

Model Pembelajaran Mesin untuk Deteksi Gangguan Tidur: Perbandingan Logistic Regression dan Random Forest

Berliano Kurniawan¹, Fikri Budiman¹

¹ Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Indonesia
Correspondence: 111202113601@mhs.dinus.ac.id

Copyright: © 2025 by the authors

Received: 2 Oktober 2025 | Revised: 26 Oktober 2025 | Accepted: 23 November 2025 | Published: 6 Desember 2025

Abstrak

Gangguan tidur harus dikelompokkan dengan benar agar bisa terdeteksi lebih dini dan ditangani dengan baik. Namun, penelitian sebelumnya masih kurang lengkap karena hanya menggunakan satu metode atau menghadapi masalah data yang tidak seimbang. Penelitian ini bertujuan membandingkan kemampuan *Logistic Regression* dan *Random Forest* dalam mengklasifikasikan gangguan tidur menggunakan dataset Sleep Health dari Kaggle. Dataset berisi 374 contoh dan 12 variabel yang terbagi menjadi tiga kategori yaitu Insomnia, Tanpa Gangguan, dan Sleep Apnea. Proses data meliputi pembersihan, pembagian dataset, serta penyeimbangan kelas dengan teknik SMOTE untuk menangani ketidakseimbangan distribusi. Penilaian dilakukan dengan menggunakan metrik akurasi, precision, recall, F1-score, dan confusion matrix. Hasil temuan kami menunjukkan bahwa *Logistic Regression* mencapai akurasi sebesar 96%, sedangkan *Random Forest* 92%. Secara ilmiah, *Logistic Regression* lebih stabil pada data yang seimbang, sedangkan *Random Forest* lebih rentan terhadap variasi akibat proses oversampling. Analisis model menunjukkan bahwa frekuensi terbangun, durasi tidur total, kualitas tidur yang dirasakan, serta gejala seperti mendengkur atau henti napas merupakan faktor paling berpengaruh dalam mendeteksi gangguan tidur. Hasil ini menunjukkan bahwa model yang sederhana bisa lebih efektif dibandingkan model yang rumit setelah data diimbang, dan penemuan ini belum banyak dibahas dalam literatur tentang kesehatan tidur.

Kata kunci: regresi logistik; random forest; klasifikasi; gangguan tidur; smote

Abstract

Sleep disorders must be classified accurately to enable early detection and proper treatment. However, previous studies remain limited because they often rely on a single method or suffer from imbalanced datasets. This study aims to compare the performance of Logistic Regression and Random Forest in classifying sleep disorders using the Sleep Health dataset from Kaggle. The dataset consists of 374 samples and 12 variables categorized into three classes: Insomnia, No Disorder, and Sleep Apnea. The data processing steps include cleaning, splitting the dataset, and balancing the classes using the SMOTE technique to address distribution imbalance. Model evaluation was conducted using Accuracy, precision, recall, F1-score, and confusion matrix. Our findings show that Logistic Regression achieved an Accuracy of 96%, while Random Forest obtained 92%. Scientifically, Logistic Regression demonstrated greater stability on balanced data, whereas Random Forest was more sensitive to variations introduced by oversampling. Model analysis revealed that frequency of nighttime awakenings, total sleep duration, perceived sleep quality, and symptoms such as snoring or breathing pauses were the most influential factors in detecting sleep disorders. These results indicate that simpler models can outperform more complex ones after class balancing, a finding that has not been widely reported in the existing sleep health literature.

Keywords: logistic regression; random forest; classification; sleep disorders; smote



PENDAHULUAN

Gangguan tidur adalah masalah kesehatan yang terjadi di seluruh dunia dan bisa memengaruhi kesehatan tubuh serta pikiran. Secara internasional, sekitar 16% orang dewasa mengalami insomnia secara medis, dan diperkirakan lebih dari 900 juta orang dewasa menderita gangguan tidur menghambat pernapasan (OSA). Kualitas tidur yang tidak cukup bisa membuat kita kurang berenergi, lebih rentan cedera, dan mengganggu suasana hati serta hubungan dengan orang lain. (Astari et al., 2024; Hartini et al., 2021; Hutagalung et al., 2022; Ruriyanty.R et al., 2023; Wardani & Tiastiningsih, 2023). Dalam bidang informatika kesehatan, penggunaan machine learning sangat penting karena mampu menganalisis data seperti lama dan kualitas tidur untuk mendeteksi serta mengelompokkan gangguan tidur dengan lebih cepat dan tepat, sehingga membantu dalam mendiagnosis lebih awal dan membuat keputusan berdasarkan informasi yang akurat (Andira et al., 2022; Faradillah et al., 2024; Monarisa et al., 2021; Rohmah, 2023).

