

KLASIFIKASI TINGKAT STRES BERDASARKAN FAKTOR GAYA HIDUP MENGGUNAKAN METODE *RANDOM FOREST*

Annisa Mei Galuh Ageng, Bondan Wahyu Pamekas, Hanifah Permatasari

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No.55, Tipes, Kec. Serengan, Kota Surakarta, Jawa Tengah - Indonesia 57154

220103006@mhs.udb.ac.id

ABSTRAK

Tingkat stres merupakan salah satu kondisi yang dapat dipengaruhi oleh berbagai faktor gaya hidup, seperti durasi tidur, kualitas tidur, dan aktivitas fisik. Namun, faktor-faktor tersebut masih sering dianalisis secara terpisah sehingga hubungan antarvariabel dalam menentukan tingkat stres belum tergambarkan secara menyeluruh. Penelitian ini bertujuan untuk mengklasifikasikan tingkat stres berdasarkan faktor gaya hidup menggunakan algoritma *Random Forest*. Penelitian memanfaatkan *Sleep Health and Lifestyle Dataset* yang tersedia secara publik. Tahapan yang dilakukan meliputi praproses data, pembagian dataset menjadi data latih dan data uji, pelatihan model menggunakan algoritma *Random Forest*, serta evaluasi performa model berdasarkan nilai akurasi. Hasil pengujian menunjukkan bahwa model mampu mencapai tingkat akurasi pada rentang 96% hingga 97%. Selain itu, hasil pengujian berulang menunjukkan nilai akurasi yang relatif konsisten sehingga mengindikasikan kestabilan model yang baik. Berdasarkan hasil penelitian, algoritma *Random Forest* dapat digunakan secara efektif untuk mengklasifikasikan tingkat stres berdasarkan faktor gaya hidup dengan performa yang tinggi dan stabil.

Kata kunci : *Random Forest, Klasifikasi, Tingkat Stres, Machine Learning, Gaya Hidup*

1. PENDAHULUAN

Stres adalah kondisi psikologis yang sering dialami seseorang sebagai respons terhadap tekanan dari berbagai aspek kehidupan. Dalam jangka panjang, dampak stres tidak hanya terbatas pada kesehatan mental, tetapi juga dapat memengaruhi kondisi fisik, kualitas tidur, serta produktivitas individu. Sejumlah penelitian menunjukkan bahwa aspek gaya hidup, seperti durasi tidur, kualitas tidur, dan aktivitas fisik, memiliki keterkaitan dengan tingkat stres seseorang.

Meskipun demikian, berbagai penelitian terdahulu masih cenderung menganalisis faktor-faktor tersebut secara terpisah sehingga belum mampu memberikan gambaran yang menyeluruh mengenai kondisi stres individu. Padahal, faktor-faktor gaya hidup memiliki keterkaitan dan dapat saling memengaruhi satu sama lain. Oleh sebab itu, diperlukan metode yang mampu mengolah berbagai variabel secara simultan agar hubungan antar faktor dapat dianalisis secara lebih komprehensif.

Perkembangan teknologi *machine learning* memberikan peluang untuk mengatasi permasalahan tersebut. Salah satu algoritma yang banyak digunakan dalam tugas klasifikasi adalah *Random Forest* karena mampu mengelola data multivariabel dan memiliki ketahanan yang baik terhadap *overfitting* melalui mekanisme *ensemble*. Karakteristik tersebut menjadikan *Random Forest* sesuai untuk digunakan dalam analisis tingkat stres yang dipengaruhi oleh berbagai faktor gaya hidup.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengklasifikasikan tingkat stres menggunakan algoritma *Random Forest* dengan memanfaatkan faktor-faktor gaya hidup sebagai variabel prediktor. Data penelitian diperoleh dari *Sleep*

Health and Lifestyle Dataset yang tersedia secara publik. Melalui penelitian ini diharapkan diperoleh model klasifikasi yang memiliki tingkat akurasi yang baik serta mampu memberikan gambaran mengenai hubungan antara gaya hidup dan tingkat stres.

