



Prediksi Gangguan Tidur pada Sleep Health and Lifestyle Menggunakan Support Vector Machine dan Neural Network

^{1,*}Diana Sari

^{1,*}Prodi Teknik Informatika, Sekolah Tinggi Teknologi Bandung, Bandung, Indonesia

^{1,*}Coressponding auther e-mail: dianasari25jr@gmail.com

Abstrak

Tidur dan gaya hidup sehat berperan penting dalam menjaga keseimbangan fisik dan mental seseorang. Faktor yang dapat mempengaruhi gangguan tidur seseorang dapat disebabkan karena pekerjaan individu, durasi tidur, kualitas tidur, tingkat aktivitas, dan faktor gaya hidup lainnya. Tujuan dari penelitian ini menganalisis terjadinya gangguan tidur pada seseorang dengan melihat data faktor-faktor yang mempengaruhi. Maka dari itu diperlukannya suatu sistem yang dapat menganalisis terjadinya gangguan tidur pada seseorang. Penelitian ini mengelompokkan data menggunakan data mining dan membandingkan dua metode prediksi untuk memperkirakan faktor gangguan tidur berdasarkan dataset kesehatan tidur dan gaya hidup yang diimplementasikan pada aplikasi Orange. Dua metode penelitian ini menggunakan algoritma *Support Vector Machine* dan algoritma Neural Network, untuk mengembangkan model prediksi yang akurat. Berdasarkan analisis dari dua metode algoritma tersebut akurasi pada *Support Vector Machine* persentasenya 90.1% dan akurasi algoritma Neural Network persentasenya 91.2%. Hasil penelitian menunjukkan bahwa kedua model memiliki tingkat akurasi yang memuaskan dalam memprediksi variabel kesehatan tidur dan gaya hidup. Hasil dari perbandingan dua algoritma, Neural Network yang memiliki persentase yang lebih baik dari klasifikasi data kesehatan tidur dan gaya hidup yang memiliki akurasi AUC 0.907, CA 0.912, F1 0.912, Presisi 0.912, dan Recall 0.912.

Kata kunci : Gangguan Tidur, Support Vector Machine, Neural Network, Prediksi.

Abstract

Sleep and a healthy lifestyle play an important role in maintaining a person's physical and mental balance. Factors that can influence a person's sleep disorders can be caused by the individual's work, sleep duration, sleep quality, activity level, and other lifestyle factors. The aim of this research is to analyze the occurrence of sleep disorders in a person by looking at data on influencing factors. Therefore, we need a system that can analyze the occurrence of sleep disorders in a person. This research groups data using data mining and compares two prediction methods to estimate sleep disturbance factors based on sleep health and lifestyle datasets implemented in the orange application. These two research methods use the Support Vector Machine algorithm and the Neural Network algorithm, to develop accurate prediction models. Based on the analysis of the two algorithm methods, the accuracy of the Support Vector Machine percentage is 90.1% and the accuracy of the Neural Network algorithm is 91.2%. The results showed that both models had a satisfactory level of accuracy in predicting sleep health and lifestyle variables. The results of the comparison of two algorithms, Neural Network which has a better percentage of classification of sleep health and lifestyle data which has an accuracy of AUC 0.907, CA 0.912, F1 0.912, Precision 0.912, and Recall 0.912.

Keywords: Sleep Disorders, Support Vector Machine, Neural Network, Prediction.



Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

1. Pendahuluan

Gangguan pada tidur adalah sekelompok kondisi yang ditandai dengan adanya gangguan kuantitas, kualitas, atau durasi tidur [1]. Gangguan tidur merupakan masalah kesehatan yang signifikan dengan dampak yang merugikan pada kualitas hidup dan kesejahteraan individu. Faktor yang dapat mempengaruhi gangguan tidur adalah dari kualitas tidur, durasi tidur, tingkat stres, pekerjaan, tingkat aktivitas, dan faktor lainnya. Gangguan tidur pada data ini terdiri dari insomnia, sleep apnea, dan tidak ada gangguan tidur. Insomnia kondisi gangguan tidur dimana seseorang kesulitan untuk tidur. Sleep Apnea dimana kondisi klinis ditandai dengan adanya kesulitan bernafas saat tidur, sehingga bisa menurunkan kualitas tidur seseorang [2].