Gangguan tidur menjadi salah satu masalah kesehatan yang semakin banyak ditemukan belakangan ini, terutama pada kelompok usia yang aktif dan sibuk. Beberapa penelitian menunjukkan peningkatan jumlah orang yang mengalami gejala seperti susah tidur, merasa sangat lelah, hingga pola tidur yang tidak teratur, hal ini bisa membuat perhatian, kemampuan bekerja, dan kondisi mental seseorang menurun (Sahri & Hasibuan, 2023). Meski demikian, dalam penelitian sebelumnya, cara mengenali gangguan tidur masih menggunakan metode tradisional yang kurang objektif, memakan waktu, dan tidak efektif untuk dianalisis dalam skala besar. Oleh karena itu, para peneliti kini mulai menggunakan pendekatan machine learning untuk menganalisis data tidur secara otomatis. Beberapa algoritma seperti *Logistic Regression* dan *Random Forest* telah terbukti efektif dalam mengklasifikasikan jenis tidur dan mendeteksi gangguan seperti insomnia atau sleep apnea, sementara teknik SMOTE sering digunakan untuk membuat data lebih seimbang agar hasil prediksi lebih akurat (Hidayat, 2023; Kivrak et al., 2024; Rao et al., 2025).

Meskipun banyak penelitian sudah menggunakan algoritma machine learning untuk mengklasifikasikan gangguan tidur, masih sedikit penelitian yang secara rapi membandingkan hasil antara model-model yang sederhana dan yang lebih rumit setelah data diimbang secara tepat. Penelitian ini menggunakan dataset Sleep Health (Maulidah & Hidayati, 2024; Setyowati et al., 2023) untuk mengklasifikasikan gangguan tidur dengan dua algoritma, yaitu *Logistic Regression* dan *Random Forest*. *Logistic Regression* dipilih karena mampu memodelkan hubungan antara fitur dan label secara sederhana dan efisien, serta mudah dipahami, sehingga cocok untuk dataset kesehatan dengan jumlah fitur yang tidak terlalu banyak (Cahyani & Budiman, 2025; Mufid & Kismiantini, 2024). Sementara itu, *Random Forest* dipilih karena mampu menangani pola yang tidak linear dan interaksi antar variabel (Syahreza et al., 2024) seperti durasi tidur, tingkat stres, dan kebiasaan sehari-hari, serta tahan terhadap masalah overfitting (Larasati et al., 2024). Beberapa penelitian di bidang health informatics menunjukkan bahwa kedua algoritma ini memiliki performa yang baik dalam mendeteksi penyakit dan menganalisis pola kesehatan berbasis data (Irfannandhy et al., 2024; Pangestu et al., 2025). Untuk mengatasi ketidakseimbangan kelas, digunakan teknik SMOTE, sedangkan penilaian kinerja model dilakukan dengan menggunakan beberapa metrik seperti *Accuracy*, *precision*, *recall*, dan *f1-score* untuk melihat efektivitas model secara menyeluruh.

Ketidakseimbangan antar kelas menjadi tantangan besar dalam menganalisis data kesehatan, termasuk data gangguan tidur yang umumnya didominasi oleh kategori "normal". Situasi ini menyebabkan model *machine learning* lebih mudah mengenali pola dari kelas yang paling banyak (Haikal et al., 2023; Triginandri & Subhiyakto 2024), sementara mengabaikan kelas yang lebih sedikit, seperti insomnia atau sleep apnea. Hal ini membuat model kurang sensitif dalam mendeteksi kasus dari kelas minoritas (Rachmat et al., 2023). Untuk mengatasi masalah tersebut, penelitian ini menggunakan teknik SMOTE, yaitu cara untuk

menyeimbangkan data dengan membuat sampel baru dari kelas yang jumlahnya sedikit tanpa mengubah distribusi data terlalu banyak (A'la, 2025). Setelah menerapkan SMOTE, penelitian ini membandingkan hasil dua jenis algoritma klasifikasi, yaitu *Logistic Regression* dan *Random Forest*, menggunakan beberapa ukuran penilaian seperti akurasi, presisi, recall, dan f1-score agar dapat mengevaluasi bagaimana efektifnya model tersebut secara menyeluruh.