2. TINJAUAN PUSTAKA

Klasifikasi tingkat stres telah mengalami kemajuan yang signifikan sebagaimana diuraikan oleh Widjaya F dkk [1], seiring dengan meningkatnya penerapan metode *machine learning* dalam sektor kesehatan. Berbagai algoritma seperti *Support Vector Machine (SVM)*, *K-Nearest Neighbor (KNN)*, dan *Random Forest* telah digunakan untuk mengidentifikasi kondisi stres berdasarkan data fisiologis maupun gaya hidup. Septyanto dkk [2] mendemonstrasikan bahwa algoritma klasifikasi meliputi *Support Vector Machine (SVM)*, *K-Nearest Neighbor (KNN)*, dan *Random Forest* yang diterapkan untuk mengidentifikasi kondisi stres berbasis data fisiologis dan gaya hidup, memperoleh performa klasifikasi yang lebih unggul dibandingkan pendekatan konvensional.

Dalam bidang klasifikasi data, algoritma *Random Forest* sering diterapkan sebab dapat menangani dataset dengan jumlah fitur yang cukup banyak, termasuk pola keterkaitan antarvariabel yang bersifat kompleks. Berdasarkan kajian Khasanah dkk [3] dan Putra Argadianata dkk [4], performa klasifikasi yang dihasilkan algoritma *Random Forest* dinilai lebih unggul dibandingkan pendekatan konvensional, terutama dari sisi akurasi prediksi.

Dalam konteks kesehatan mental, khususnya tingkat stres, beberapa penelitian menunjukkan bahwa faktor gaya hidup memiliki peran yang signifikan.

Penelitian yang dilakukan oleh Muhammad Yusuf Gunawan dkk [5] menyatakan bahwa kualitas tidur dan durasi tidur memiliki hubungan terhadap kondisi psikologis individu. Selain itu, Sultan dkk [6] juga menjelaskan bahwa aktivitas fisik yang baik dapat membantu menurunkan tingkat tekanan psikologis dan stres.

Rusliadi dkk [7] menunjukkan bahwa hubungan antara faktor gaya hidup dan tingkat stres tidak bersifat satu dimensi, melainkan bahwa faktor-faktor tersebut saling terkait dan saling memengaruhi. Keterkaitan antarberbagai faktor tersebut menuntut penggunaan metode analisis yang dapat memproses seluruh variabel secara simultan sehingga diperoleh hasil kajian yang lebih menyeluruh.

Berbagai penelitian sebelumnya menunjukkan bahwa algoritma *Random Forest* mampu memberikan hasil yang baik pada klasifikasi data di bidang kesehatan. Oktavia dkk [8] nilai akurasi yang tinggi pada klasifikasi data kesehatan menunjukkan bahwa algoritma *Random Forest* bekerja secara efektif. Lebih lanjut, Fakhri dkk [9] menegaskan bahwa metode ini juga mampu menganalisis faktor dominan yang memengaruhi suatu kondisi tertentu melalui pendekatan *feature importance*.

Berdasarkan sintesis dari berbagai penelitian yang telah dilakukan, Anissa dkk [10] menyatakan bahwa klasifikasi tingkat stres yang memanfaatkan faktor gaya hidup dapat dilakukan secara efektif dengan menerapkan algoritma *Random Forest*. Pendekatan tersebut dinilai relevan karena memberikan hasil prediksi yang memiliki akurasi tinggi serta kestabilan yang baik.

3. METODE PENELITIAN

Pendekatan *machine learning* diterapkan pada penelitian ini untuk menganalisis tingkat stres menggunakan algoritma *Random Forest* dan berbagai indikator gaya hidup. Pendekatan ini bertujuan mengidentifikasi keterhubungan antara aspek gaya hidup dengan kondisi stres yang dialami individu. *Random Forest* adalah metode yang didasarkan pada konsep pembelajaran ensemble, yang dikenal karena kinerjanya yang unggul dalam tugas klasifikasi serta kemampuannya yang efektif dalam meminimalkan *overfitting* pada kumpulan data yang kompleks [11], [12].

