

PROPOSAL METODOLOGI PENELITIAN

**PREDIKSI TINGKAT STRESS REMAJA BERDASARKAN POLA
TIDUR MENGGUNAKAN ALGORITMA *RANDOM FOREST***



Disusun oleh:

Zsa Zsa Aulia Az Zahrah

09021382328136

Dosen Pengampu

Samsuryadi, M.Kom., Ph.D.

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2025

BAB I

PENDAHULUAN

1.1 Latar Belakang

Masa remaja (*adolescence*) adalah periode perkembangan yang sangat dinamis, namun juga penuh tantangan. Perubahan masif pada aspek biologis, sosial, dan kognitif membuat remaja menjadi kelompok yang rentan terhadap tingginya tingkat stres. Stres yang berkepanjangan pada remaja telah dikaitkan dengan konsekuensi negatif yang luas, termasuk penurunan kinerja kognitif, disregulasi emosi, dan peningkatan risiko masalah kesehatan mental yang serius (Stromájer et al., 2023). Urgensi masalah ini terlihat dari laporan global yang menunjukkan bahwa sekitar satu dari tujuh remaja di seluruh dunia diperkirakan mengalami kondisi kesehatan mental yang membutuhkan perhatian (WHO, 2025).

Kondisi stres ini memiliki korelasi kuat dengan faktor perilaku fundamental, yaitu pola tidur. Tidur yang cukup dan berkualitas adalah pilar bagi pemulihan neurokognitif dan stabilitas emosional. Sebaliknya, gangguan pada pola tidur, seperti kesulitan memulai atau mempertahankan tidur, secara signifikan meningkatkan kerentanan terhadap stres. Penelitian telah mendeskripsikan adanya mekanisme *sleep-stress spiral*, di mana kualitas tidur yang buruk secara timbal balik memperkuat persepsi stres yang dialami sehari-hari (ten Brink et al., 2021). Studi juga menggarisbawahi bahwa stres, khususnya yang berkaitan dengan tuntutan akademik, memiliki efek langsung yang merusak kualitas tidur remaja (Wang & Fan, 2023). Oleh karena itu, kemampuan untuk memprediksi tingkat stres remaja berdasarkan parameter pola tidur menawarkan peluang intervensi preventif yang efektif.

Dalam upaya membangun prediksi yang akurat, metode analisis yang digunakan harus mampu mengatasi kompleksitas data psikologis multidimensi. Metode statistik konvensional seringkali menghadapi kesulitan dalam memodelkan hubungan non-linear dan interaksi kompleks antar variabel. Kebutuhan akan alat yang lebih canggih dan efisien untuk memprediksi kondisi mental memicu adopsi teknologi Kecerdasan Buatan (Machine Learning) dalam bidang kesehatan.

Random Forest dipilih sebagai algoritma kunci untuk penelitian ini. Algoritma *ensemble learning* ini telah terbukti kuat, stabil, dan unggul dalam tugas klasifikasi data biomedis dan perilaku (Jiarui Liu, 2023). Keunggulan Random Forest, dalam menghasilkan prediksi yang akurat, mampu menangani banyak fitur, dan memitigasi risiko *overfitting*, mendukung pilihan ini (Astari et al., 2023). Namun, sebagian besar penelitian yang telah menerapkan Random Forest untuk prediksi stres dari pola tidur cenderung berfokus pada sinyal fisiologis atau total skor sederhana, sementara eksplorasi mendalam terhadap *sub-komponen* pola tidur (seperti durasi tidur, latensi, atau higiene tidur) pada populasi *remaja* masih terbatas.

Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut. Fokus utama penelitian ini adalah mengimplementasikan Random Forest secara komprehensif untuk tidak hanya memprediksi tingkat stres remaja dari pola tidur, tetapi juga melakukan analisis *Feature Importance*. Analisis ini akan mengidentifikasi secara spesifik sub-komponen pola tidur mana yang paling berpengaruh terhadap tingkat stres. Hasil penelitian ini diharapkan memberikan kontribusi metodologis dan praktis dalam pengembangan sistem deteksi dini dan intervensi kesehatan mental remaja.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini, yaitu:

1. Bagaimana gambaran pola tidur dan tingkat stres yang dialami oleh remaja?
2. Apakah terdapat hubungan yang signifikan antara pola tidur dengan tingkat stres?
3. Bagaimana penerapan algoritma Random Forest dalam memprediksi tingkat stres berdasarkan pola tidur remaja?
4. Variabel pola tidur manakah yang memiliki *Feature Importance* (kepentingan fitur) paling dominan dalam memprediksi tingkat stres remaja menggunakan algoritma Random Forest?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini ini, yaitu:

1. Menganalisis gambaran pola tidur dan tingkat stres yang dialami remaja.
2. Mengetahui hubungan antara pola tidur remaja dengan tingkat stres.
3. Menerapkan algoritma Random Forest untuk memprediksi tingkat stres remaja berdasarkan pola tidur.
4. Mengevaluasi hasil prediksi algoritma Random Forest terhadap data pola tidur remaja.
5. Mengidentifikasi *Feature Importance* untuk menentukan variabel pola tidur mana yang paling berpengaruh dalam model prediksi tingkat stres remaja.

1.4 Manfaat Penelitian

1.4.1 Manfaat Teoretis

Adapun manfaat teoretis dari penelitian ini, yaitu:

1. Memberikan kontribusi ilmiah dalam penerapan machine learning, khususnya algoritma Random Forest, untuk prediksi tingkat stres berdasarkan pola tidur.
2. Menjadi dasar bagi penelitian lanjutan dalam bidang kesehatan mental dan data science yang menghubungkan variabel gaya hidup dengan kondisi psikologis remaja.

1.4.2 Manfaat Praktis

Adapun manfaat praktis dari penelitian ini, yaitu:

1. Meningkatkan kesadaran para remaja terhadap pentingnya menjaga kualitas tidur sebagai faktor kunci dalam manajemen stres.

2. Menyediakan model prediktif berbasis teknologi yang dapat dimanfaatkan oleh institusi pendidikan, praktisi kesehatan, atau pengembang aplikasi digital dalam upaya skrining dini dan penanganan stres remaja.
3. Menjadi contoh penerapan teknologi kecerdasan buatan untuk mendukung kesehatan mental secara preventif.