Penelitian yang dilakukan oleh Ariyanto et al. (2024) menunjukkan bahwa metode *Logistic Regression* mampu mengklasifikasikan kualitas tidur dengan akurasi sebesar 77%, tetapi kinerjanya terbatas karena hanya menggunakan satu jenis algoritma. Di sisi lain, penelitian Khasanah et al., (2025) menemukan bahwa *Random Forest* memiliki akurasi lebih tinggi yaitu 89,69% dibandingkan dengan K-Nearest Neighbors (KNN) yang hanya mencapai 87,02%. Meskipun begitu, kedua metode tersebut masih mengalami kesulitan dalam mengenali kategori Sleep Apnea karena ketidakseimbangan jumlah data. *Logistic Regression* kurang efektif ketika digunakan pada data yang memiliki pola non-linear dan terdapat interaksi antar variabel seperti tingkat stres, durasi tidur, serta kebiasaan harian, sedangkan *Random Forest* cenderung memihak kelas mayoritas. Situasi ini menunjukkan adanya celah dalam penelitian, yaitu belum ada studi yang secara menyeluruh membandingkan kedua algoritma tersebut pada data yang sudah seimbang menggunakan teknik SMOTE untuk meningkatkan pengenalan terhadap kelas minoritas.

Penelitian ini bertujuan membandingkan kemampuan algoritma *Logistic Regression* dan *Random Forest* pada dataset Sleep Health yang telah diimbangi menggunakan teknik SMOTE. Dengan pendekatan ini, penelitian ini berusaha memberikan analisis yang lebih lengkap mengenai bagaimana efektivitas kedua algoritma berfungsi pada data yang tidak seimbang. Selain itu, penelitian ini juga bertujuan meningkatkan tingkat akurasi dan sensitivitas model dalam mengenali kelas minoritas seperti Sleep Apnea. Hasil yang diperoleh diharapkan dapat menjadi pedoman dalam memilih metode klasifikasi yang paling sesuai untuk menganalisis gangguan tidur, serta menjadi dasar pengembangan sistem deteksi otomatis yang lebih andal dalam bidang kesehatan tidur.

METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental berbasis komputer dengan menerapkan dua algoritma pembelajaran mesin, yaitu *Logistic Regression* dan *Random Forest*, untuk mengklasifikasikan jenis gangguan tidur. Dataset yang digunakan dalam penelitian ini berasal dari Kaggle <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset> yang berjudul Sleep Health and Lifestyle dan isi dataset terdiri dari 374 data dan 12 variabel, yaitu Person ID, Gender, Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Blood Pressure, Heart Rate, dan Daily Steps. Variabel target yang digunakan untuk klasifikasi adalah Sleep Disorder, yang dibagi menjadi tiga kategori, yaitu None, Insomnia, dan Sleep Apnea.

Sebelum memulai pemodelan, ada beberapa langkah persiapan data yang dilakukan yaitu data berupa angka seperti durasi tidur, detak jantung, tingkat stres, dan langkah harian diubah agar skala semua fitur sama menggunakan metode StandardScaler. Data berupa kategori seperti jenis kelamin, pekerjaan, dan kategori BMI diubah menjadi angka dengan metode one-hot encoding. Nilai yang kosong atau tidak terisi dicek dan diisi kembali dengan nilai rata-rata untuk data angka, serta dengan nilai paling umum untuk data kategori. Karena jumlah data pada setiap kategori target tidak sama, metode SMOTE digunakan untuk menambah data pada kategori yang sedikit, yaitu None, Insomnia, dan Sleep Apnea. Penambahan data ini hanya dilakukan pada data latih agar tidak terjadi kebocoran data. Setelah data seimbang, seluruh dataset dibagi menjadi 80% untuk latihan dan 20% untuk pengujian, dengan menjaga jumlah kategori tetap seimbang.

