

Prediksi Kualitas Tidur: Pendekatan Machine Learning yang Mengintegrasikan Faktor Kesehatan dan Lingkungan

Jordy Lasmana Putra^{1*}, Wahyutama Fitri Hidayat²

¹Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri
Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar, Jakarta Timur, Indonesia

²Program Studi Sistem Informasi Kota Pontianak, Universitas Bina Sarana Informatika
Jl. Abdulrahman Saleh No.18.A, Pontianak, Indonesia

e-mail: jordy.jlp@nusamandiri.ac.id, wahyutama.wfh@bsi.ac.id

(*) Corresponding Author

Artikel Info : Diterima : 17-07-2024 | Direvisi : 24-07-2024 | Disetujui : 29-07-2024

Abstrak - Gangguan tidur sangat mempengaruhi kualitas tidur seseorang yang dapat menyebabkan masalah kesehatan yang serius, untuk kalangan lansia kualitas tidur yang buruk dapat menyebabkan tingkat harapan hidup menurun drastis. Permasalahan utama adalah kurangnya alat prediksi yang efektif untuk meningkatkan kualitas tidur lansia, ditambah dengan banyak sekali faktor-faktor yang dapat mempengaruhi kualitas tidur lansia, yang perlu dilakukan analisis dan prediksi agar dapat meningkatkan kualitas tidur yang baik. Penelitian ini bertujuan untuk mengembangkan dan menguji model prediksi kualitas tidur lansia dengan mengintegrasikan faktor kesehatan dan lingkungan menggunakan pendekatan machine learning, data set yang digunakan adalah data set baru yang ada di website kaggle.com yaitu data National Poll on Healthy Aging (NPHA) yang berisikan wawasan tentang isu kesehatan, perawatan kesehatan, dan kebijakan kesehatan yang memengaruhi orang Amerika berusia 50 tahun ke atas, hal ini dilakukan agar dapat memperbaiki kualitas tidur dikalangan lansia. Metode machine learning, yaitu deep learning dengan algoritma Random Forest digunakan pada penelitian ini dan menunjukkan hasil yang baik dengan nilai akurasi 94,00% dan 44,44% akurasi yang didapatkan dengan menggunakan data latih. Hasil penelitian ini diharapkan dapat memberikan alat prediksi yang dapat digunakan oleh praktisi kesehatan untuk meningkatkan kualitas tidur lansia, sehingga berdampak positif terhadap kesehatan dan harapan hidup mereka.

Kata Kunci : Kualitas Tidur, Machine Learning, Deep Learning, Random Forest, NPHA

Abstracts - Sleep disorders significantly affect an individual's sleep quality, which can lead to serious health problems. For the elderly, poor sleep quality can drastically reduce life expectancy. The main problem is the lack of effective predictive tools to improve sleep quality among the elderly, compounded by the numerous factors that can influence their sleep quality. Therefore, analysis and prediction are necessary to enhance sleep quality. This study aims to develop and test a predictive model for sleep quality in the elderly by integrating health and environmental factors using a machine learning approach. The dataset used is a new one available on the website Kaggle.com, namely the National Poll on Healthy Aging (NPHA) data, which provides insights into health issues, healthcare, and health policies affecting Americans aged 50 and above. The aim is to improve sleep quality among the elderly. A machine learning method, specifically deep learning with the Random Forest algorithm, was used in this study and showed good results with an accuracy rate of 94.00% and a training data accuracy of 44.44%. The results of this study are expected to provide a predictive tool that can be used by healthcare practitioners to improve the sleep quality of the elderly, thereby positively impacting their health and life expectancy.

Keywords Sleep Quality, Machine Learning, Deep Learning, Random Forest, NPHA.