1.5 Batasan Masalah

Penelitian ini memiliki batasan agar tujuan penelitian dapat tercapai secara terarah dan fokus, yaitu:

1. Penelitian menggunakan dataset sekunder yang berisi data hasil kuesioner dari populasi remaja (*adolescents*).
2. Variabel target yang diprediksi adalah tingkat stres yang diklasifikasikan menjadi kategori diskrit (rendah, sedang, tinggi) berdasarkan skor total *Perceived Stress Scale* (PSS).
3. Variabel prediktor utama dibatasi pada data yang berkaitan dengan pola tidur dan kebersihan tidur (termasuk skor total dan sub-komponen dari *Pittsburgh Sleep Quality Index* (PSQI), *Insomnia Severity Index* (ISI), dan *Adolescent Sleep Hygiene Scale* (ASHS)).
4. Algoritma *Machine Learning* yang digunakan untuk klasifikasi dan analisis *Feature Importance* hanya dibatasi pada algoritma Random Forest.
5. Penelitian ini berfokus pada klasifikasi dan analisis *Feature Importance*; penelitian ini tidak mencakup pengujian komparatif dengan model *Machine Learning* lain.

BAB II

TINJAUAN PUSTAKA

2.1 Konsep Stres

2.1.1 Definisi Stres

Stres merupakan fenomena yang bersifat universal. Stres didefinisikan sebagai respons non-spesifik tubuh terhadap tuntutan (*demands*) yang diberikan kepadanya, baik tuntutan yang bersifat fisik maupun psikologis (WHO, 2025). Definisi ini, yang awalnya dikembangkan oleh Hans Selye dalam *General Adaptation Syndrome* (GAS), melihat stres sebagai reaksi biologis universal tubuh terhadap perubahan. Reaksi ini melibatkan tiga tahap: *alarm*, *resistance*, dan *exhaustion*.

Dalam konteks psikologi kontemporer, stres juga dikategorikan menjadi dua jenis utama, yang mencerminkan sifat tuntutan:

1. *Eustress*: Stres positif atau bermanfaat (misalnya, stres menjelang pertandingan atau presentasi) yang berfungsi sebagai motivator.
2. *Distress*: Stres negatif yang bersifat merusak dan berkepanjangan, yang jika tidak ditangani dapat mengarah pada masalah kesehatan.

Penelitian ini berfokus pada bentuk *distress*, karena jenis stres ini berhubungan langsung dengan gangguan kesejahteraan psikologis. Stres psikologis muncul ketika tuntutan eksternal (misalnya beban akademik, tekanan sosial) atau tuntutan internal (misalnya standar diri yang tinggi) mengganggu keseimbangan (*homeostasis*) individu. Apabila stres tidak ditangani dengan cara yang sehat, kondisi ini dapat berlangsung kronis dan memicu dampak jangka panjang terhadap kesehatan mental maupun fisik.

2.1.2 Model Stres Transaksional

Dalam kajian psikologi kontemporer, stres dijelaskan melalui Model Stres Transaksional, yang memandang stres bukan sebagai respons otomatis terhadap stimulus eksternal, tetapi sebagai hasil interaksi dinamis antara

individu dan lingkungannya. Stres muncul ketika individu menilai bahwa tuntutan atau tekanan yang dihadapi melampaui kemampuan dirinya untuk mengelola atau mengatasinya. Proses ini berlangsung melalui mekanisme penilaian kognitif (*cognitive appraisal*) yang terdiri dari dua tahapan utama.

Tahap pertama adalah Penilaian Primer (*Primary Appraisal*), yaitu evaluasi awal individu terhadap relevansi dan makna suatu peristiwa. Pada tahap ini, situasi dapat diinterpretasikan sebagai ancaman, kerugian, atau tantangan. Tahap kedua adalah Penilaian Sekunder (*Secondary Appraisal*), yaitu proses evaluasi terhadap kemampuan dan sumber daya yang dimiliki untuk menghadapi situasi tersebut, termasuk dukungan sosial, keyakinan diri, dan strategi koping. Tingkat stres menjadi lebih tinggi apabila individu menilai bahwa sumber daya yang tersedia tidak memadai dibandingkan dengan tuntutan yang muncul.

Instrumen *Perceived Stress Scale (PSS)* dirancang untuk mengukur persepsi individu mengenai situasi kehidupan yang dianggap menekan, terutama terkait kegagalan dalam tahapan penilaian sekunder, yaitu ketika individu merasa situasi berada di luar kendali atau melebihi kapasitas dirinya.

2.1.3 Stres pada Remaja

Masa remaja (*adolescence*) merupakan fase perkembangan neurobiologis yang rentan. Pada periode ini, individu menghadapi peningkatan tekanan yang signifikan, baik yang bersifat internal maupun eksternal. Stres pada kelompok usia ini sering dikaitkan dengan *academic burnout* (kelelahan akademik), yang merupakan kelelahan emosional, sinisme, dan penurunan efikasi diri akibat tuntutan sekolah yang intensif (Gao, 2023). Stres kronis ini memicu *allostatic load* (beban sistemik) dan disfungsi regulasi emosi (Stromájer et al., 2023).

Secara kontekstual, studi literatur di Indonesia menggarisbawahi bahwa faktor eksternal seperti tekanan akademik, pola asuh, dan isu sosial-ekonomi adalah prediktor penting yang memengaruhi kesehatan mental remaja. Hal ini menunjukkan urgensi untuk mengidentifikasi pemicu stres utama pada populasi ini melalui analisis data (Wahyudi et al., 2025). Stres yang tidak dikelola pada

remaja berpotensi menimbulkan dampak negatif jangka panjang pada kesehatan dan perkembangan mereka.

2.1.4 Pengukuran Tingkat Stres (*Perceived Stress Scale*)

Pengukuran tingkat stres pada penelitian ini menggunakan *Perceived Stress Scale* versi 10 item (PSS-10). Instrumen ini merupakan alat ukur swalapor yang paling luas digunakan secara global dalam menilai persepsi stres sebagai variabel psikologis subjektif. PSS-10 dipilih karena memiliki karakteristik psikometri yang kuat, reliabilitas stabil, serta sensitivitas yang baik dalam mendeteksi variasi stres pada kelompok usia remaja.

Penggunaan PSS-10 pada populasi remaja telah divalidasi melalui beberapa studi. Temuan penelitian menunjukkan bahwa struktur faktor PSS-10 menunjukkan validitas konstruk yang memadai untuk digunakan pada kelompok usia ini (Liu et al., 2020). Selain itu, instrumen ini juga terbukti mampu mengidentifikasi hubungan antara tingkat stres dan berbagai indikator kesehatan fisik maupun mental pada remaja (Thorsen et al., 2022).

Dalam penelitian ini, skor total PSS-10 (rentang 0-40) diklasifikasikan menjadi tiga kategori berdasarkan panduan klinis dan temuan literatur sebelumnya, yaitu:

1. Stres Rendah: Skor 0 - 13
2. Stres Sedang: Skor 14 - 26
3. Stres Tinggi: Skor 27 - 40

Kategorisasi ini digunakan sebagai label output dalam proses klasifikasi *Machine Learning*, serta didukung oleh penelitian yang telah menetapkan batas skor tersebut sebagai cut-off yang layak secara empiris (Bio-protocol, 2021).