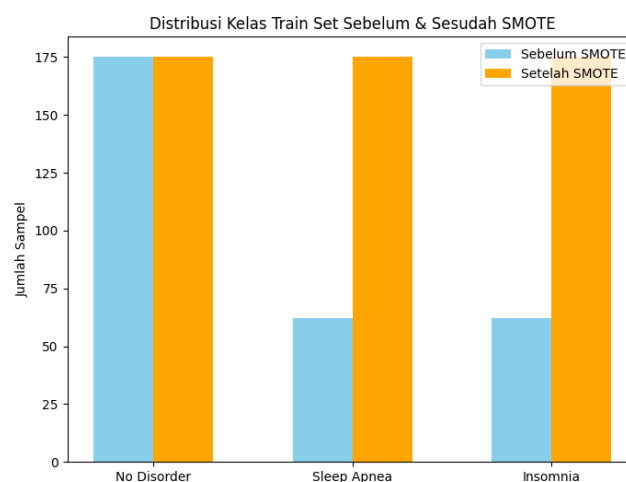
Model *Logistic Regression* dipilih karena kemampuannya dalam menjelaskan hubungan probabilitas antara fitur dan label dengan pendekatan linear, serta sering digunakan dalam studi kesehatan prediktif. Model ini menggunakan metode regularisasi L2 dengan batasan iterasi maksimal sebanyak 1000 dan nilai `random_state = 42`. Di sisi lain, *Random Forest* dipilih karena kemampuannya dalam menangkap hubungan non-linear antar variabel dan mengurangi risiko overfitting melalui mekanisme bagging. Model ini dibuat dengan 100 pohon keputusan (`n_estimators = 100`) menggunakan parameter default. Kedua model dilatih dengan library `scikit-learn` tanpa melakukan penyesuaian parameter (hyperparameter tuning) agar hasil perbandingan tetap terkontrol dan bisa direplikasi kembali.

Performa model dievaluasi menggunakan empat metrik utama, yaitu akurasi, presisi, *recall*, dan *f1-score*, untuk memberikan gambaran lengkap tentang kemampuan model dalam mengklasifikasikan data multikelas. Selain itu, Confusion Matrix digunakan untuk mengetahui jenis kesalahan klasifikasi pada setiap jenis gangguan tidur. Secara metode, kontribusi penelitian ini terletak pada desain eksperimen yang membandingkan dua model dengan tingkat kompleksitas berbeda pada data yang sudah disamakan menggunakan SMOTE, sehingga hasilnya memberikan gambaran nyata mengenai perbedaan efektivitas model linear dan non-linear dalam analisis kesehatan tidur.

HASIL DAN PEMBAHASAN

Hasil

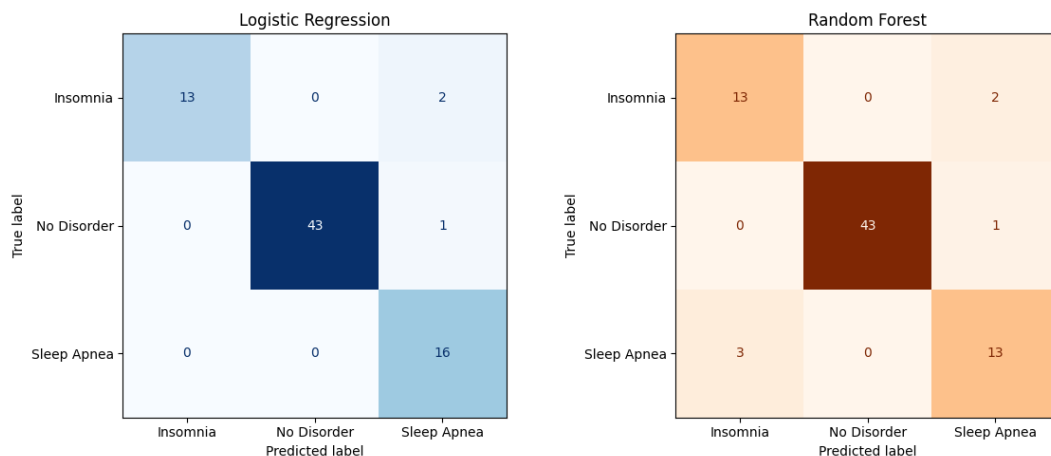
Penelitian ini menggunakan dataset *Sleep Health and Lifestyle* yang diambil dari platform Kaggle. Dataset tersebut berisi data dari 374 responden dengan variabel target yaitu Sleep Disorder yang dibagi menjadi tiga kategori: *No Disorder* (219 responden), *Insomnia* (77 responden), dan *Sleep Apnea* (78 responden). Karena terdapat ketidakseimbangan jumlah responden antar kategori, maka digunakan metode SMOTE untuk menjadikan jumlah data di setiap kategori seimbang. Sebelum membangun model, data dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%. Pembagian ini dilakukan dengan parameter `random_state = 42` agar hasil dapat direplikasi secara konsisten. Analisis dilakukan menggunakan Python versi 3.10 di platform Google Colab dengan beberapa library seperti `pandas`, `scikit-learn`, dan `imbalanced-learn`.