Penelitian ini memberikan wawasan mengenai jenis-jenis gangguan tidur, antara lain insomnia, sleep apnea, dan no sleep disorder. Pada penelitian ini prediksi gangguan tidur menggunakan algoritma SVM dan Neural Network aplikasi Orange. Data yang digunakan diambil dari repositori Kaggle. Dataset ini berisi informasi tentang identitas, jenis kelamin, usia, pekerjaan, waktu tidur, kualitas tidur, tingkat aktivitas fisik, tingkat stres, BMI, tekanan darah, detak jantung, jumlah langkah harian, dan gangguan tidur. Data targetnya akan diprediksi sebagai gangguan tidur.

SVM merupakan algoritma pembelajaran mesin [3], [4]. Dalam pembelajaran mesin, algoritma SVM dengan pembelajaran terkait untuk menganalisa data dan klasifikasi serta analisis regresi. SVM dicirikan oleh kemampuannya untuk meminimalkan konkurensi [5]. Neural Network sebagai fungsi matematika yang menggambarkan fungsi otak pada manusia untuk melakukan perhitungan paralel. NN melakukan perhitungan paralel dengan kompleks seperti prediksi, pengenalan pola, dan pemodelan. Proses yang terjadi mirip dengan perintah yang dikirimkan ke otak manusia, yaitu bertindak seperti dendrit, kemudian hasil dari perintah tersebut bertindak seperti akson dengan fungsi aktivasi [6].

Pada proses menganalisa data menggunakan data mining dimulai dari dataset sebagai data yang akan dianalisis, kemudian menggunakan algoritma SMV dan NN, dan mengevaluasi dengan confusion matrix dan analisis ROC.

2. Tinjauan Pustaka

2.1. Algoritma

Pertama kalinya algoritma ditemukan oleh seorang ilmuwan dari Persia, yaitu Abu Ja'fzr Mohammad bin Musa al-Khawarizmi. Algoritma adalah suatu

langkah- langkah, cara atau urutan untuk memecahkan suatu masalah [7].

2.2. Data Mining

Data Mining merupakan proses menganalisis data untuk menemukan dan merangkum informasi yang tidak diinginkan dan dapat dimengerti. Data mining juga sebuah proses mengekstraksi informasi dan pola dari sejumlah besar data tersebut. Data mining mencakup dari pengumpulan data, ekstraksi data, statistik data dan analisis data [8]. Data mining melibatkan kecerdasan buatan untuk mengidentifikasi pola yang dapat digunakan untuk membuat prediksi tentang perilaku masa depan. Analisis data mining melebihi apa yang dilakukan oleh sistem pendukung lainnya. Data mining tidak membutuhkan banyak waktu dan biaya yang murah. Data mining mencari informasi untuk memprediksi apa yang mungkin dilupakan atau tersembunyi pada data itu sendiri [9].

2.3. Support Vector Machine (SVM)

SVM dicirikan oleh kemampuannya dalam mengambil kumpulan data masukan dan membuat prediksi berdasarkan fitur-fiturnya [10]. SVM digunakan dalam klasifikasi linier atau non-linier. SVM nonlinier merupakan solusi permasalahan linier dengan mengimplementasikan fungsi kernel pada ruang fitur berdimensi tinggi [11].