3.1. Data Collection

Pada tahap awal, penelitian diawali dengan pemanfaatan data yang bersumber dari *Sleep Health and Lifestyle Dataset* sebagai dataset utama. Dataset tersebut diperoleh dari sumber data publik dan digunakan sebagai dasar dalam proses analisis. Kumpulan data tersebut berisi informasi terkait gaya hidup dan kondisi kesehatan individu, yang kemudian digunakan sebagai dasar untuk mengklasifikasikan tingkat stres.

3.2. Data Preprocessing

Tahapan *preprocessing data* meliputi beberapa proses kritis, yaitu *data cleaning*, seleksi atribut, serta transformasi data. Tahapan pra-proses dilakukan dengan membersihkan data agar hanya atribut yang sesuai dengan tujuan penelitian yang digunakan. Beberapa atribut yang tidak berkaitan langsung, seperti *Person ID*, *Occupation*, *BMI Category*, *Blood Pressure*, *Heart Rate*, *Daily Steps*, dan *Sleep Disorder*, dikeluarkan dari dataset. Selanjutnya, data kategorikal diubah menjadi nilai numerik dengan metode *encoding* agar dapat diproses oleh model *machine learning* [13].

3.3. Data Splitting

Setelah data selesai diproses, dilakukan pembagian dataset menjadi data latih serta data uji dengan perbandingan 80% berbanding 20%. Bagian data latih digunakan dalam proses pembentukan model, sedangkan data pengujian digunakan untuk menilai kemampuan model dalam memprediksi data baru yang tidak disertakan pada proses pelatihan [14].

3.4. Model Training

Pelatihan model dilakukan dengan menerapkan algoritma *Random Forest* yang menghasilkan banyak *decision tree* dari variasi data pelatihan. Masing-masing *decision tree* memberikan hasil prediksi, kemudian seluruh prediksi tersebut digabungkan menggunakan mekanisme voting untuk menentukan output akhir. Cara ini dinilai mampu memperbaiki ketepatan prediksi dan menekan risiko *overfitting* [15].

3.5. Model Evaluation

Penilaian terhadap model dilakukan guna mengetahui tingkat kemampuan sistem dalam melakukan klasifikasi. Kinerja model diukur menggunakan beberapa indikator evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* [16]. Selain itu, *confusion matrix* digunakan untuk menampilkan distribusi hasil prediksi secara lebih rinci. Berbagai parameter evaluasi yang digunakan untuk mengukur performa model dihitung berdasarkan *confusion matrix*. Parameter evaluasi tersebut digunakan guna menggambarkan performa hasil klasifikasi model dari berbagai sisi.

Accuracy:

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

Precision:

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

Recall:

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

F1-Score:

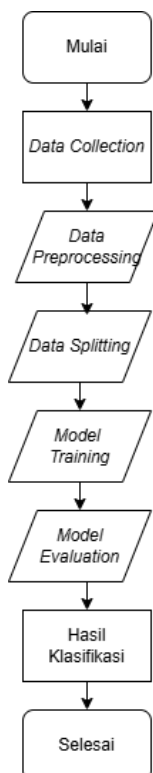
$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

True Positive (TP) menandakan jumlah data yang diprediksi secara akurat sebagai kelas positif, dimana *True Negative* (TN) merepresentasikan jumlah data yang diprediksi secara tepat sebagai kelas negatif. Sebaliknya, *False Positive* (FP) menunjukkan jumlah data yang diprediksi secara tidak tepat sebagai kelas positif, dan *False Negative* (FN) mengindikasikan jumlah data yang diprediksi secara keliru sebagai kelas negatif.

Guna mengevaluasi kestabilan model, dilakukan pengujian berulang dengan memvariasikan nilai *random_state*. Proses pengujian ini bertujuan untuk mengevaluasi konsistensi performa model ketika digunakan pada pembagian dataset yang bervariasi [17].

3.6. Alur Penelitian

Tahapan penelitian yang digunakan pada studi ini disajikan dalam bentuk *flowchart* untuk mempermudah pemahaman terhadap setiap proses yang dilakukan. Diagram tersebut menunjukkan urutan kegiatan penelitian, dimulai dari pengumpulan data, praproses data, pembagian dataset, pelatihan model dengan algoritma *Random Forest*, hingga tahap evaluasi model.