Gambar 1. Distribusi antar kelas

Gambar 1 memperlihatkan perbandingan cara distribusi kelas pada data latih sebelum dan setelah menggunakan SMOTE. Sebelum diterapkan SMOTE, data terlihat tidak seimbang karena kelas *No Disorder* mendominasi, sedangkan kelas *Sleep Apnea* dan *Insomnia* memiliki jumlah sampel yang lebih sedikit. Setelah menggunakan SMOTE, jumlah sampel di setiap

kelas menjadi sama, masing-masing sebanyak 175 data. Proses penyeimbangan ini membantu model belajar dengan lebih adil terhadap setiap kategori, sehingga mengurangi pengaruh kelas yang jumlahnya lebih banyak.



Gambar 2. *Confusion matrix*

Gambar 2 menampilkan matriks kebingungan dari kedua model. Pada *Logistic Regression*, sebagian besar prediksi benar, dengan total 43 data No Disorder, 13 Insomnia, dan 16 Sleep Apnea berhasil diklasifikasikan tepat, serta hanya ada tiga kesalahan klasifikasi. Sementara itu, *Random Forest* juga menunjukkan performa yang baik, namun sedikit menurun, dengan tiga data Sleep Apnea yang salah diklasifikasikan sebagai Insomnia. Secara keseluruhan, kedua model mampu mengenali pola data dengan baik, terutama pada kelas No Disorder yang terdeteksi sempurna. Namun, *Logistic Regression* lebih konsisten dalam mengenali kelas minoritas setelah penerapan SMOTE, sedangkan *Random Forest* lebih rentan terhadap perubahan data meskipun sedikit lebih sensitif terhadap variasi antar kelas.

Tabel 1. Hasil klasifikasi

Model	Accuracy	Precision	Recall	F1-Score
<i>Logistic Regression</i>	96%	95%	95%	94%
<i>Random Forest</i>	92%	88%	89%	88%

Hasil model *logistic regression* sesuai dengan tabel 1 menunjukkan akurasi tertinggi yaitu 96% dengan nilai *precision*, *recall*, dan *f1-score* semuanya di atas 94%. Metode ini mampu mengklasifikasikan semua kelas secara konsisten termasuk kategori minoritas seperti Sleep Apnea. Sementara itu, *Random Forest* mencapai akurasi 92%, tetapi performanya menurun pada kelas minoritas, menunjukkan adanya kemungkinan overfitting setelah proses SMOTE dilakukan. Analisis variabel menunjukkan bahwa durasi tidur, tingkat stres harian, dan kualitas tidur subjektif merupakan faktor-faktor yang paling berpengaruh dalam mendeteksi gangguan tidur. Hasil ini menegaskan bahwa faktor-faktor perilaku dan kondisi psikologis harian berperan penting dalam membedakan berbagai kategori gangguan tidur. Secara umum, *Logistic Regression* lebih efektif digunakan pada data yang telah diseimbangkan, sedangkan *Random Forest* lebih cocok digunakan pada dataset yang besar dengan distribusi alami.

Pembahasan

Hasil penelitian menunjukkan ada perbedaan nyata dalam kemampuan *Logistic Regression* dan *Random Forest* dalam mengklasifikasikan gangguan tidur. *Logistic Regression*

mencapai akurasi 96% dengan nilai *precision*, *recall*, dan *f1-score* di atas 94%, sedangkan *Random Forest* hanya mencapai akurasi 92% dan performa turun pada kelas minoritas. Perbedaan ini tidak hanya terlihat dari angka, tapi juga mencerminkan cara kerja algoritma dan bagaimana mereka berinteraksi dengan sifat data. *Logistic Regression* bekerja terbaik ketika ada hubungan linear antar variabel. Setelah data diimbang menggunakan metode SMOTE, distribusi kelas menjadi lebih merata, sehingga model bisa menentukan batasan keputusan lebih jelas, termasuk pada kategori minoritas seperti *Sleep Apnea*. Hasil yang stabil menunjukkan bahwa model sederhana yang sesuai dengan struktur data bisa menangkap pola utama tanpa terganggu oleh fluktuasi atau noise dari data sintetis. Sebaliknya, meskipun *Random Forest* mampu menangani pola yang tidak linear dan hubungan kompleks antar variabel, performanya cenderung menurun pada kelas minoritas karena terjadi *overfitting* pada pohon-pohon individu. Data yang dihasilkan dari SMOTE, meskipun sudah seimbang dalam jumlah sampel, tidak selalu mencerminkan pola alami secara sempurna, sehingga model bisa mengklasifikasikan beberapa sampel *Sleep Apnea* secara salah. Analisis fitur menunjukkan bahwa variabel yang paling berpengaruh dalam mendeteksi gangguan tidur adalah durasi tidur, tingkat stres harian, dan kualitas tidur yang dirasakan subjektif, karena ketiga variabel ini berperan penting dalam membedakan setiap jenis gangguan tidur. *Logistic Regression* menekankan variabel-variabel tersebut melalui koefisien yang signifikan, sementara *Random Forest* menempatkan variabel-variabel tersebut sebagai fitur dengan tingkat kepentingan tertinggi.