Rumus ilmiah berikut digunakan untuk Prediction Evaluation dengan analisis AUC, CA, F1, Precision, dan Recall dengan masing-masing teknik:

(1)

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

(2)

$$\text{Specificity} = \frac{TN}{TN+F}$$

(3)

$$F1 = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

(4)

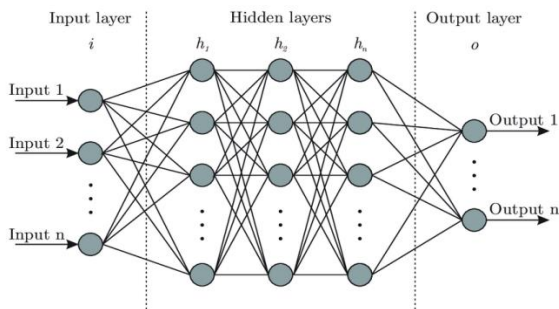
$$\text{Precision} = \frac{TP}{TP+FP}$$

(5)

$$\text{Recall} = \frac{TP}{TP+FN}$$

2.4. Neural Network

NN pertama kali ditemukan oleh McCulloch dan Pitts, sistem ini berkembang pesat dan banyak digunakan oleh banyak aplikasi, Jaringan Syaraf Tiruan merupakan jaringan yang digunakan untuk memodelkan perilaku sistem manusia. sistem saraf (otak) dengan melakukan tugas-tugas tertentu [12]. NN merupakan sistem pemrosesan informasi yang dapat mensimulasikan cara kerjanya otak manusia untuk memecahkan masalah-maslah dengan melakukan proses pembelajaran dengan mengubah bobot rangkuman [13].



Gambar 1. Algoritma Neural Network

2.5. Aplikasi Orange

Orange atau biasa dikenal dengan orange data mining merupakan aplikasi yang bersifat open source yang biasa digunakan untuk melakukan proses pengolahan data atau analytic [14].

3. Metode Penelitian

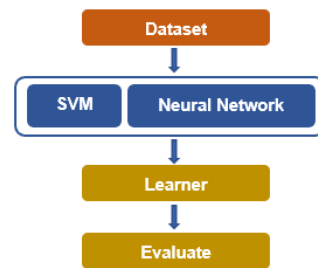
Penelitian ini melibatkan penyelidikan terhadap parameter dan variabel yang bergantung pada penelitian ini. Dataset Sleep Health and Lifestyle terdiri dari 374 baris dan 13 kolom, yang mencakup variabel terkait kesehatan tidur dan gaya hidup. Berdasarkan penelitian sebelumnya pada klasifikasi Sleep Health and Lifestyle menggunakan algoritma Random Forest sudah baik. Maka dari itu penelitian ini mencoba membandingkan hasil algoritma SVM dan neural network untuk menemukan algoritma terbaik untuk memprediksi gangguan tidur. Properti dataset dapat dilihat seperti pada Tabel 1:

Tabel 1. Tabel Atribut Dataset Sleep Health and Lifestyle

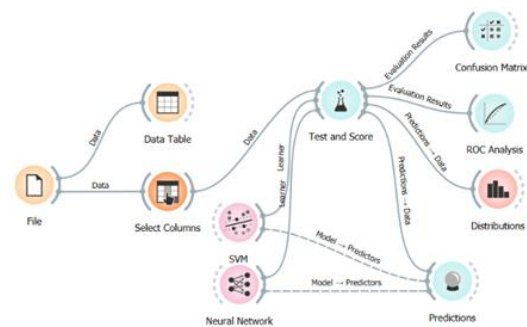
Atribut	Informasi Atribut	Keterangan
Id	Integer	Id Individu
Gender	Integer	Jenis kelamin (Male/Female)
Age	Integer	Usia dalam tahun
Occupation	Integer	Pekerjaan atau profesi orang
Sleep Duration	Integer	Jumlah jam tidur perhari
Quality of Sleep	Integer	Penilaian subjektif kualitas hidup (1-10)

Atribut	Informasi Atribut	Keterangan
Physical Activity Level	Integer	Jumlah menit aktivitas fisik setiap hari
Stress Level	Integer	Penilaian subjektif tingkat stress (1-10)
BMI Category	Integer	Kategori BMI (berat badan kurang, normal, kelebihan berat badan)
Blood Pressure	Integer	Pengukuran tekanan darah
Heart Rate	Integer	Denyut jantung (BPM) permenit
Daily Steps	Integer	Langkah perhari
Sleep Disorder	Integer	Target