Gambar 1. Flowchart Metode Penelitian

Mengacu pada Gambar 1, langkah pertama penelitian adalah memperoleh dataset sebelum memasuki tahap pengolahan awal data. Setelah itu, dataset dipisahkan ke dalam dua kelompok, yaitu data pelatihan dan data pengujian. Model *Random Forest* dibentuk menggunakan data pelatihan, kemudian performanya dievaluasi melalui data pengujian. Evaluasi model menjadi proses terakhir yang

dilakukan untuk mengukur kemampuan klasifikasi yang dihasilkan.

4. HASIL DAN PEMBAHASAN

Pada studi ini, algoritma *Random Forest* digunakan untuk mengelompokkan tingkat stres berdasarkan berbagai faktor gaya hidup. Tahapan penelitian meliputi proses pengumpulan data, *preprocessing*, pembagian dataset, pelatihan model, serta evaluasi hasil klasifikasi.

Hasil dari penelitian ini disajikan sesuai dengan urutan tahapan yang telah dilakukan agar proses analisis data dan performa model dapat dijelaskan secara lebih terstruktur. Penyajian tersebut bertujuan untuk memudahkan pemahaman terhadap setiap proses yang dilakukan dalam penelitian.

4.1. Collecting Data

Penelitian dimulai dengan menggunakan *Sleep Health and Lifestyle Dataset* yang diperoleh melalui sumber data publik. Berbagai data mengenai gaya hidup serta kondisi kesehatan individu tersedia pada dataset tersebut dan dimanfaatkan dalam proses analisis tingkat stres.

Dalam penelitian ini, variabel *Stress Level* digunakan sebagai target klasifikasi untuk merepresentasikan tingkat stres pada setiap individu. Model *Random Forest* dilatih untuk memprediksi kategori tingkat stres berdasarkan sejumlah faktor gaya hidup yang tersedia pada dataset tersebut.

Tabel 1. Deskripsi Variabel Dataset

Variabel	Keterangan
<i>Age</i>	Usia individu
<i>Gender</i>	Jenis kelamin
<i>Sleep Duration</i>	Lama tidur perhari
<i>Quality of Sleep</i>	Tingkat kualitas tidur
<i>Physical Activity</i>	Tingkat aktivitas fisik
<i>Stress Level</i>	Tingkat stres

Mengacu pada Tabel 1, dataset yang digunakan memuat sejumlah atribut yang merepresentasikan aspek gaya hidup serta kondisi kesehatan individu. Atribut tersebut dimanfaatkan sebagai fitur untuk mendukung proses klasifikasi tingkat stres.

4.2. Data Preprocessing

Pada tahap *preprocessing*, data dibersihkan dan atribut diseleksi agar hanya variabel yang relevan dengan penelitian yang digunakan. Atribut yang tidak berkaitan langsung dengan analisis, seperti *Person ID*, *Occupation*, *BMI Category*, *Blood Pressure*, *Heart Rate*, *Daily Steps*, dan *Sleep Disorder*, dikeluarkan dari dataset.

	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level
0	Male	27	6.1	6	42	6
1	Male	28	6.2	6	60	8
2	Male	28	6.2	6	60	8
3	Male	28	5.9	4	30	8
4	Male	28	5.9	4	30	8

Gambar 2. Hasil *Preprocessing* Data

Berdasarkan Gambar 2, dapat dilihat bahwa beberapa atribut yang tidak digunakan dalam penelitian telah dihapus dari dataset. Data yang digunakan menjadi lebih sederhana sehingga mempermudah proses pengolahan dan pelatihan model.

Agar dapat diproses oleh algoritma *Random Forest*, data kategorikal terlebih dahulu diubah menjadi format numerik melalui teknik *encoding*. Pengolahan awal data dilakukan untuk memperbaiki kualitas dataset sehingga kinerja model menjadi lebih optimal. Adri Surya Kusuma [18] menjelaskan dalam penelitiannya bahwa prapemrosesan data memainkan peran penting dalam meningkatkan kinerja model *machine learning*.

4.3. Data Splitting

Setelah tahap *preprocessing* selesai dilakukan, dataset dipisahkan ke dalam data pelatihan dan data pengujian dengan rasio 80% dan 20%. Data latih digunakan dalam proses pembentukan model, sedangkan data uji dimanfaatkan untuk mengevaluasi performanya.