2.2 Konsep Pola Tidur

Pola tidur merupakan bentuk keteraturan dalam aktivitas tidur seseorang yang mencakup durasi, waktu mulai tidur, waktu bangun, kualitas tidur, serta konsistensi jadwal tidur setiap hari. Dalam konteks perkembangan manusia, tidur bukan hanya berfungsi sebagai waktu istirahat, tetapi merupakan proses biologis penting yang

berperan dalam pemulihan fisik, penguatan sistem imun, konsolidasi memori, dan pengaturan emosi (Uccella, 2023).

Pada masa remaja, kebutuhan tidur ideal adalah 8–10 jam per malam agar fungsi fisik dan mental dapat berjalan optimal (National Sleep Foundation, 2024). Namun, banyak remaja mengalami kurang tidur karena beban akademik, penggunaan gadget sebelum tidur, aktivitas sosial, serta perubahan ritme biologis yang membuat mereka cenderung tidur lebih larut (Johri et al., 2025).

Penelitian terdahulu menunjukkan bahwa pola tidur yang tidak teratur dan durasi tidur yang kurang dapat meningkatkan kelelahan di siang hari, kesulitan fokus, dan perubahan suasana hati. Kondisi tersebut juga dapat berdampak pada kesejahteraan psikologis dan meningkatkan risiko stres maupun gangguan emosional (Uccella, 2023). Dengan demikian, pola tidur bukan hanya soal lamanya tidur, tetapi juga konsistensi dan kualitas tidur yang dimiliki remaja.

2.2.1 Instrumen Pengukuran Pola Tidur

Pola tidur dalam penelitian psikologi dan kesehatan remaja umumnya diukur menggunakan instrumen terstandarisasi yang mampu menangkap variasi kualitas, durasi, dan kebiasaan tidur. Penggunaan instrumen ini tidak hanya memberikan gambaran subjektif mengenai pengalaman tidur remaja, tetapi juga memungkinkan integrasi data dalam analisis komputasi lanjutan seperti machine learning.

Instrumen yang banyak digunakan dalam penelitian pola tidur pada remaja meliputi:

1. *Pittsburgh Sleep Quality Index* (PSQI): Mengukur kualitas tidur melalui tujuh komponen, termasuk durasi, latensi, efisiensi, gangguan tidur, penggunaan obat tidur, disfungsi di siang hari, serta kualitas tidur secara keseluruhan (Chehri et al., 2020).
2. *Insomnia Severity Index* (ISI): Menilai tingkat keparahan insomnia berdasarkan kesulitan memulai tidur, mempertahankan tidur, dampak terhadap fungsi harian, dan persepsi individu mengenai gangguan tidur (Cerri, 2023).

3. *Adolescent Sleep Hygiene Scale* (ASHS): Mengukur kebiasaan tidur yang berkaitan dengan perilaku dan faktor lingkungan, seperti penggunaan perangkat elektronik sebelum tidur, paparan cahaya, konsumsi kafein, dan rutinitas tidur (Chehri et al., 2023).

Dalam konteks penelitian ini, instrumen-instrumen tersebut berperan sebagai sumber data yang dapat digunakan dalam pemodelan Random Forest untuk memprediksi tingkat stres pada remaja. Selain menghasilkan prediksi, variabel dari instrumen ini juga akan dianalisis menggunakan feature importance untuk mengidentifikasi komponen pola tidur yang paling berpengaruh terhadap tingkat stres.

2.3 Hubungan Pola Tidur dan Stres pada Remaja

Tidur memiliki peran penting dalam perkembangan fisiologis dan psikologis remaja. Periode ini ditandai dengan perubahan neurologis yang kompleks, terutama pada area korteks prefrontal dan sistem limbik, yang berperan dalam pengaturan emosi, perilaku, dan respons stres. Ketika kualitas atau durasi tidur terganggu, fungsi regulasi emosi menjadi kurang optimal sehingga remaja lebih rentan mengalami peningkatan tingkat stres (Kim, 2022).

Studi empiris menunjukkan adanya korelasi signifikan antara pola tidur yang buruk dan peningkatan stres pada remaja. Remaja dengan kualitas tidur rendah cenderung mengalami peningkatan iritabilitas, kelelahan kognitif, dan penurunan kemampuan akademik yang kemudian meningkatkan tekanan psikologis (Pradnyaparamita et al., 2023). Selain itu, perilaku seperti penggunaan gadget sebelum tidur dan pola tidur yang tidak teratur dapat meningkatkan tingkat stres dengan memengaruhi ritme sirkadian.

Temuan longitudinal terbaru juga menunjukkan adanya hubungan dua arah (bidirectional). Tidur yang kurang dapat meningkatkan persepsi stres, sementara stres yang tinggi memperburuk gangguan tidur seperti insomnia, kesulitan memulai tidur, dan fragmentasi tidur (Johri et al., 2025). Pola hubungan ini dikenal sebagai *sleep-stress feedback loop*, dan menjadi dasar pendekatan prediktif berbasis data karena melibatkan banyak variabel multidimensi.

Dalam konteks penelitian ini, hubungan antara pola tidur dan stres tidak hanya dianalisis secara deskriptif, tetapi juga dimodelkan menggunakan algoritma Random Forest. Dengan memanfaatkan sub-komponen instrumen seperti PSQI, ISI, dan ASHS sebagai variabel input, model memungkinkan identifikasi fitur yang paling berkontribusi terhadap prediksi tingkat stres melalui analisis feature importance. Pendekatan ini dapat memberikan pemahaman lebih spesifik mengenai faktor tidur mana yang memiliki relevansi terbesar pada kondisi stres remaja secara empiris.

2.4 Algoritma Random Forest

Random Forest adalah algoritma ensemble learning yang menggunakan banyak pohon keputusan (decision tree) untuk melakukan klasifikasi atau regresi. Algoritma ini diperkenalkan oleh Breiman (2001) dan dikenal karena kemampuannya menangani dataset besar, fitur yang banyak, serta mengurangi risiko overfitting dibandingkan pohon keputusan tunggal. Random Forest bekerja dengan membangun beberapa pohon keputusan secara independen menggunakan sampling acak data (bootstrapping) dan subset acak fitur untuk setiap split. Prediksi akhir dihasilkan dari voting mayoritas (untuk klasifikasi) atau rata-rata (untuk regresi), sehingga model menjadi lebih stabil dan akurat (Breiman, 2001; Biau, 2010).

Random Forest telah diterapkan luas di berbagai bidang, mulai dari analisis lingkungan, kesehatan, hingga ekonomi dan ilmu komputer. Penelitian terbaru menunjukkan efektivitas Random Forest dalam menangani dataset yang kompleks, termasuk data tidak seimbang dan data dengan noise tinggi, sambil tetap mempertahankan kemampuan interpretasi model menggunakan teknik seperti SHAP (Shapley Additive Explanations) atau partial dependence plots (Simon et al., 2023; Dube & Verster, 2024; Maças, 2024).