Langkah selanjutnya adalah memilih model dan algoritma dari peneliti sebelumnya yaitu SVM dan NN. Kemudian evaluasi prediksi menggunakan AUC, CA, F1, recall presisi, matriks konfusi, dan analisis ROC. Proses ini menggunakan Data Mining dan diimplementasikan pada aplikasi Orange seperti terlihat pada Gambar 2 dan Gambar 3:



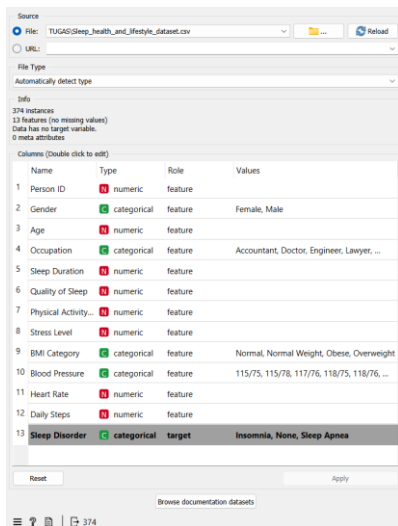
Gambar 2. Proses Data Mining



Gambar 3. Alur dari Proses Sleep Health and Lifestyle

4. Hasil dan Pembahasan

Metode yang digunakan dalam penelitian ini adalah SVM dan NN yang digunakan untuk mengembangkan sistem prediksi untuk menganalisis dan memprediksi kemungkinan terjadinya gangguan tidur. Dataset yang digunakan adalah Sleep Health and Lifestyle yang memiliki 13 atribut yang terdiri dari numeric dan categorical. Data target pada sleep disorder yang terlihat pada Gambar 4:

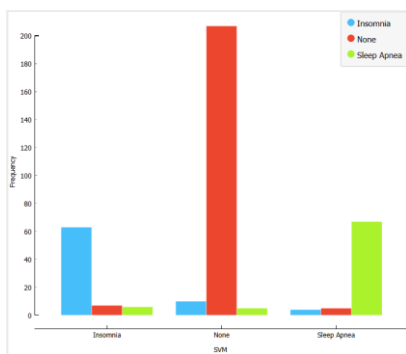


Gambar 4. Dataset yang digunakan

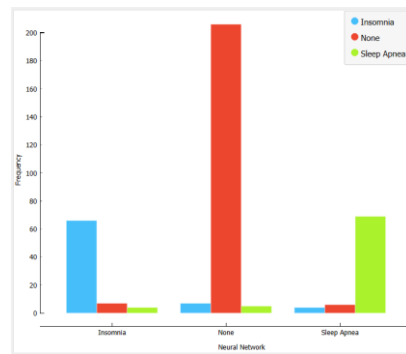
Sleep Disorder	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps
None	1	Male	29	Doctor	8.2	6	20	6	Normal	115/75	75	10000
None	2	Male	28	Doctor	8.2	6	20	6	Normal	115/75	75	10000
None	3	Male	28	Self-employed	5.9	4	30	6	Obese	140/90	85	3000
None	4	Male	28	Self-employed	5.9	4	30	6	Obese	140/90	85	3000
None	5	Male	28	Self-employed	5.9	4	30	6	Obese	140/90	85	3000
None	6	Male	28	Self-employed	5.9	4	30	6	Obese	140/90	85	3000
None	7	Male	29	Teacher	6.3	6	40	7	Obese	140/90	85	2000
None	8	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	9	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	10	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	11	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	12	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	13	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	14	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	15	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	16	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	17	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	18	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	19	Female	28	Teacher	6.5	5	40	7	Normal Weight	120/80	80	4000
None	20	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	21	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	22	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	23	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	24	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	25	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	26	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	27	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000
None	28	Male	29	Doctor	7.8	7	35	6	Normal	120/80	70	8000

Gambar 5. Features used untuk Prediksi Gangguan Tidur

Pada Gambar 5 terdapat 374 data, 12 features dan memiliki target data tiga value.