Tabel 2. Pembagian Data Training dan Testing

Jenis Data	Jumlah
Data Training	299
Data Testing	75

Berdasarkan Tabel 2, sebagian besar dataset dialokasikan untuk tahap pelatihan agar model dapat mengenali pola pada data secara optimal. Adapun sisa dataset dimanfaatkan pada proses pengujian guna menilai performa model yang telah dibuat.

4.4. Model Training

Pada tahap *preprocessing*, dilakukan pembersihan data disertai pemilihan atribut agar hanya variabel yang sesuai dengan tujuan penelitian yang digunakan. Sejumlah atribut yang tidak berkaitan langsung dengan analisis, seperti *Person ID*, *Occupation*, *BMI Category*, *Blood Pressure*, *Heart Rate*, *Daily Steps*, dan *Sleep Disorder*, dikeluarkan dari dataset.

	Gender	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level
0	Male	27	6.1	6	42	6
1	Male	28	6.2	6	60	8
2	Male	28	6.2	6	60	8
3	Male	28	5.9	4	30	8
4	Male	28	5.9	4	30	8

Gambar 3. Hasil *Preprocessing* Data

Berdasarkan Gambar 3, atribut-atribut yang tidak relevan dengan penelitian ini telah dihapus dari kumpulan data. Akibatnya, kumpulan data menjadi lebih ringkas, sehingga memudahkan tahap pemrosesan data dan pelatihan model.

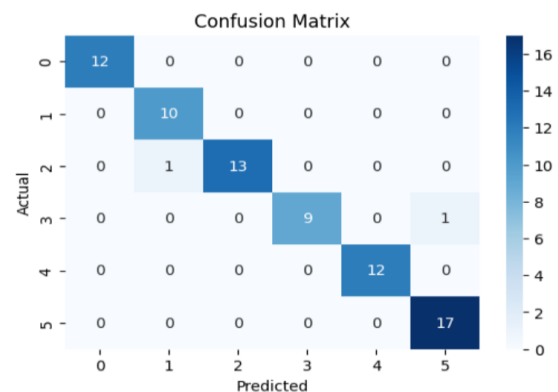
4.5. Model Evaluation

Beberapa indikator evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* digunakan untuk menilai kemampuan model. Di samping itu, *confusion matrix* dimanfaatkan guna menganalisis sebaran prediksi yang dihasilkan.

Tabel 3. Hasil Evaluasi Model *Random Forest*

Metrik	Nilai (%)
<i>Accuracy</i>	97.00
<i>Precision</i>	96.80
<i>Recall</i>	96.50
<i>F1-Score</i>	96.65

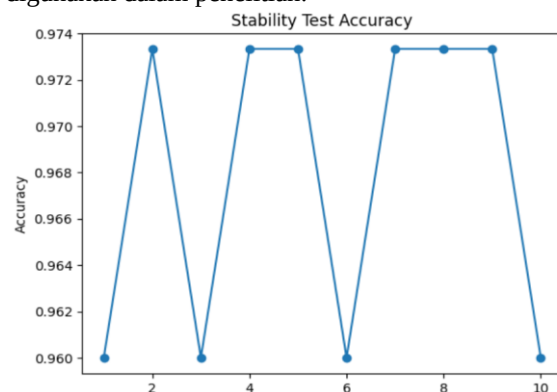
Berdasarkan Tabel 3, model *Random Forest* mampu menghasilkan performa klasifikasi yang sangat baik dengan nilai akurasi mencapai sekitar 97%. Hasil *precision*, *recall*, dan *F1-score* yang tinggi juga memperlihatkan bahwa model dapat mengelompokkan tingkat stres secara tepat.