2.4.1 Cara Kerja *Random Forest*

Proses kerja Random Forest secara umum meliputi beberapa langkah berikut:

1. **Bootstrapping data:** Random Forest memilih sampel acak dari dataset asli dengan pengembalian (sampling with replacement) untuk membangun setiap pohon keputusan. Setiap pohon memiliki subset data yang berbeda, sehingga menghasilkan variasi antar pohon.

2. Pemilihan subset fitur secara acak: Untuk setiap split node dalam pohon, algoritma memilih subset acak dari fitur untuk menentukan split terbaik, bukan menggunakan semua fitur. Ini membantu mengurangi korelasi antar pohon dan meningkatkan generalisasi model.
3. Pembangunan pohon keputusan: Setiap pohon dibangun hingga mencapai kriteria berhenti tertentu, misalnya kedalaman maksimum atau jumlah sampel minimum per leaf node. Pohon dalam Random Forest biasanya tidak dipruning, agar setiap pohon tetap kompleks dan bervariasi.
4. Prediksi akhir: Setelah semua pohon terbentuk, prediksi untuk instance baru ditentukan dengan voting mayoritas (klasifikasi) atau rata-rata nilai output (regresi) dari semua pohon. Metode ini mengurangi variansi dan membuat model lebih stabil dibandingkan pohon tunggal.
5. Interpretabilitas model: Meskipun Random Forest dianggap “black-box”, teknik interpretasi seperti feature importance, SHAP, dan partial dependence plots dapat digunakan untuk memahami pengaruh tiap fitur terhadap prediksi (Xiao et al., 2024; Jurczuk et al., 2023).

2.4.2 Feature Importance

Feature importance adalah ukuran yang menunjukkan seberapa besar kontribusi masing-masing fitur terhadap hasil prediksi model *Random Forest*. Dengan menghitung feature importance, peneliti dapat mengidentifikasi fitur-fitur yang paling berpengaruh, meningkatkan interpretabilitas model, dan memberikan insight untuk pengambilan keputusan atau perbaikan dataset.

Dalam *Random Forest*, *feature importance* dapat dihitung dengan beberapa metode:

1. *Mean Decrease Impurity* (MDI)
Mengukur penurunan impuritas (seperti Gini atau entropy) pada setiap node split di semua pohon dalam hutan. Fitur yang sering dipakai untuk split dan menghasilkan penurunan impuritas besar dianggap lebih penting.

2. *Mean Decrease Accuracy* (MDA)
Mengukur penurunan akurasi model ketika nilai fitur tertentu diacak. Semakin besar penurunan akurasi, semakin penting fitur tersebut bagi model.
3. *Shapley Additive Explanations* (SHAP)
Metode berbasis teori permainan yang menghitung kontribusi tiap fitur terhadap prediksi secara individual dan global, sehingga interpretasinya lebih akurat dan konsisten, terutama untuk dataset yang kompleks atau tidak seimbang.

Beberapa penelitian terbaru menunjukkan bahwa analisis feature importance membantu memahami model *Random Forest* dalam konteks nyata, misalnya dalam klasifikasi data lingkungan, kesehatan, dan bencana alam (Simon et al., 2023; Dube & Verster, 2024; Xiao et al., 2024; Maças, 2024). Dengan demikian, *feature importance* tidak hanya meningkatkan interpretabilitas, tetapi juga dapat menjadi alat untuk seleksi fitur dan optimasi model.

2.5 Penelitian Terdahulu

Studi-studi mengenai hubungan antara pola tidur dan stres telah dilakukan dengan berbagai pendekatan, mulai dari analisis statistik hingga pemodelan *Machine Learning*. Kajian penelitian terdahulu ini bertujuan untuk memosisikan kontribusi unik dari penelitian ini.

1. Analisis Temporal Stres dan Tidur pada Remaja

Penelitian oleh ten Brink et al. (2021), yang berfokus pada dinamika harian stres dan tidur pada remaja, menggunakan *Dynamic Structural Equation Modeling (DSEM)* untuk memvalidasi fenomena *sleep-stress spiral*. Hasil studi ini menunjukkan bahwa peningkatan stres pada suatu hari memperburuk kualitas tidur pada malam hari, yang kemudian memperkuat tingkat stres keesokan harinya. Meskipun penelitian ini sangat relevan untuk konteks remaja, metodenya bersifat analisis korelasi dan mediasi statistik dan tidak menghasilkan model klasifikasi prediktif berbasis *Machine Learning* yang dapat diimplementasikan untuk interpretasi fitur.

2. Prediksi Stres menggunakan Algoritma *Machine Learning*

Penelitian oleh Astari et al. (2023) menerapkan algoritma Random Forest (dibandingkan dengan Decision Tree) untuk klasifikasi tingkat stres. Model yang dibangun menghasilkan akurasi yang tinggi, menegaskan keunggulan RF dalam menangani data kesehatan yang kompleks. Namun, studi ini difokuskan pada populasi dewasa/umum dan tidak secara spesifik mengeksplorasi kontribusi multidimensi dari sub-komponen pola tidur (seperti Latensi, Durasi, atau Kebersihan Tidur) pada remaja. Model tersebut lebih ditujukan pada akurasi klasifikasi daripada interpretasi mendalam.

3. Keterkaitan Stres, *Burnout*, dan Tidur dengan Metode SEM

Studi oleh Wang & Fan (2023) menyelidiki hubungan antara stres akademik, *burnout*, dan kualitas tidur pada remaja menggunakan *Structural Equation Modeling* (SEM). Temuan mereka menunjukkan bahwa *academic burnout* dan kecemasan adalah mediator yang signifikan antara stres akademik dan kualitas tidur yang buruk. Penelitian ini menguatkan urgensi topik pada remaja, tetapi analisisnya terbatas pada pembuktian hubungan statistik dan belum memanfaatkan potensi *Machine Learning* untuk prediksi multikelas dan interpretasi fitur.

4. Aplikasi *Random Forest* untuk *Feature Importance*

Studi oleh ResearchGate (2024) mengenai prediksi stres menggunakan Random Forest menunjukkan bahwa algoritma ini mampu mencapai akurasi tinggi dan juga menyediakan informasi *Feature Importance*. Penelitian ini memvalidasi bahwa analisis *Feature Importance* adalah alat yang efektif untuk mengidentifikasi variabel paling berpengaruh. Meskipun demikian, studi aplikasi *Feature Importance* yang komprehensif pada data psikometri multidimensi (PSQI/ISI/ASHS) dan berfokus pada populasi Remaja masih terbatas.