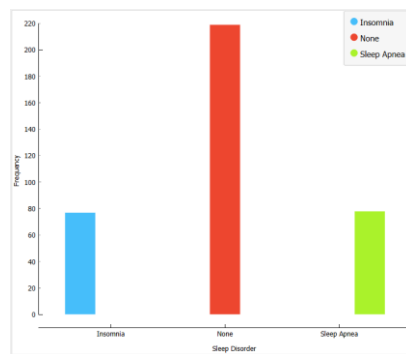


Gambar 6. Distribusi SVM



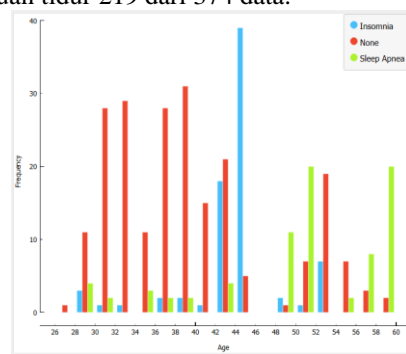
Gambar 7. Distribusi Neural Network

Gambar 6 dan Gambar 7 menunjukkan data distribusi dari SVM dan NN dengan target sleep disorder.



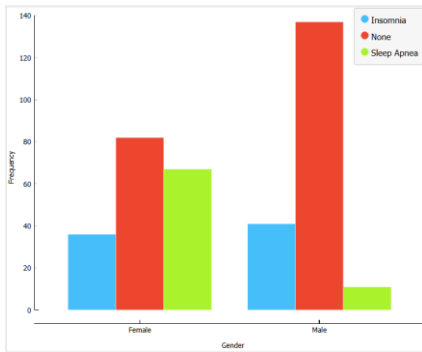
Gambar 8. Distribusi Sleep Disorder

Dapat dilihat pada Gambar 8 distribusi sleep disorder terdapat data yang memiliki gangguan tidur insomnia sebanyak 77, sleep apnea 78, dan tidak ada gangguan tidur 219 dari 374 data.



Gambar 9. Distribusi Usia

Pada Gambar 9 menunjukkan terjadinya gangguan tidur dan tidak terjadinya gangguan tidur berdasarkan usia.



Gambar 10. Distribusi Jenis Kelamin

Pada Gambar 10 perempuan lebih banyak mengalami gangguan tidur dari pada laki-laki. Diantara gangguan tidur yang terjadi pada perempuan sleep apne lebih banyak dibandingkan dengan insomnia.

Kemudian data diklasifikasikan menggunakan SVM dan NN kemudian melakukan validasi untuk melihat performa algoritma AUC, CA, F1, Precision, dan Recall pada masing-masing algoritma sehingga dapat diperoleh algoritma dengan akurasi terbaik. Berikut ini adalah matriks SVM dan NN.

		Predicted			Σ
		Insomnia	None	Sleep Apnea	
Actual	Insomnia	63	10	4	77
	None	7	207	5	219
	Sleep Apnea	6	5	67	78
	Σ	76	222	76	374

Gambar 11. Matriks SVM

SVM pada Gambar 5 berisi 63 + 207 + 67 prediksi benar dan 10 + 4 + 7 + 5 + 6 + 5 prediksi yang salah dari total 370 data pengujian. Dengan demikian kecuratan algoritma ini dapat dihitung sebagai $337/374 * 100\% = 90.1\%$.