Hasil temuan menunjukkan bahwa *logistic regression* memiliki hasil yang lebih baik, sesuai dengan penelitian yang dilakukan oleh Ariyanto et al. (2024) yang juga menggunakan algoritma yang sama untuk memprediksi kualitas tidur dengan akurasi mencapai 77%. Peningkatan kinerja dalam penelitian ini berkat penerapan SMOTE yang berhasil membuat distribusi kelas lebih seimbang, sehingga model bisa lebih baik mengenali kategori minoritas seperti *Sleep Apnea*. Berbeda dengan itu, penelitian Khasanah et al. (2025) menyatakan bahwa *Random Forest* lebih unggul dibandingkan KNN dengan akurasi sebesar 89,69%. Keunggulan *Random Forest* terjadi karena kemampuannya dalam mengolah data yang memiliki pola non-linear dan hubungan antar variabel yang kompleks, yang sesuai dengan karakteristik data yang digunakan. Namun meskipun akurasinya tinggi, nilai *recall* pada kelas minoritas tetap rendah karena dataset tidak seimbang, sehingga model cenderung lebih fokus pada kelas mayoritas.

Temuan ini memberikan beberapa manfaat. Pertama, penelitian ini menekankan pentingnya kesesuaian antara karakteristik data dan tingkat kompleksitas model machine learning: model sederhana bisa lebih efektif pada dataset kecil yang sudah seimbang, sedangkan model kompleks cocok untuk dataset besar dengan distribusi alami. Kedua, hasil penelitian ini dapat menjadi dasar pengembangan sistem pendukung keputusan digital yang membantu dokter mendeteksi gangguan tidur seperti insomnia dan *Sleep Apnea* secara lebih cepat dan efisien. Ketiga, penelitian ini menyoroti variabel-variabel perilaku dan psikologis harian sebagai faktor utama yang memengaruhi klasifikasi gangguan tidur, sehingga bisa menjadi fokus dalam upaya intervensi kesehatan.

Keterbatasan dari penelitian ini antara lain jumlah sampel yang tidak terlalu banyak, yaitu hanya 374 sampel, serta sumber data yang hanya satu, sehingga belum mampu mencakup berbagai ragam pola tidur di masyarakat yang lebih luas. Selain itu, data yang dihasilkan melalui metode SMOTE bersifat buatan dan mungkin tidak sepenuhnya mencerminkan keadaan sebenarnya, sehingga masih ada kemungkinan terjadinya kesalahan dalam mengklasifikasikan data. Untuk penelitian berikutnya, disarankan menggunakan dataset yang lebih besar dan lebih beragam, termasuk berbagai kelompok orang dan faktor lingkungan yang bisa memengaruhi kebiasaan tidur. Dapat juga dibandingkan kinerja algoritma lain seperti *Gradient Boosting* atau *XGBoost* pada data yang seimbang maupun tidak seimbang. Penambahan data fisiologis nyata, seperti detak jantung atau pola aktivitas tidur, akan membuat

model lebih cocok untuk digunakan dalam aplikasi kesehatan yang membutuhkan prediksi gangguan tidur menggunakan machine learning.

SIMPULAN

Penelitian ini menunjukkan bahwa *Logistic Regression* adalah model yang paling efektif untuk mendeteksi gangguan tidur setelah data diimbangi menggunakan metode SMOTE, dengan akurasi mencapai 96%, lebih baik dibandingkan *Random Forest* yang hanya mencapai 92%. Hasil ini menunjukkan bahwa model yang lebih sederhana bisa memberikan performa yang lebih baik pada data kesehatan yang sudah seimbang, sehingga tidak semua model yang kompleks pasti lebih akurat. Secara praktis, penemuan ini membuka kemungkinan *Logistic Regression* digunakan sebagai model yang efisien dan handal dalam sistem deteksi dini gangguan tidur. Namun, karena ukuran dataset yang terbatas serta penggunaan data sintesis SMOTE, hasil ini belum bisa diterapkan secara luas. Oleh karena itu, penelitian berikutnya perlu menggunakan dataset yang lebih besar, melakukan validasi eksternal, serta menggunakan metode yang lebih canggih seperti *Gradient Boosting* atau *XGBoost* untuk meningkatkan keandalan model.

