Bayes | 90,90% | 90,29% | 88,78% |
| Support Vector Machine | 90,92% | 89,00% | 89,83% |
| Neural Network | 93.08% | 91,63% | 91,87% |

Berdasarkan tabel 2, algoritma Neural Network menunjukkan akurasi yang lebih tinggi dibandingkan dengan Naïve Bayes dan Support Vector Machine dengan akurasi sebesar 93.08%. Selain memiliki akurasi yang lebih tinggi, Neural Network juga mampu mengklasifikasikan lebih banyak penderita gangguan tidur dibandingkan dengan algoritma lainnya. Selain pengujian performa menggunakan Confusion Matrix, pengujian juga dilakukan menggunakan Compare ROCs untuk mengetahui perbandingan tingkat kehandalan dari algoritma Naïve Bayes, Support Vector Machine (SVM), dan Neural Network. Compare ROCs adalah grafik yang menunjukkan performa model klasifikasi pada berbagai threshold. Grafik ini memplot True Positive Rate (TPR) di sepanjang sumbu y dan False Positive Rate (FPR) di sepanjang sumbu x (Witten et al., 2016).



Gambar 4. Hasil Compare ROSS

Keandalan nilai AUC dapat diklasifikasikan menjadi lima kelompok: (1) 0.90 – 1.00 = Klasifikasi Luar Biasa (*Excellent Classification*); (2) 0.80 – 0.90 = Klasifikasi Baik (*Good Classification*); (3) 0.70 – 0.80 = Klasifikasi Cukup (*Fair Classification*); (4) 0.60 – 0.70 = Klasifikasi Buruk (*Poor Classification*); dan (5) 0.50 – 0.60 = Gagal (*Failure*) (G., 2015).

Berdasarkan hasil Compare ROCs pada gambar 4, terlihat bahwa algoritma Neural Network memiliki Compare ROCs yang tepat angka 1.0 pada sumbu Y, yaitu kelas True Positive (TP). Hal ini membuktikan bahwa Neural Network

adalah algoritma dengan kualitas klasifikasi “*Excellent*” dengan rentang akurasi 0.90 – 1.00, menunjukkan bahwa algoritma ini memiliki hasil klasifikasi yang sangat baik dalam menyelesaikan masalah klasifikasi gangguan tidur.

KESIMPULAN

Dari pengujian tiga algoritma, *Neural Network* terbukti efektif dalam mengklasifikasi gangguan tidur dengan akurasi 93,08% dan dikategorikan sebagai klasifikasi “*Excellent*” dengan nilai tepat 1.00 pada sumbu Y pada pengujian *Compare ROCs*. Model ini memiliki kemampuan yang baik dalam mengenali gangguan tidur, sehingga dapat membantu meningkatkan akurasi diagnosis dan kualitas penanganan.

DAFTAR PUSTAKA

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (1st ed.).
- Faturrahman, M. S., & Kurniasih, A. (2023). *Penggunaan Metode NearMiss, SMOTE, dan Naïve Bayes untuk Klasifikasi Gangguan Tidur Berdasarkan Kualitas Tidur dan Gaya Hidup*.
- G., G. (2015). *Data Mining: Concept, Models and Techniques*.
- Huda, M. (2020). *Mengatasi Insomnia Secara Alami*.
- Indahsari, L. N., Yunita, I., & Fatah, Z. (2024). INFORMATION SYSTEM DESIGN USING CUTOMER RELATIONSHIP MANAGEMENT (CRM) METHOD AT PAGLAK PETUNG CAFE AND ART IN BANYUWANGI DISTRICT. *Jurnal Teknik Informatika (Jutif)*.
- Kryger, M. H., Roth, T., & Dement, W. C. (2017). *Principles and Practice of Sleep Medicine* (6th ed.).
- Mulla, R. (2021). *From Zero to Kaggle Hero* (2nd ed.).
- Nisa, I. C. (2024). *PENERAPAN ALGORITMA K-MEANS DALAM PENGKATEGORIAN INSOMNIA*. 8, 117–122.
- Olson, D., & Shi, Y. (2007). *Introduction to Business Data Mining*.
- Prabowo, I. A., Remawati, D., & Wardana, A. P. W. (2020). Klasifikasi Tingkat Gangguan Tidur Menggunakan Algoritma Naïve Bayes. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.519>
- Sari, D. (2021). *Prediksi Gangguan Tidur pada Sleep Health and Lifestyle Menggunakan Support Vector Machine dan Neural Network*. 1(3), 23–30.
- Sleep Health and Lifestyle Dataset*. (n.d.). Retrieved October 30, 2024, from <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>
- Suyanto. (2019). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*.
- Wibisono, A. (2023). *FILTERING SPAM EMAIL MENGGUNAKAN METODE NAIVE BAYES*.
- Witten, H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.).
- Wulandari, R. T. (2017). *Data Mining: Teori dan Aplikasi* (1st ed.).