		Predicted			Σ
		Insomnia	None	Sleep Apnea	
Actual	Insomnia	66	7	4	77
	None	7	206	6	219
	Sleep Apnea	4	5	69	78
	Σ	77	218	79	374

Gambar 12. Matriks Neural Network

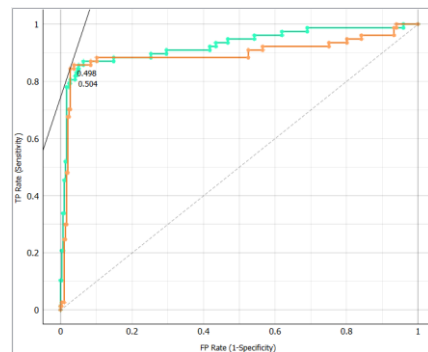
Neural Network pada Gambar 6 berisi 66 + 206 + 69 prediksi benar dan 7 + 4 + 7 + 6 + 4 + 5 prediksi salah dari total 374 data pengujian. Dengan

demikian kecuratan algoritma ini dapat di hitung sebagai $341/374 * 100\% = 91.2\%$.

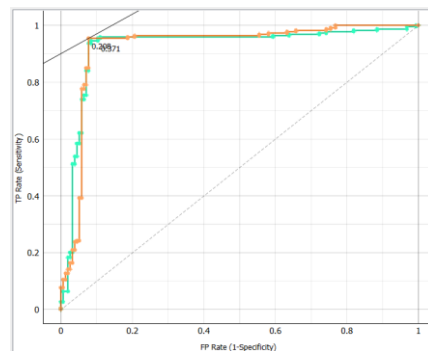
Tabel 2. Tabel Performa Algoritma

Model	AUC	CA	F1	Precision	Recall
SVM	0.923	0.901	0.901	0.901	0.901
Neural Network	0.907	0.912	0.912	0.912	0.912

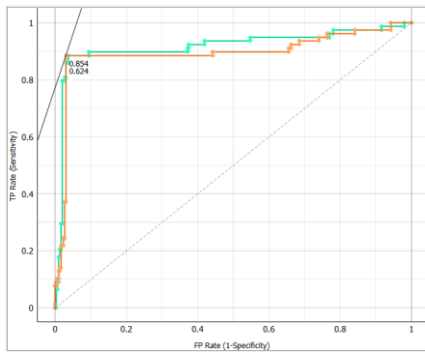
Pada Tabel 2 performa algoritma, ACU, CA, F1, Precision dan Recall adalah metode yang digunakan dalam pengujian data. CA bekerja karena kecuratan kumpulan data. Akurasi adalah kecuratan data yang memungkinkan terjadinya tiga kejadian yaitu None, Insomnia, dan Sleep Apnea. Pengingat fungsi pengukuran rasio. F1 adalah perbandingan antara recall dan presisi. AUC digunakan untuk menyatakan probabilitas. SVM memiliki akurasi AUC 0.923, CA 0.901, F1 0,901, Presisi 0.901, dan Recall 0.901. Sedangkan NN memiliki akurasi AUC 0.907, CA 0.912, F1 0.912, Presisi 0.912, dan Recall 0.912. Dapat dilihat dari akurasi tersebut NN memiliki tingkat akurasi yang lebih baik daripada SVM untuk data Sleep Health and Lifestyle.



Gambar 13. Analisis ROC Target Insomnia



Gambar 14. Analisis ROC Target None



Gambar 15. Analisis ROC Target Sleep Apnea

Pada Gambar 13, 14, dan 15, analisis ROC setiap metode disorot dengan kurva berwarna berbeda. SVM ditandai dengan warna hijau dan Neural Network ditandai dengan warna oranye. Untuk kedua metode dalam satu grafik, nilai sensitivitas dan spesifisitasnya tidak lebih besar dari 1, dan jika nilai targetnya berubah maka untuk semua metode, perubahan sensitivitas dan spesifisitasnya tidak akan terlihat jelas. Pada setiap kelas terlihat rata-rata kurva SVM berangsur-angsur mendekati sumbu Y.

5. Kesimpulan

Dari hasil penelitian ini dapat disimpulkan berdasarkan data gangguan tidur yang sudah dianalisis menggunakan algoritma SVM dan algoritma Neural Network dengan nilai akurasi 90.1% pada SVM dan 91.2% pada nilai akurasi Neural Network. Maka pada penelitian dataset Sleep Health and Lifestyle, algoritma Neural Network memiliki tingkat akurasi yang lebih baik dalam memprediksi data dibandingkan SVM dengan nilai keseluruhan dari hasil prediksi akurasi AUC 0.907, CA 0.912, F1 0.912, Presisi 0.912, dan Recall 0.912 yang lebih unggul dari algoritma SVM.