REFERENSI

- A'la, F. Y. (2025). Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia: IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 256-265. <https://doi.org/10.29408/edumatic.v9i1.29666>
- Andira, A. D., Usman, A. M., & Wowor, T. J. F. (2022). Hubungan Penggunaan Gadget Terhadap Kualitas Tidur Pada Mahasiswa Keperawatan di Universitas Nasional. *Jurnal Promotif Preventif*, 4(2), 51–56. <https://doi.org/10.47650/jpp.v4i2.354>
- Ariyanto, D., Sofro, A. Y., Hanifah, A. I. N., Prihanto, J. B., Maulana, D. A., & Romadhonia, R. W. (2024). Logistic and probit regression modeling to predict the opportunities of diabetes in prospective athletes. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 18(3), 1391-1402. <https://doi.org/10.30598/barekengvol18iss3pp1391-1402>
- Astari, D. F., Chrisnanto, Y. H., & Melina, M. (2024). Klasifikasi Tingkat Stres Saat Tidur Menggunakan Algoritma Random Forest. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3676–3684. <https://doi.org/10.36040/jati.v7i5.7750>
- Cahyani, O. N., & Budiman, F. (2025). Performa *Logistic Regression* dan Naive Bayes dalam Klasifikasi Berita Hoax di Indonesia. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 60–68. <https://doi.org/10.29408/edumatic.v9i1.28987>
- Faradillah, F., Alie, M. F., & Rahmanda, R. (2024). Model Prediksi Gangguan Tidur berdasarkan Beberapa Faktor menggunakan Machine Learning. *CESS (Journal of Computer Engineering, System and Science)*, 9(2), 504–517. <https://doi.org/10.24114/cess.v9i2.61068>
- Haikal, F. M., Akrom, M., & Trisnapradika, G. A. (2023). Perbandingan Algoritma Multilinear Regression dan Decision Tree Regressor dalam Memprediksi Efisiensi Penghambatan Korosi Piridazin. *Edumatic: Jurnal Pendidikan Informatika*, 7(2), 307-315. <https://doi.org/10.29408/edumatic.v7i2.22127>
- Hartini, S., Nisa, K., & Herini, E. S. (2021). Faktor-Faktor yang Berhubungan dengan Masalah Tidur Remaja Selama Pandemi Covid-19. *Sari Pediatri*, 22(5), 311–317. <https://doi.org/10.14238/sp22.5.2021.311-7>
- Hidayat, I. A. (2023). Classification of Sleep Disorders Using *Random Forest* on Sleep Health and Lifestyle Dataset. *Journal of Dinda : Data Science, Information Technology, and Data Analytics*, 3(2), 71–76. <https://doi.org/10.20895/dinda.v3i2.1215>
- Hutagalung, N. A., Marni, E., & Erianti, S. (2022). Faktor-Faktor Yang Mempengaruhi Kualitas Tidur Pada Mahasiswa Tingkat Satu Program Studi Keperawatan STIKes Hang