BAB III

METODOLOGI PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini menggunakan metode penelitian kuantitatif dengan pendekatan prediktif berbasis pembelajaran mesin (Machine Learning). Pendekatan ini dipilih karena penelitian difokuskan pada pembangunan model klasifikasi yang mampu memprediksi kategori tingkat stres remaja (Rendah, Sedang, atau Tinggi) berdasarkan data fitur pola tidur. Model yang digunakan adalah Random Forest Classifier, yang merupakan algoritma ensemble learning dengan kemampuan menangani struktur data yang kompleks serta hubungan antar variabel yang bersifat non-linear. Hal ini relevan mengingat indikator psikologi, termasuk gejala stres, sering kali tidak mengikuti pola linier sederhana dan dipengaruhi oleh banyak faktor yang saling berinteraksi.

Penggunaan pendekatan Machine Learning memberikan keunggulan dalam aspek akurasi prediksi dan kemampuan generalisasi terhadap data baru. Selain itu, algoritma Random Forest memiliki ketahanan terhadap overfitting, mampu menangani data dengan jumlah fitur yang beragam, dan secara otomatis melakukan estimasi variabel yang paling berkontribusi dalam proses klasifikasi. Karena itu, metode ini sesuai untuk menganalisis hubungan antara variabel pola tidur, seperti durasi tidur, kualitas tidur, jam mulai tidur, hingga tingkat insomnia, dengan tingkat stres remaja.

Penelitian ini juga bersifat analitis interpretatif karena tidak hanya berfokus pada performa prediktif model, tetapi juga pada interpretasi hasil melalui analisis Feature Importance. Analisis ini membantu mengidentifikasi fitur mana yang memiliki kontribusi paling besar terhadap prediksi tingkat stres, sehingga temuan tidak hanya menghasilkan model prediksi, tetapi juga memberikan pemahaman yang lebih mendalam mengenai faktor tidur yang paling berpengaruh terhadap kondisi stres remaja. Hasil interpretatif ini dapat menjadi dasar bagi intervensi praktis pada bidang psikologi pendidikan maupun kesehatan mental remaja, seperti penyusunan strategi pencegahan atau rekomendasi kebiasaan tidur yang lebih sehat.

3.2 Tahapan Penelitian

Untuk mencapai tujuan penelitian dan memastikan proses berjalan secara sistematis serta terukur, penelitian ini dilaksanakan melalui serangkaian tahapan yang disusun secara runtut. Adapun langkah-langkah penelitian yang dilakukan adalah sebagai berikut:

3.2.1 Penjadwalan Penelitian

Penjadwalan penelitian ini disusun secara sistematis, dengan rentang waktu pelaksanaan selama dua bulan (Oktober–November). Jadwal ini berfungsi sebagai acuan waktu yang terstruktur untuk memastikan setiap tahap, mulai dari pengolahan data, pemodelan, evaluasi, hingga penyusunan laporan, dapat berjalan secara terencana dan efisien.

No	Kegiatan	Oktober				November	
		I	II	III	IV	I	II
1.	Persiapan Penelitian						
2.	Pengumpulan dan Persiapan Data						
3.	<i>Data Preprocessing</i>						
4.	Pemodelan & Evaluasi Random Forest						
5.	<i>Analisis Feature Importance</i>						
6.	Analisis Hasil						
7.	Laporan dan Dokumentasi Akhir						

3.2.2 Pengumpulan Data

Pada penelitian ini, data diperoleh dari dataset pola tidur dan gejala insomnia pada remaja yang dipublikasikan melalui platform *Data in Brief*.

Dataset tersebut menjadi sumber utama untuk seluruh proses analisis, mulai dari persiapan data, eksplorasi statistik, hingga pelatihan dan evaluasi model Machine Learning. Setiap variabel dalam dataset diperiksa terlebih dahulu untuk memastikan kelengkapan, konsistensi, dan kesesuaiannya dengan tujuan penelitian, yaitu memprediksi tingkat stres remaja berdasarkan parameter pola tidur menggunakan algoritma Random Forest.

3.2.2.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder kuantitatif yang telah dikumpulkan dan dipublikasikan oleh peneliti sebelumnya. Penggunaan data sekunder dipilih karena instrumen psikometri yang digunakan pada dataset tersebut telah melalui proses validasi dan standarisasi internasional, sehingga layak digunakan untuk analisis ilmiah.

Deskripsi	Detail
Jenis Data	Data sekunder kuantitatif
Sumber Penelitian	Orsolya Kiss, Dilara Yuksel, Devin E. Prouty, Fiona C Baker, Massimiliano de Zambotti. (2022).
Judul Studi Asli	<i>A dataset reflecting the multidimensionality of insomnia symptomatology in adolescence using standardized questionnaires</i>
Nama File Dataset	insomnia_data.csv

Dataset ini berisi variabel terkait pola tidur dan parameter psikometri yang relevan untuk analisis prediksi tingkat stres pada remaja.

3.2.2.2 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah remaja (*adolescents*) yang menjadi subjek pengukuran gejala insomnia dan kelompok kontrol. Sampel yang digunakan berjumlah 95 responden, sesuai dengan data yang tersedia pada file insomnia_data.csv.

Pemilihan jumlah sampel ini dinilai memadai untuk analisis menggunakan Random Forest, karena algoritma tersebut memiliki performa yang baik meskipun ukuran sampel tidak besar. Selain itu, fokus utama penelitian bukan hanya pada performa prediksi, tetapi juga pada interpretasi model melalui analisis Feature Importance serta validasi hasil model.

3.2.3 Perencanaan Metode Penelitian

Tahap perencanaan metode penelitian dilakukan untuk merancang keseluruhan alur kerja penelitian secara sistematis dan terstruktur. Perencanaan ini mencakup penetapan pendekatan, teknik analisis, serta tools yang digunakan agar proses penelitian berjalan efektif dan sesuai dengan tujuan penelitian. Adapun perencanaan metode penelitian meliputi beberapa tahapan berikut, yaitu:

1. Studi Literatur

Tahap ini dilakukan dengan menelusuri teori dan penelitian sebelumnya terkait stres remaja, pola tidur (PSQI, ISI, ASHS), penggunaan data psikometri, serta penerapan algoritma Random Forest dalam penelitian prediktif. Hasil studi literatur menjadi dasar untuk menyusun kerangka teori, menentukan variabel penelitian, serta merancang strategi analisis.

2. Penentuan Metode dan Pendekatan

Pada tahap ini ditetapkan bahwa penelitian menggunakan pendekatan kuantitatif dengan jenis analisis prediktif berbasis Machine Learning. Algoritma utama yang digunakan adalah Random Forest Classifier, karena mampu menangani data multidimensi dan memberikan interpretasi melalui Feature Importance. Teknik evaluasi yang digunakan meliputi akurasi, precision, recall, F1-score, serta confusion matrix.