Gambar 4. *Confusion Matrix*

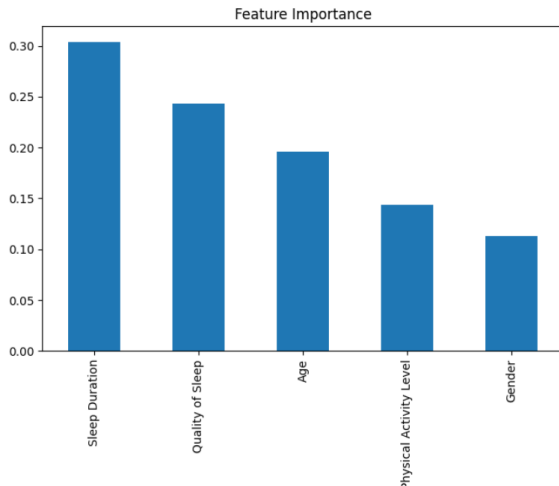
Pada Gambar 4 terlihat bahwa algoritma *Random Forest* mampu memprediksi sebagian besar data secara tepat. Angka yang terdapat pada diagonal *confusion matrix* menggambarkan hasil prediksi yang cocok dengan kelas asli data, sedangkan angka di luar diagonal menunjukkan kesalahan dalam proses klasifikasi.

Analisis *confusion matrix* mengindikasikan bahwa model dapat melakukan klasifikasi tingkat stres secara efektif berdasarkan aspek gaya hidup yang digunakan dalam penelitian.



Gambar 5. Grafik *Stability Test Accuracy*

Berdasarkan Gambar 5, dapat dilihat bahwa nilai akurasi model berada pada kisaran 96% hingga 97.3% pada setiap iterasi pengujian. Nilai yang mengalami perubahan hanya sedikit, sehingga menunjukkan bahwa model tetap konsisten terhadap variasi pada proses pembagian dataset.



Gambar 6. Feature Importance

Gambar 6 menunjukkan tingkat pengaruh masing-masing variabel terhadap proses klasifikasi tingkat stres. Beberapa variabel memiliki kontribusi yang lebih besar dibandingkan variabel lainnya, sehingga dapat diketahui bahwa faktor gaya hidup tertentu memiliki pengaruh yang signifikan terhadap tingkat stres individu.

4.6. Pembahasan

Berdasarkan tahap pengumpulan data, *Sleep Health and Lifestyle Dataset* dipilih karena memuat berbagai atribut yang sesuai dengan kebutuhan penelitian. Beberapa atribut yang digunakan meliputi durasi tidur, kualitas tidur, aktivitas fisik, serta tingkat stres sebagai variabel yang dianalisis.

Variabel-variabel ini dipilih untuk menggambarkan hubungan antara gaya hidup seseorang dan tingkat stresnya. Dengan memasukkan atribut target “*Stress Level*”, klasifikasi langsung dapat dilakukan untuk memprediksi tingkat stres berdasarkan data gaya hidup.

Pada tahap prapemrosesan, beberapa atribut dihilangkan sesuai dengan kebutuhan penelitian. Kolom-kolom seperti *Person ID* dan *Occupation* dikecualikan karena tidak memiliki dampak langsung terhadap klasifikasi tingkat stres. Selain itu, proses pengkodean untuk data kategorikal memungkinkan model menafsirkan semua variabel dalam format numerik, sehingga memastikan bahwa proses pelatihan dapat berjalan secara optimal. Proses ini menjadi tahap yang krusial, sebab kondisi data awal sangat menentukan tingkat akurasi hasil klasifikasi yang diperoleh.

Rasio 80:20 pada pembagian data training dan data testing dinilai telah memberikan distribusi data

yang cukup seimbang untuk kebutuhan penelitian. Sebagian besar dataset digunakan pada proses pelatihan agar model dapat mempelajari pola hubungan antarvariabel, sementara bagian data lainnya dimanfaatkan untuk mengevaluasi kemampuan model terhadap data baru. Dengan pembagian tersebut, hasil evaluasi menjadi lebih objektif karena pengujian dilakukan menggunakan data yang terpisah dari proses pelatihan.

Selama proses pelatihan, algoritma *Random Forest* berhasil menghasilkan model klasifikasi dengan kinerja yang optimal. Metode tersebut bekerja dengan membentuk beberapa *decision tree* menggunakan data pelatihan, kemudian seluruh hasil prediksi digabungkan untuk menentukan klasifikasi akhir. Mekanisme tersebut membuat model menjadi lebih konsisten karena hasil keputusan diperoleh dari kombinasi beberapa pohon, bukan hanya bergantung pada satu *decision tree*.