6. Daftar Rujukan

- [1] A. Haryono et al., "Prevalensi Gangguan Tidur pada Remaja Usia 12-15 Tahun di Sekolah Lanjutan Tingkat Pertama," *Sari Pediatri*, vol. 11, no. 3, 2016, doi: 10.14238/sp11.3.2009.149-54.
- [2] S. S. Azzahra, "Obstructive Sleep Apnea (OSA) Sebagai Faktor Resiko Hipertensi," *Jurnal Ilmiah Kesehatan Sandi Husada*, vol. 10, no. 2, 2019, doi: 10.35816/jiskh.v10i2.180.
- [3] Y. Wendong, Z. Lou, and J. Bo, "A multi-factor analysis model of quantitative investment based on GA and SVM," in 2017 2nd International Conference on Image, Vision and Computing, ICIVC 2017, 2017, doi: 10.1109/ICIVC.2017.7984734.
- [4] D. M. Abdullah and A. M. Abdulazeez, "Machine Learning Applications based on SVM Classification: A Review," *Qubahan Academic Journal*, vol. 1, no. 2, 2021, doi: 10.48161/qaj.v1n2a50.
- [5] S. Long, X. Huang, Z. Chen, S. Pardhan, D. Zheng, and F. Scalzo, "Automatic detection of hard exudates in color retinal images using dynamic threshold and SVM classification: Algorithm development and evaluation," *Biomed Res Int*, vol. 2019, 2019, doi: 10.1155/2019/3926930.
- [6] M. A. Setyadji, A. Faqih, and Y. A. Wijaya, "Peramalan Harga Komoditas Beras Di Kalimantan Timur Menggunakan Algoritma Neural Network," 2023.
- [7] Intana Purnama Sari, "Pengantar Algoritma Dan Pemrograman," *Osf Preprints*, vol. 4, no. 1, 2021.
- [8] M. Adjie Setyadj, A. Faqih, and Y. Arie Wijaya, "Peramalan Harga Komoditas Beras Di Kalimantan Timur Menggunakan Algoritma Neural Network," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, 2023, doi: 10.36040/jati.v7i1.6327.
- [9] R. Ridwan, H. Lubis, and P. Kustanto, "Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, p. 286, Apr. 2020, doi: 10.30865/mib.v4i2.2035.
- [10] A. Khaleel Faieq and M. M. Mijwil, "Prediction of heart diseases utilising support vector machine and artificial neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, 2022, doi: 10.11591/ijeecs.v26.i1.pp374-380.
- [11] M. M. Mutoffar, "Prediction of Heart Disease Using Random Forest Algorithm, Support Vector Machine and Neural Network," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 99, no. 1, pp. 1–1, 2099, doi: 10.12928/TELKOMNIKA.v99i1.paperID.
- [12] A. Lumute Unihehu and I. Suharjo, "Klasifikasi Jenis Ikan Berbasis Jaringan Saraf Tiruan Menggunakan Algoritma Principal Component Analysis (PCA)," vol. 7, no. 2, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [13] J. Veronika and A. Andri, "Penerapan Metode Algoritma Neural Network Untuk Memprediksi Penjualan Bahan Bakar Minyak," *Journal of Information Technology Ampera*, vol. 3, no. 2, 2022, doi: 10.51519/journalita.volume3.iss2.issue2.year2022.page235-243.
- [14] I. Indriyanti, N. Ichsan, H. Fatah, T. Wahyuni, and E. Ermawati, "Implementasi Orange Data

Mining Untuk Prediksi Harga Bitcoin,” Jurnal Responsif: Riset Sains dan Informatika, vol. 4, no. 2, 2022, doi: 10.51977/jti.v4i2.762