3. Persiapan Tools dan Environment

Tahap ini bertujuan menyiapkan seluruh perangkat dan lingkungan pengembangan yang diperlukan untuk proses analisis, antara lain:

- a. Python sebagai bahasa pemrograman utama.
- b. Library seperti pandas, numpy, scikit-learn, dan matplotlib,
- c. platform komputasi seperti Google Colab untuk menjalankan eksperimen
- d. penyimpanan dataset dan dokumentasi kerja.

Persiapan environment memastikan semua kebutuhan teknis siap sebelum pengembangan dimulai.

4. Perancangan Metrik Evaluasi

Pada tahap ini ditentukan metrik evaluasi model yang digunakan untuk menilai performa prediksi tingkat stres, yaitu:

- a. Accuracy
- b. Precision
- c. Recall
- d. F1-score
- e. Confusion Matrix

Pemilihan metrik ini bertujuan untuk memastikan bahwa kualitas model dapat diukur secara objektif dan dapat dibandingkan antar kelas (Rendah, Sedang, Tinggi).

Selain itu, tahap ini juga mencakup perencanaan analisis interpretatif melalui Feature Importance, sehingga seluruh proses pemodelan diarahkan untuk tidak hanya menghasilkan prediksi yang akurat, tetapi juga memberikan pemahaman mengenai variabel pola tidur.

3.2.4 Pengembangan Sistem

Tahap pengembangan sistem fokus pada implementasi proses analisis data dan pembangunan model prediksi berbasis Random Forest. Aktivitas utama pada tahap ini meliputi:

1. Implementasi Data Preprocessing
Melakukan pembersihan data, penanganan missing values, encoding variabel kategorikal, dan kategorisasi variabel target sesuai definisi operasional.
2. Pembuatan Pipeline Modeling
Merancang alur sistem yang mencakup pemisahan data, pelatihan model Random Forest, hyperparameter tuning, serta penyimpanan model dalam format yang dapat digunakan ulang.
3. Implementasi Evaluasi Otomatis
Menyusun kode evaluasi model untuk menghasilkan akurasi, confusion matrix, serta laporan classification metrics yang menampilkan precision, recall, dan F1-score pada setiap kelas.
4. Integrasi Analisis Feature Importance
Menghasilkan visualisasi importance pada model Random Forest, serta menyediakan opsi penggunaan SHAP jika analisis lanjutan diperlukan.

Tahap ini menghasilkan sistem analisis yang dapat dijalankan secara konsisten untuk mengolah dataset hingga menghasilkan prediksi dan interpretasi model.

3.2.5 Analisis Hasil Penelitian

Tahap ini bertujuan menginterpretasikan seluruh keluaran model dan analisis statistik awal. Analisis dilakukan dalam beberapa bagian:

1. Analisis Statistik Awal
Menginterpretasikan hasil deskriptif dan korelasi untuk menjawab rumusan masalah mengenai gambaran pola tidur dan hubungan dengan tingkat stres.

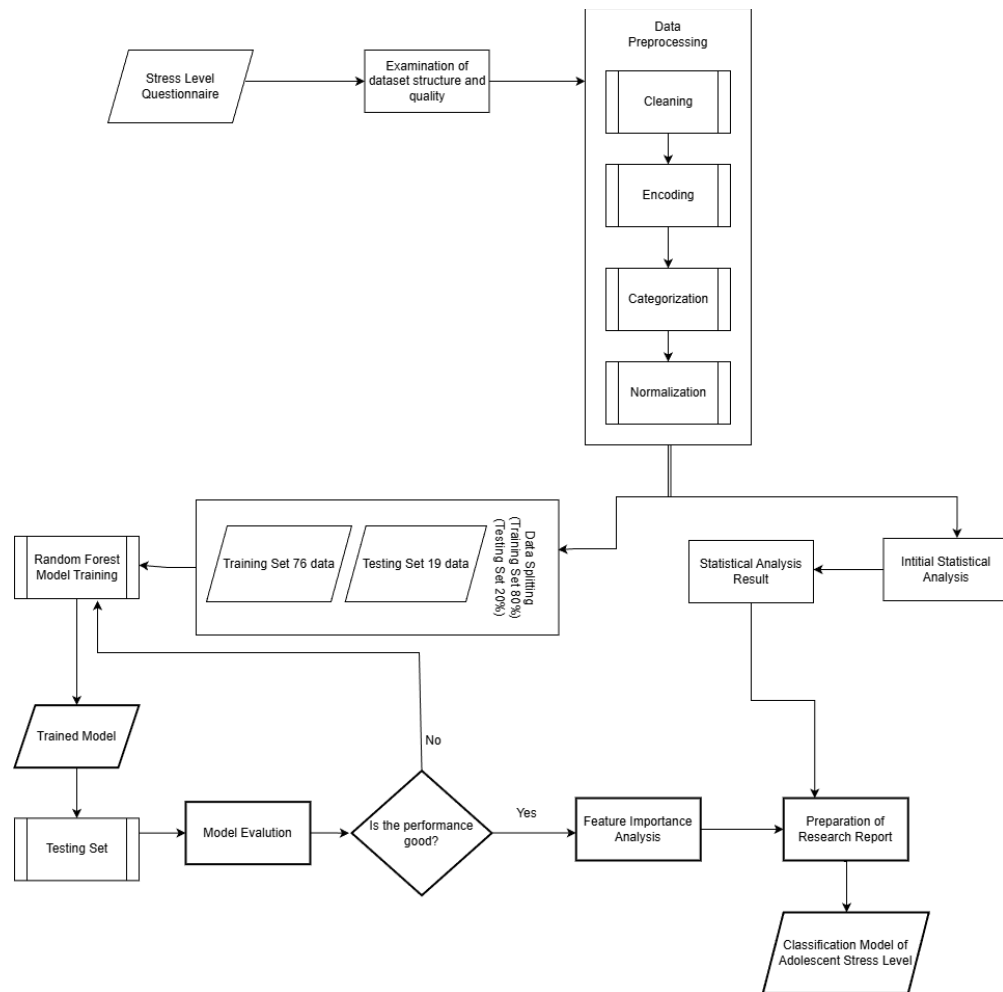
2. Analisis Kinerja Model Random Forest
Membahas nilai akurasi, distribusi prediksi pada confusion matrix, serta performa setiap kelas melalui precision, recall, dan F1-score.
3. Analisis Feature Importance
Menafsirkan fitur-fitur yang memiliki nilai importance paling tinggi untuk memahami variabel pola tidur mana yang paling memengaruhi prediksi tingkat stres.
4. Sintesis Temuan
Menggabungkan seluruh hasil analisis statistik dan model prediktif untuk menjawab seluruh rumusan masalah penelitian secara komprehensif.