PENDAHULUAN

Menjaga kualitas tidur merupakan hal penting dalam memelihara kesehatan manusia (Jasmine & Martdianty, 2022), kualitas tidur yang baik telah terbukti memiliki dampak positif terhadap kinerja fisik, kognitif, dan emosional seseorang (Mehta, 2022), sementara gangguan tidur atau tidur yang tidak memadai dapat mengarah pada masalah kesehatan serius seperti obesitas, penyakit jantung, dan gangguan mental (Aisyah & Wijayani, 2023)



(Lyons et al., 2022). Pada usia lansia masalah gangguan tidur menjadi persoalan yang serius yang dapat memicu penyakit-penyakit kronis (Li et al., 2024), yang dapat mempengaruhi tingkat harapan hidup lansia akibat adanya gangguan dan penurunan kualitas tidur (Yuliadarwati & Utami, 2022)

Penelitian untuk meningkatkan kualitas tidur lansia kerap dilakukan oleh peneliti di dunia, seperti penelitian yang pernah dilakukan oleh (Alghwiri et al., 2021) yang melakukan penilaian prevalensi gangguan tidur pada mahasiswa dan melakukan identifikasi faktor-faktor yang dapat mempengaruhi kualitas tidur, penelitian selanjutnya pernah dilakukan oleh (Arora et al., 2020) yang menggunakan model prediksi Deep Learning, yaitu Convolutional Neural Network (CNN) dan Multilayer Perceptron (MLP) untuk melakukan prediksi kualitas tidur berdasarkan sensor yang didapatkan dari perangkat smartwatch, gelang kebugaran, dan sepatu GPS dengan hasil prediksi tinggi 97,30% untuk CNN, penelitian serupa yang menggunakan model deep learning juga pernah dilakukan oleh (Phan et al., 2020) dengan hasil penelitian menyatakan model Deep Learning dapat melakukan prediksi kualitas tidur dengan baik, pada penelitian yang dilakukan oleh (Alshammari, 2024) dilakukan penelitian dengan fokus mengklasifikasi gangguan tidur dengan membandingkan beberapa algoritma deep learning, yaitu : k-nearest neighbours, support vector machine, decision tree, random forest and artificial neural network (ANN), penelitian lainnya mengembangkan arsitektur pembelajaran dalam hibrida untuk secara otomatis mengklasifikasikan tahap tidur menggunakan sinyal polysomnografi multimodal, guna meningkatkan evaluasi dan diagnosis kualitas tidur (Lin et al., 2024)

Meskipun pentingnya kualitas tidur telah diakui secara luas dengan mencermati dari penelitian-penelitian terdahulu, memahami faktor-faktor yang mempengaruhi kualitas tidur dan kemampuan untuk memprediksi kualitas tidur seseorang menjadi sebuah tantangan tersendiri yang dapat menghasilkan kepuasan dalam tidur (Nugraha et al., 2023). Dalam konteks ini, teknik-teknik machine learning, khususnya deep learning, menawarkan pendekatan yang unik untuk menganalisis data multidimensi yang kompleks dan mengidentifikasi pola yang tersembunyi (Nasir & Sassani, 2021)(Mahmud et al., 2021). Dengan menggunakan model-model ini, kita dapat mengintegrasikan data kesehatan dan lingkungan dari individu dan membangun sistem yang dapat memprediksi kualitas tidur, khususnya lansia. Perbedaan penelitian ini dengan penelitian sebelumnya yaitu menggunakan data baru yang belum dipublikasikan hasil penelitiannya sehingga nantinya dapat ditarik kesimpulan mengenai data yang digunakan maupun hasil prediksi tidur.