Hasil evaluasi memperlihatkan bahwa model yang diterapkan mampu menghasilkan kinerja klasifikasi dengan akurasi yang tinggi. Berdasarkan hasil analisis *confusion matrix*, mayoritas data dapat diprediksi dengan benar sesuai kelas sebenarnya sehingga model mampu mengidentifikasi tingkat stres dengan cukup efektif.

Selain itu, hasil analisis *feature importance* memperlihatkan bahwa variabel seperti kualitas tidur, lama tidur, dan aktivitas fisik memiliki pengaruh yang lebih dominan dibandingkan variabel lainnya. Temuan tersebut menunjukkan adanya keterkaitan antara pola gaya hidup dan tingkat stres individu, sehingga algoritma *Random Forest* dinilai tepat untuk diterapkan pada penelitian ini.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang diperoleh, algoritma *Random Forest* mampu mengidentifikasi tingkat stres melalui faktor-faktor gaya hidup dengan tingkat akurasi sekitar 96%–97%. Temuan ini mengindikasikan bahwa model yang dikembangkan memiliki kapabilitas yang efektif dalam mengenali korelasi antara variabel gaya hidup dan tingkat stres. Sebagai tambahan, pengujian mendemonstrasikan bahwa model menunjukkan tingkat kestabilan yang memadai, yang direpresentasikan oleh nilai akurasi yang relatif konsisten pada setiap iterasi. Kondisi tersebut menandakan bahwa performa model tidak banyak dipengaruhi oleh perbedaan pembagian dataset selama proses pengujian. Berdasarkan temuan tersebut, algoritma *Random Forest* terbukti efektif dan memiliki reliabilitas yang memadai dalam melakukan klasifikasi tingkat stres berbasis faktor gaya hidup. Untuk pengembangan selanjutnya, penelitian dapat dilakukan dengan memanfaatkan dataset yang lebih besar serta variabel yang lebih beragam agar hasil analisis menjadi lebih optimal. Sebagai pembanding, metode lain seperti *Support Vector Machine* atau *K-Nearest Neighbor* dapat diterapkan guna membandingkan performa model secara lebih

menyeluruh. Pengembangan lanjutan juga dapat dilakukan dengan mengintegrasikan model ke dalam sistem berbasis aplikasi, sehingga pengguna dapat memanfaatkannya secara langsung untuk memantau tingkat stres berdasarkan pola gaya hidup.