3.2.6 Penyusunan Laporan Kegiatan

Tahap penyusunan laporan penelitian merupakan fase penutup dari seluruh rangkaian kegiatan penelitian. Pada tahap ini, setiap proses, hasil analisis, dan temuan utama disusun serta didokumentasikan secara teratur, menyeluruh, dan sesuai kaidah ilmiah ke dalam bentuk laporan penelitian yang utuh.

3.3 Kerangka Penelitian (*Research Framework*)

Kerangka kerja penelitian ini menggambarkan tahapan sistematis yang digunakan dalam proses pengembangan dan evaluasi model klasifikasi tingkat stres remaja. Penelitian ini terdiri dari beberapa tahap utama, yaitu *Data Preprocessing*, *Model Development*, *Model Evaluation*, dan *Feature Importance Analysis*. Alur lengkap penelitian ditunjukkan pada diagram kerangka kerja yang disajikan pada Gambar 3.1



Gambar 3.1 Kerangka Penelitian

3.4 Pengembangan Perangkat Lunak

Pada tahap pengembangan perangkat lunak, seluruh proses teknis yang berkaitan dengan pengolahan data, pembangunan model, pelatihan, serta evaluasi performa sistem dilakukan secara terstruktur. Tahapan ini berfokus pada implementasi algoritma Random Forest sebagai model utama dalam memprediksi tingkat stres remaja berdasarkan pola tidur. Proses pengembangan mencakup langkah-langkah seperti preprocessing dataset insomnia remaja, pembagian data menjadi set pelatihan dan pengujian, pelatihan model Random Forest beserta penyesuaian hyperparameter, serta evaluasi performa model menggunakan metrik akurasi, confusion matrix, precision, recall, dan F1-score.

Selain evaluasi performa, tahap pengembangan sistem juga mencakup analisis Feature Importance untuk menilai kontribusi relatif setiap variabel pola tidur terhadap

prediksi model, sehingga hasil pemodelan tidak hanya menghasilkan akurasi yang baik, tetapi juga memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang paling berpengaruh.

3.4.1 Tahap *Data Preprocessing*

Tahap ini bertujuan untuk menyiapkan data mentah (*insomnia_data.csv*) agar siap diproses oleh algoritma Random Forest.

1. *Pemilihan Fitur*: Memilih kolom-kolom yang relevan dengan variabel X dan Y (seperti yang didefinisikan di 3.3.2), serta mengabaikan variabel yang tidak terkait (misalnya, skor item level, variabel *coping* yang tidak relevan).
2. *Penanganan Missing Values*: Melakukan pengecekan terhadap nilai data yang hilang (NaN). Berdasarkan ukuran sampel yang kecil (95 responden), setiap *missing value* yang ditemukan akan diatasi dengan metode yang paling sesuai (misalnya, imputasi dengan nilai rata-rata/median) atau penghapusan baris jika *missing value* ditemukan pada variabel kunci.
3. *Transformasi Variabel Target (Kategorisasi PSS)*: Variabel kontinu PSS_total dikonversi menjadi variabel kategorikal diskret (Rendah, Sedang, Tinggi) menggunakan kriteria *cut-off* yang ditentukan dalam Sub-bab 3.3.2.1
4. *Encoding Fitur Kategorikal*: Variabel fitur kategorikal (misalnya, *sex*/jenis kelamin) akan diubah menjadi format numerik menggunakan *One-Hot Encoding* atau *Label Encoding* agar dapat diproses oleh model.
5. *Normalisasi/Standardisasi*: Karena Random Forest adalah algoritma berbasis pohon yang tidak terlalu sensitif terhadap skala, normalisasi tidak diwajibkan tetapi dapat dipertimbangkan jika diperlukan oleh proses *hyperparameter tuning*.

3.4.2 Tahap Pengembangan Model (*Random Forest*)

1. Pembagian Data (*Data Splitting*): Dataset yang sudah bersih dibagi menjadi dua set untuk menghindari *overfitting* dan memastikan generalisasi model. Digunakan rasio:
 - a. Data Latih (*Training Set*): 80% dari total data (sekitar 76 responden).
 - b. Data Uji (*Testing Set*): 20% dari total data (sekitar 19 responden).
2. Pelatihan Model: Model Random Forest Classifier diinisialisasi dan dilatih menggunakan *Training Set*.
3. *Hyperparameter Tuning*: Untuk optimasi model, akan dilakukan *tuning* pada parameter kunci, seperti jumlah pohon (*n_estimators*) (misalnya 100-500) dan kedalaman maksimum pohon (*max_depth*). Metode *tuning* seperti *Grid Search CV* atau *Randomized Search CV* dapat digunakan untuk menemukan kombinasi *hyperparameter* terbaik.

3.4.3 Tahap Evaluasi Model

Evaluasi kinerja model Random Forest dilakukan menggunakan *Testing Set* dengan beberapa metrik untuk menilai performa prediksi pada setiap kelas kategori stres. Evaluasi dilakukan secara komprehensif agar hasil prediksi tidak hanya mengukur akurasi global, tetapi juga kualitas model dalam membedakan tiap kelas.

Metrik evaluasi yang digunakan meliputi:

1. Akurasi (*Accuracy*)

Mengukur proporsi prediksi yang benar dari keseluruhan data uji.

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

Sumber: Medium

2. Matriks Kebingungan (*Confusion Matrix*)

Matriks Kebingungan adalah tabel spesifik yang menyajikan visualisasi kinerja algoritma klasifikasi dengan mencatat jumlah hasil prediksi model dibandingkan dengan label aktual (nilai kebenaran dasar).

Elemen dasar Matriks Kebingungan digunakan untuk menghitung metrik kinerja:

1. TP (*True Positive*): Jumlah sampel yang benar diprediksi sebagai kelas positif.
2. TN (*True Negative*): Jumlah sampel yang benar diprediksi sebagai kelas negatif.
3. FP (*False Positive*): Jumlah sampel yang salah diprediksi sebagai kelas positif (*Type I Error*).
4. FN (*False Negative*): Jumlah sampel yang salah diprediksi sebagai kelas negatif (*Type II Error*).

3. Metrik Per-Kelas

Metrik ini dihitung untuk setiap kelas kategori stres:

a. Presisi (*Precision*)

Presisi (atau *Positive Predictive Value*) adalah proporsi sampel positif yang diprediksi dengan benar terhadap total sampel yang diprediksi sebagai positif.

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

Sumber: Medium

b. Recall

Recall (atau *Sensitivity*) adalah proporsi sampel positif yang berhasil diprediksi dengan benar terhadap total sampel yang sebenarnya positif (label aktualnya positif).

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

Sumber: Medium

c. *F1-Score*

F1-Score adalah *harmonic mean* (rata-rata harmonis) dari Presisi dan *Recall*. Metrik ini secara simetris mewakili kedua metrik dan sering digunakan untuk mengevaluasi kinerja pada dataset multikelas.