Penelitian ini bertujuan untuk mengisi kesenjangan ini dengan mengembangkan model prediktif menggunakan deep learning dengan algoritma random forest untuk memprediksi kualitas tidur berdasarkan data kesehatan fisik, kesehatan mental, dan faktor-faktor lingkungan. Metode Random Forest digunakan karena mampu memberikann akurasi yang baik (Renata & Ayub, 2020). Selain itu algoritma Random Forest, juga menggunakan parameter yang diujikan maupun disesuaikan sebagai optimasi kinerja (Hidayat et al., 2023). Diharapkan hasil penelitian ini dapat memberikan wawasan yang lebih dalam tentang hubungan antara faktor-faktor tersebut dengan kualitas tidur dan memberikan kontribusi pada pengembangan intervensi yang lebih efektif dalam meningkatkan tidur yang sehat dan berkualitas

METODE PENELITIAN

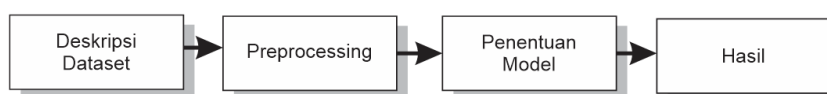
Penelitian ini merupakan proses prediksi kualitas tidur untuk mengklasifikasikan apakah kualitas tidur baik atau kurang baik serta menentukan hasil akurasi menggunakan metode Random Forest.

Subjek Penelitian

Subyek penelitian ini adalah menggunakan data hal-hal yang mempengaruhi kualitas tidur berdasarkan kunjungan ke dokter yang diambil dari National Poll on Healthy Aging (NPHA).

Tipe penelitian

Penelitian ini merupakan eksperimen menemukan model terbaik untuk prediksi kualitas tidur. Pengumpulan data untuk memperoleh sumber data yang digunakan adalah metode pengumpulan data publik dengan mengambil data pada National Poll on Healthy Aging (NPHA). Data yang sudah dikumpulkan diproses menggunakan Google Colabs dan bahasa Python. Untuk mempermudah alur penelitian dibuat alur penelitian yang digambarkan menggunakan Gambar 1.



Sumber : Hasil Penelitian (2024)

Gambar 1. Alur Penelitian

Tahapan dimulai dengan mendeskripsikan dataset yang digunakan hal tersebut bertujuan untuk mengetahui jenis serta label yang nantinya akan dilanjutkan pada tahap preprocessing data. Tahapan kedua yaitu melakukan *preprocessing* sehingga nantinya dataset yang digunakan tidak terdapat kerusakan atau ketidaksesuaian data. Penentuan akurasi digunakan model *machine learning* random forest sehingga menghasilkan dua akurasi berdasarkan data latih dan data uji. Hasil dari akurasi nantinya dapat digunakan sebagai tolak ukur untuk dibandingkan dengan model machine learning lainnya.

HASIL DAN PEMBAHASAN

Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini menggunakan dataset publik yang didapatkan dari website Kaggle.com. yang dapat diakses melalui laman : <https://www.kaggle.com/datasets/joebeachcapital/national-poll-on-healthy-aging-npha>

	Number of Doctors Visited	Age	Physical Health	Mental Health	Dental Health	Employment	Stress Keeps Patient from Sleeping	Medication Keeps Patient from Sleeping	Pain Keeps Patient from Sleeping	Bathroom Needs Keeps Patient from Sleeping	Unknown Keeps Patient from Sleeping	Trouble Sleeping	Prescription Sleep Medication	Race	Gender
count	714.000000	714.0	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000	714.000000
mean	2.112045	2.0	2.794118	1.988796	3.009804	2.805723	0.247899	0.056022	0.218487	0.504202	0.417367	2.407563	2.829132	1.425770	1.55042
std	0.683441	0.0	0.909939	0.939928	1.361117	0.586582	0.432096	0.230126	0.413510	0.500333	0.493470	0.670349	0.546767	1.003896	0.49780
min	1.000000	2.0	-1.000000	-1.000000	-1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-1.000000	-1.000000	1.000000	1.000000
25%	2.000000	2.0	2.000000	1.000000	2.000000	3.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.000000	3.000000	1.000000	1.000000
50%	2.000000	2.0	3.000000	2.000000	3.000000	3.000000	0.000000	0.000000	0.000000	1.000000	1.000000	3.000000	3.000000	1.000000	2.000000
75%	3.000000	2.0	3.000000	3.000000	4.000000	3.000000	0.000000	0.000000	0.000000	1.000000	1.000000	3.000000	3.000000	1.000000	2.000000
max	3.000000	2.0	5.000000	5.000000	6.000000	4.000000	1.000000	1.000000	1.000000	1.000000	1.000000	3.000000	3.000000	5.000000	2.000000

Sumber : (Arvidsson, n.d.)