DAFTAR PUSTAKA

- [1] F. A. Widjaya, M. S. Ngete, and D. S. Tumanggor, "Perbandingan Algoritma Random Forest dengan K-Nearest Neighbor pada Klasifikasi Tingkat Stress Pelajar," 2025. [Online]. Available: <http://jurnal.mdp.ac.id>
- [2] A. W. Septyanto, H. Lugo Hariyanto, and H. Permatasari, "Utilizing The SVM Approach For Toddler Nutrition Status Classification Based On Anthropometric Measurements," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 4, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [3] N. Khasanah, D. Uki Eka Saputri, F. Aziz, T. Hidayat, U. Nusa Mandiri Jl, and C. Author, "Studi Perbandingan Algoritma Random Forest dan K-Nearest Neighbors (KNN) dalam Klasifikasi Gangguan Tidur," 2025. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/co-science>
- [4] A. Putra Argadianata, D. Abdul Fatah, H. Sukri, J. Raya Telang, K. Kamal, and J. Timur, "KLASIFIKASI KUALITAS BUAH APEL MENGGUNAKAN METODE RANDOM FOREST," 2025.
- [5] Muhammad Yusuf Gunawan, Naya Naisyia Sahda Aulia, Adystia Milkyna Puteri Gicella, and Sri Mulyeni, "Pengaruh Stres Akademik terhadap Kualitas Pola Tidur Mahasiswa," *JURNAL PENDIDIKAN DAN ILMU SOSIAL (JUPENDIS)*, vol. 4, no. 1, pp. 47–60, Jan. 2026, doi: 10.54066/jupendis.v4i1.3726.
- [6] U. Sultan, A. Tirtayasa, N. R. Haryana, R. Rosmiati, E. M. Purba, and H. Firmansyah, "Generation Z Lifestyle in Aspect of Eating Behavior, Stress, Sleep Quality and Its Relation to Nutritional Status: Literature Review," *Jurnal Gizi Kerja dan Produktivitas*, vol. 4, pp. 267–282, 2023.
- [7] Rusliadi Rusliadi and Mohamad Sofie, "Analisis Statistika Multivariat untuk Menilai Keterkaitan Gaya Hidup Digital dengan Tingkat Stres Mahasiswa," *Journal of New Trends in Sciences*, vol. 3, no. 1, pp. 46–59, Feb. 2025, doi: 10.59031/jnts.v3i1.750.
- [8] T. Oktavia, M. Ula, and A. Razi, "ANALISIS DATA MINING PERBANDINGAN ALGORITMA SUPPORT VEKTOR MACHINE DAN RANDOM FOREST PADA KLASIFIKASI SUBTIPE ANEMIA," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1405–1414, Jul. 2025, doi: 10.36341/rabit.v10i2.6565.
- [9] D. Fakhri, A. Ananda Sutrisno, N. Afza Zain, G. Angga Mukti, and A. Setiawan, "IMPLEMENTASI ALGORITMA MACHINE LEARNING MENGGUNAKAN MODEL RANDOM FOREST UNTUK KLASIFIKASI OBESITAS," 2025.
- [10] A. I. Anissa and A. Qoiriah, "Prediksi Tingkat Stres Berdasarkan Pola Hidup Menggunakan Machine Learning," *Journal of Informatics and Computer Science*, vol. 07, 2025.
- [11] I. Setiawan, I. Fatah Yasin, Y. Tri Desianti, P. Studi Sistem Dan Teknologi Informasi, F. Sains Dan Teknologi, and A. Surakarta, "Komparasi Kinerja Algoritma Random Forest, Decision Tree, Naïve Bayes, dan KNN dalam Prediksi Tingkat Depresi Mahasiswa Menggunakan Student Depression Dataset," 2025. [Online]. Available: <http://creativecommons.org/licenses/by/4.0/>
- [12] M. Hussein Umar, R. Daud Antony Pangaribuan, R. Primadana, I. Budiawan, and R. Pakpahan, "JURNAL MEDIA INFORMATIKA [JUMIN] Penerapan Algoritma Random Forest Untuk Prediksi Tingkat Stres Dari Aktivitas Media Sosial," 2025.
- [13] M. Maulidah and N. Hidayati, "PREDIKSI KESEHATAN TIDUR DAN GAYA HIDUP MENGGUNAKAN MACHINE LEARNING," *CONTEN : Computer and Network Technology*, vol. 4, no. 1, 2024, [Online]. Available: <http://jurnal.bsi.ac.id/index.php/conten>
- [14] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215–224, Feb. 2024, doi: 10.25126/jtiik.20241118074.
- [15] A. Rachmat Raharja, A. Pramudianto, and Y. Muchsam, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data 'Framingham' Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang," 2024.
- [16] S. Helmiyah and R. Pramestiawan, "Analisis Komparatif Algoritma Machine Learning dengan Metrik Akurasi, Presisi, Recall, dan F1-Score pada Dataset Kacang Kering," *Jurnal Ilmu Komputer dan Teknologi*, vol. 6, no. 3, pp. 152–159, Oct. 2025, doi: 10.35960/ikomti.v6i3.2031.
- [17] V. Yoga Pudya Ardhana, S. Lonang, D. Tejo Kumoro, M. Dermawan Mulyodiputro, and U. Qamarul Huda Badaruddin, "Benchmarking Model Machine Learning untuk Prediksi Data Berdasarkan Akurasi dan Error Benchmarking Machine Learning Models for Data Prediction Based on Accuracy and Error," 2025.
- [18] A. I. P. B. W. P. Adri Surya Kusuma1*, "Pengembangan Sistem Perhitungan Jumlah Kendaraan Berdasarkan Jenis Kendaraan Menggunakan Algoritma YOLO Secara Realtime," 2024