- Tuah Pekanbaru. *Jurnal Keperawatan Hang Tuah (Hang Tuah Nursing Journal)*, 2(1), 77–89. <https://doi.org/10.25311/jkh.Vol2.Iss1.535>
- Irfannandhy, R., Handoko, L. B., & Ariyanto, N. (2024). Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 714–723. <https://doi.org/10.29408/edumatic.v8i2.27990>
- Khasanah, N., Saputri, D. U. E., Aziz, F., & Hidayat, T. (2025). Studi Perbandingan Algoritma *Random Forest* dan K-Nearest Neighbors (KNN) dalam Klasifikasi Gangguan Tidur. *Computer Science (CO-SCIENCE)*, 5(1), 17–25. <https://doi.org/10.31294/coscience.v5i1.5522>
- Kivrak, M., Avci, U., Uzun, H., & Ardic, C. (2024). The Impact of the SMOTE Method on Machine Learning and Ensemble Learning Performance Results in Addressing Class Imbalance in Data Used for Predicting Total Testosterone Deficiency in Type 2 Diabetes Patients. *Diagnostics*, 14(23), 2–16. <https://doi.org/10.3390/diagnostics14232634>
- Larasati, S. S. A., Dewi, E. N. K., Farhansyah, B. H., Bachtiar, F. A., & Pradana, F. (2024). Penerapan Decision Tree dan *Random Forest* dalam Deteksi Tingkat Stres Manusia Berdasarkan Kondisi Tidur. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(5), 1043–1050. <https://doi.org/10.25126/jtiik.2024117993>
- Maulidah, M., & Hidayati, N. (2024). Prediksi Kesehatan Tidur Dan Gaya Hidup Menggunakan Machine Learning. *CONTEN: Computer and Network Technology*, 4(1), 81–86.
- Monarisa, M., Indriani, S., & Suci, H. (2021). Tingkat Kecemasan dan Kualitas Tidur Petugas Kesehatan Pada Masa Pandemi Covid-19. *JIK JURNAL ILMU KESEHATAN*, 5(2), 268–274. <https://doi.org/10.33757/jik.v5i2.439>
- Mufid, N. A., & Kismiantini, K. (2024). Analisis Kualitas Tidur Penduduk Usia Produktif di Indonesia dengan Model Regresi Logistik Ordinal. *Indonesian Journal of Applied Statistics*, 7(2), 98–111. <https://doi.org/10.13057/ijas.v7i2.86659>
- Pangestu, A. G., Winarno, S., Nugraha, A., & Muttaqin, A. N. I. (2025). DiabTrack: Sistem Prediksi Dini Diabetes Melitus Tipe 2 berbasis Web menggunakan Algoritma K-Nearest Neighbors. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 284–293. <https://doi.org/10.29408/edumatic.v9i1.29691>
- Rachmat, B. K., Suwarisman, A., Afriyanti, I., Wahyudi, A., & Saputra, D. D. (2023). Analisis Sentimen Complain dan Bukan Complain pada Twitter Telkomsel dengan SMOTE dan Naïve Bayes. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(1), 106–113. <https://doi.org/10.35870/jtik.v7i1.691>
- Rao, B. H., Srinadh, C. V., Prashanthi, Y. L., Karthik, I., & Kiran, T. P. (2025). Sleep Disorder Detection Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 13(4), 3668–3671. <https://doi.org/10.22214/ijraset.2025.68924>
- Rohmah, H. N. F. (2023). Pelatihan Teknik Peningkatan Kualitas Tidur Pada Kader Kesehatan Di Desa Bantarjaya. *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, 7(1), 414–421. <https://doi.org/10.31764/jpmb.v7i1.13630>
- Ruriyanty, R. N. R., Subhannur Rahman, Mohtar, M. S., & Mohammad Basit. (2023). Hubungan Tingkat Stres Dengan Kualitas Tidur Mahasiswa Kesehatan Tingkat Akhir Di Universitas Sari Mulia. *Jurnal Keperawatan Suaka Insan (JKSI)*, 8(2), 184–189. <https://doi.org/10.51143/jksi.v8i2.493>
- Sahri, R. P., & Hasibuan, M. T. D. (2023). Hubungan Kualitas Tidur dengan Prestasi Belajar pada Mahasiswa. *Jurnal Penelitian Perawat Profesional*, 5(2), 473–480. <https://doi.org/10.37287/jppp.v5i2.1459>

- Setyowati, A., Rahmayanti, D., & Agianto, A. (2023). “Sleep Hygiene” untuk Kesehatan Fisik dan Mental yang Berkualitas pada Masyarakat yang Tinggal di Area Lahan Basah. *Jurnal Pengabdian ILUNG (Inovasi Lahan Basah Unggul*, 3(Agustus), 137–145. <https://doi.org/10.20527/ilung.v3i1>
- Syahreza, A., Ningrum, N. K., & Syahrazy, M. A. (2024). Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, Support Vector Regression, dan XGBoost. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 526-534. <https://doi.org/10.29408/edumatic.v8i2.27640>
- Triginandri, R., & Subhiyakto, E. R. (2024). Deteksi Dini Cacar Monyet menggunakan Convolutional Neural Network (CNN) dalam Aplikasi Mobile. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 516-525. <https://doi.org/10.29408/edumatic.v8i2.27625>
- Wardani, I. A. K., & Tiastiningsih, N. N. (2023). Gangguan Tidur Pada Penderita Gangguan Afektif Bipolar. *Jurnal Hasil Penelitian Dan Pengembangan (JHPP)*, 1(3), 177–183. <https://doi.org/10.61116/jhpp.v1i3.158>