Gambar 2. Deskripsi Dataset

Berdasarkan Gambar 2 Deskripsi Dataset yang digunakan menggunakan lima belas variabel yang terdiri dari variabel Number of Doctors Visited, Age, Physical Health, Mental Health, Dental Health, Stress Keeps Patient from Sleeping, Medication Keeps Patient from Sleeping, Bathroom Needs Keeps Patient from Sleeping, Pain Keeps Patient from Sleeping, Unknown Keeps Patient from Sleeping, Trouble Sleeping, Prescription Sleep Medication, Race, Gender.

Preprocessing

Preprocessing mengubah data awal mentah supaya menjadikan data yang lebih dan layak untuk diolah pada tahapan pemodelan (Alghifari & Juardi, 2021). Pada tahapan ini juga ditujukan untuk mengatasi masalah seperti data yang hilang, duplikasi, dan inkonsistensi. Tahapan ini juga dilakukan konversi data ke dalam format yang lebih cocok untuk analisis lebih lanjut.

```
datasets.isnull().sum()
Number of Doctors Visited    0
Age                          0
Physical Health              0
Mental Health                0
Dental Health                0
Employment                   0
Stress Keeps Patient from Sleeping    0
Medication Keeps Patient from Sleeping    0
Pain Keeps Patient from Sleeping    0
Bathroom Needs Keeps Patient from Sleeping    0
Unknown Keeps Patient from Sleeping    0
Trouble Sleeping             0
Prescription Sleep Medication    0
Race                         0
Gender                       0
dtype: int64
```

Sumber : Hasil Penelitian (2024)

Gambar 3. Pengecekan Nilai Kosong

Gambar 3 menjelaskan pengecekan apakah terdapat nilai kosong pada masing-masing nilai variabel. Berdasarkan Gambar 3 tidak ditemukan nilai kosong, sehingga tidak diperlukan penghapusan data kosong. Langkah berikutnya dilakukan pengecekan apakah terdapat nilai negatif dalam dataset dengan menggunakan perintah ditunjukkan dengan Gambar 4.

```

maindf.isnull().sum()

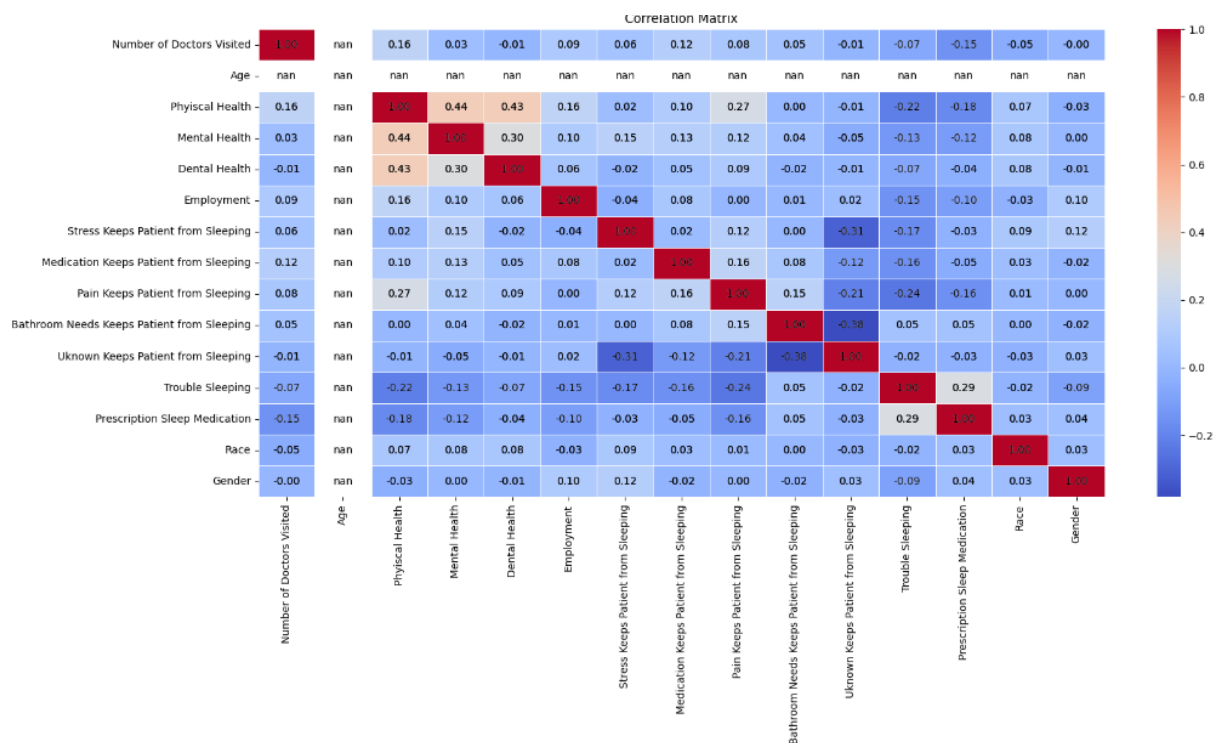
for _, col in enumerate(maindf.columns):
    count = maindf[col].isin([-1, -2]).sum()
    print(col, count)
    maindf = maindf[~maindf[col].isin([-1, -2])]
    
```

Number of Doctors Visited 0
 Age 0
 Physical Health 1
 Mental Health 10
 Dental Health 3
 Employment 0
 Stress Keeps Patient from Sleeping 0
 Medication Keeps Patient from Sleeping 0
 Pain Keeps Patient from Sleeping 0
 Bathroom Needs Keeps Patient from Sleeping 0
 Unknown Keeps Patient from Sleeping 0
 Trouble Sleeping 2
 Prescription Sleep Medication 2
 Race 0
 Gender 0

Sumber : Hasil Penelitian (2024)

Gambar 4. Penghapusan Data Negatif

Hasil dari proses penghapusan data negatif ditemukan 1 data dengan nilai negatif pada variabel Physical Health, 10 data pada variabel Mental Health, 3 data pada variabel Dental Health dan masing-masing 2 data pada variabel Trouble Sleeping dan Prescription Sleep Medication. Langkah selanjutnya adalah memeriksa korelasi antar variabel untuk menentukan apakah ada hubungan antara masing-masing pasangan variabel dan kekuatan korelasi, hasil uji korelasi yang ditunjukkan pada Gambar 5.



Sumber : Hasil Penelitian (2024)

Gambar 5. Hasil Uji Korelasi

Hasil uji korelasi menunjukkan bahwa variabel Age tidak memiliki korelasi sehingga perlu dilakukan penghapusan variabel dengan perintah Gambar 6. Setelah prapemrosesan, data dibagi menjadi dua bagian: data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi performa model, pada tahap ini rasio data latih dan data uji adalah 80:20.

```
[42] columns_to_drop = ['Age']
     X = X.drop(columns=columns_to_drop)

[43] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Sumber : Hasil Penelitian (2024)

Gambar 6. Hapus Variabel Age dan Bagi Data

Penentuan Model

Tahap ini ditunjukkan dengan Gambar 7, algoritma machine learning dipilih dan diterapkan pada data pelatihan, Penelitian ini, digunakan algoritma machine learning Random Forest untuk memodelkan data dan memprediksi kategori prediksi kualitas tidur.

```
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, BaggingClassifier
random_forest = RandomForestClassifier(n_estimators = 10)
random_forest.fit(X_train, y_train)
rf_predicted = random_forest.predict(X_test)
random_forest_score = round(random_forest.score(X_train, y_train) * 100, 2)
random_forest_score_test = round(random_forest.score(X_test, y_test) * 100, 2)
print('Random Forest Score: \n', random_forest_score)
print('Random Forest Test Score: \n', random_forest_score_test)
print('Accuracy: \n', accuracy_score(y_test, rf_predicted))
print(confusion_matrix(y_test, rf_predicted))
print(classification_report(y_test, rf_predicted))
```

Sumber : Hasil Penelitian (2024)

Gambar 7. Penentuan Model

Hasil

Hasil dari melatih model menggunakan Random Forest dihasilkan akurasi sebesar 94 % menggunakan data latih dan menghasilkan 44,44 % akurasi ditunjukkan dengan Gambar 8.

```
Random Forest Score:
94.0
Random Forest Test Score:
44.44
Accuracy:
0.4444444444444444
[[ 4 39 12]
 [23 98 18]
 [ 9 54 22]]
      precision    recall  f1-score   support

     1         0.11     0.07     0.09         55
     2         0.51     0.71     0.59        139
     3         0.42     0.26     0.32         85

 accuracy                   0.44         279
 macro avg              0.35     0.35     0.33         279
 weighted avg           0.41     0.44     0.41         279
```

Sumber : Hasil Penelitian (2024)

Gambar 8. Hasil Eksperimen

KESIMPULAN

Penelitian ini menggunakan data publik dari repository kaggle.com, dimana pada penelitian tersebut pada saat jurnal ini diterbitkan hasil penelitian menghasilkan akurasi sebesar 41%. Hasil dari pengembangan dalam penelitian ini untuk memaksimalkan menghasilkan model prediksi deep learning menggunakan algoritma Random Forest dengan akurasi 94,00% dengan menggunakan data latih dan 44,44% akurasi yang didapatkan dengan menggunakan data uji. Hasil ini dinilai baik dan menunjukkan model random forest ini dapat digunakan untuk melakukan prediksi kualitas tidur lansia dengan mempertimbangkan faktor-faktor kesehatan dan lingkungan.

Pengembangan berikutnya dapat berupa eksplorasi fitur-fitur spesifik yang signifikan dalam kontribusinya terkait peningkatan nilai akurasi, serta pengembangan metode-metode lainnya yang lebih canggih.

Peningkatan yang terjadi dari hasil pengembangan yang akan dilakukan, kedepannya perlu dilakukan juga evaluasi terhadap model prediksi yang sudah ada untuk lebih mengukur sejauh mana model prediksi yang sudah ada dapat diimplementasikan dengan baik. Penelitian ini memberikan wawasan dan juga kesempatan kepada semua peneliti di dunia, khususnya dibidang ilmu komputer dan kesehatan, agar dapat berkolaborasi dalam upaya pengembangan-pengembangan pada penelitian selanjutnya.

REFERENSI

- Aisyah, I. U., & Wijayani, Q. N. (2023). *PENGGUNAAN GADGET TERHADAP KUALITAS TIDUR DAN KESEHATAN MENTAL REMAJA*. 2(9), 31–40. <https://doi.org/https://doi.org/10.6578/triwikrama.v2i9.1421>
- Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>
- Alghwiri, A. A., Almomani, F., Alghwiri, A. A., & Whitney, S. L. (2021). Predictors of sleep quality among university students: the use of advanced machine learning techniques. *Sleep and Breathing*, 25(2), 1119–1126. <https://doi.org/10.1007/s11325-020-02150-w>
- Alshammari, T. S. (2024). Applying Machine Learning Algorithms for the Classification of Sleep Disorders. *IEEE Access*, 12. <https://doi.org/10.1109/ACCESS.2024.3374408>
- Arora, A., Chakraborty, P., & Bhatia, M. P. S. (2020). Analysis of Data from Wearable Sensors for Sleep Quality Estimation and Prediction Using Deep Learning. *Arabian Journal for Science and Engineering*, 45(12), 10793–10812. <https://doi.org/10.1007/s13369-020-04877-w>
- Arvidsson, J. (n.d.). *National Poll on Healthy Aging (NPHA)*.
- Hidayat, H., Sunyoto, A., & Al Fatta, H. (2023). Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 7(1), 31–40. <https://doi.org/10.47970/siskom-kb.v7i1.464>
- Jasmine, K., & Martdianty, F. (2022). The Analisis Pengaruh Abusive Supervision terhadap Employee Creativity dengan Sleep Deprivation, Emotional Exhaustion, dan Self Efficacy sebagai Mediator. *Jurnal Manajemen Dan Organisasi*, 13(1), 23–35. <https://doi.org/10.29244/jmo.v13i1.34343>
- Li, L., Li, X., Huang, Y., Li, H., Li, C., Ma, Y., Zhang, J., Peng, F., & Lyu, S. (2024). An RCT META analysis based on the effect of tai chi exercise therapy on the outcome of elderly patients with moderate-to-severe sleep disorders-A systematic review study. *Heliyon*, 10(2). <https://doi.org/10.1016/j.heliyon.2024.e24085>
- Lin, Y., Wang, M., Hu, F., Cheng, X., & Xu, J. (2024). Multimodal Polysomnography-Based Automatic Sleep Stage Classification via Multiview Fusion Network. *IEEE Transactions on Instrumentation and Measurement*, 73. <https://doi.org/10.1109/TIM.2023.3343781>
- Lyons, S., Strazdins, L., & Doan, T. (2022). Work intensity and workers' sleep: A case of working Australians. *Humanities and Social Sciences Communications*, 9(1). <https://doi.org/10.1057/s41599-022-01410-2>
- Mahmud, M., Kaiser, M. S., McGinnity, T. M., & Hussain, A. (2021). Deep Learning in Mining Biological Data. In *Cognitive Computation* (Vol. 13, Issue 1). Springer US. <https://doi.org/10.1007/s12559-020-09773-x>
- Mehta, K. J. (2022). Effect of sleep and mood on academic performance—at interface of physiology, psychology, and education. *Humanities and Social Sciences Communications*, 9(1), 1–13. <https://doi.org/10.1057/s41599-021-01031-1>
- Nasir, V., & Sassani, F. (2021). A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges. *International Journal of Advanced Manufacturing Technology*, 115(9–10), 2683–2709. <https://doi.org/10.1007/s00170-021-07325-7>
- Nugraha, F. A., Aisyiah, A., & Wowor, T. J. F. (2023). Analisis Faktor yang Berhubungan dengan Kualitas Tidur pada Remaja di SMPN 254 Jakarta. *Malahayati Nursing Journal*, 5(9). <https://doi.org/10.33024/mnj.v5i9.9242>
- Phan, D. Van, Chan, C. L., & Nguyen, D. K. (2020). Applying deep learning for prediction sleep quality from wearable data. *ACM International Conference Proceeding Series, class 1*, 51–55. <https://doi.org/10.1145/3418094.3418114>
- Renata, E., & Ayub, M. (2020). Penerapan Metode Random forest untuk Analisis Risiko pada dataset Peer to peer lending. *Jurnal Teknik Informatika Dan Sistem Informasi*, 6(3), 462–474. <https://doi.org/10.28932/jutisi.v6i3.2890>
- Yuliadarwati, N. M., & Utami, K. P. (2022). Pengaruh Latihan Relaksasi Otot Progresif bagi Lansia dengan Kualitas Tidur Buruk. *PengabdianMu: Jurnal Ilmiah Pengabdian Kepada Masyarakat*, 7(2), 319–323. <https://doi.org/10.33084/pengabdianmu.v7i2.2537>