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

Sumber: Medium

3.4.4 Tahap Analisis *Feature Importance*

Tahap ini berfungsi untuk memberikan pemahaman tambahan mengenai kontribusi setiap fitur pola tidur dalam proses pemodelan prediksi tingkat stres. Analisis ini dilakukan setelah model Random Forest terbentuk dan memberikan gambaran mengenai variabel mana yang paling relevan dalam menentukan hasil klasifikasi.

Langkah analisis dilakukan melalui beberapa prosedur berikut:

1. Perhitungan *Feature Importance*

Kontribusi masing-masing fitur dihitung menggunakan metode *Mean Decrease Impurity* (MDI), yaitu pengukuran penurunan impuritas pada setiap node pemisah di seluruh pohon dalam model Random Forest.

2. Analisis Interpretabilitas Tambahan

Jika diperlukan dan memungkinkan secara teknis, dilakukan analisis lanjutan menggunakan SHAP (*Shapley Additive Explanations*) untuk

memberikan interpretasi yang lebih mendetail, baik pada level keseluruhan model maupun pada prediksi individu.

3. Visualisasi dan Interpretasi Hasil

Nilai *Feature Importance* disajikan dalam bentuk grafik batang untuk memudahkan pemahaman dan penarikan kesimpulan. Interpretasi diarahkan untuk mengidentifikasi variabel atau sub-komponen pola tidur yang menunjukkan kontribusi terbesar terhadap prediksi tingkat stres remaja.

3.5 Variabel Penelitian dan Definisi Operasional

3.5.1 Variabel Penelitian

Variabel penelitian dibagi menjadi dua kategori, yaitu variabel dependen sebagai variabel target dan variabel independen sebagai fitur prediktor dalam pemodelan Machine Learning.

Kategori	Variabel	Instrumen	Tipe Data	Keterangan
Variabel Dependen (Y)	Tingkat Stres	Perceived Stress Scale (PSS_total)	Kategorikal (Diskrit)	Variabel target yang diklasifikasikan menjadi Rendah, Sedang, atau Tinggi.
Variabel Independen (X)	Pola Tidur Multidimensi	PSQI, ISI, ASHS	Numerik	Fitur input utama untuk memprediksi tingkat stres.

3.5.2 Definisi Operasional

Mendefinisikan operasional dari setiap variabel yang digunakan dalam penelitian, termasuk instrumen pengukuran dan skala data yang diterapkan.

3.5.2.1 Variabel Target (Y) / Tingkat Stres

Variabel	Definisi	Pengukuran	Skala Data
Tingkat Stres	Persepsi individu terhadap seberapa tidak terkontrol dan kewalahan dirinya terhadap situasi hidup, diukur dengan PSS-10.	Kategorisasi PSS: Skor PSS_total (0-40) dikonversi menjadi kategori diskrit: a. Rendah (0-13). b. Sedang (14-26). c. Tinggi (27-40).	Ordinal

3.5.2.2 Variabel Fitur (X) / Pola Tidur Multidimensi

Variabel independen mencakup skor utama dan sub-komponen instrumen terkait pola tidur yang digunakan sebagai prediktor dalam pemodelan.

Fitur yang Digunakan	Asal Instrumen	Keterangan
ISI_Total	<i>Insomnia Severity Index</i>	Tingkat keparahan insomnia.
PSQI_total	<i>Pittsburgh Sleep Quality Index</i>	Kualitas tidur keseluruhan.
ASHS_total	<i>Adolescent Sleep Hygiene Scale</i>	Kebersihan tidur keseluruhan.
PSQI Sub-komponen	PSQI	Durasi Tidur, Latensi Tidur, Efisiensi Tidur, Gangguan Tidur, dll.
ASHS Sub-komponen	ASHS	<i>Physiological, Cognitive, Emotional, Sleep Environment</i> , dll.
Fitur Tambahan	(Misalnya, <i>Age, Sex</i>)	Usia dan jenis kelamin (akan diuji korelasi)

		awalnya).
--	--	-----------

3.6 Etika Penelitian

Penelitian ini dilaksanakan dengan mematuhi prinsip-prinsip etika penelitian ilmiah, terutama karena menggunakan data yang berkaitan dengan kondisi psikologis remaja. Seluruh data yang digunakan berasal dari dataset sekunder yang telah dipublikasikan secara terbuka melalui *Data in Brief*, dengan informasi yang telah dianonimkan sepenuhnya oleh peneliti asli. Dengan demikian, tidak terdapat identitas pribadi atau informasi sensitif yang dapat mengarah pada individu tertentu.

Pemanfaatan dataset dilakukan sesuai dengan lisensi dan ketentuan penggunaan yang ditetapkan oleh peneliti dan penerbit. Penelitian ini juga memastikan bahwa seluruh proses analisis dilakukan untuk tujuan akademik, tanpa rekayasa data, manipulasi hasil, atau praktik yang dapat mengurangi integritas ilmiah. Keamanan data dijaga melalui penyimpanan hanya di perangkat lokal selama proses penelitian, dan tidak dibagikan kepada pihak lain.

Selain itu, interpretasi hasil penelitian dilakukan secara objektif tanpa memberikan stigma atau generalisasi berlebihan terhadap kelompok remaja. Fokus penelitian tetap pada pemodelan prediktif dan pemahaman faktor pola tidur, bukan penilaian individu. Dengan mengikuti prinsip-prinsip tersebut, penelitian ini memenuhi standar etika yang berlaku dalam penelitian psikologi, data sekunder, dan pembelajaran mesin.

DAFTAR PUSTAKA

- Astari, D. F., Chrisnanto, Y. H., & Melina. (2023). Klasifikasi tingkat stres saat tidur menggunakan algoritma Random Forest. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3676–3684.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cerri, M. (2023). Insomnia Severity Index reliability and psychometric properties in adolescent samples: A systematic review. *Sleep Medicine Reviews*, 68, 101–115.
- Chehri, A., Ahadi, H., & Asgari, P. (2020). Validation of the Pittsburgh Sleep Quality Index among adolescents. *Journal of Sleep Research*, 29(4), 1–8.
- Chehri, A., Moradi, A., & Sadeghi, K. (2023). Sleep hygiene behaviors and psychological outcomes in adolescents: A psychometric update of ASHS. *Journal of Adolescent Health*, 72(2), 233–241.
- Dube, L., & Verster, T. (2024). Interpretability of the random forest model under class imbalance. *Data Science in Finance and Economics*, 4(3), 446–468.
- Gao, X. (2023). Academic stress and academic burnout in adolescents: a moderated mediating model. *Frontiers in Psychology*, 14, 1133706.
- Johri, K., Pillai, R., Kulkarni, A., & Balkrishnan, R. (2025). Effects of sleep deprivation on the mental health of adolescents: a systematic review. *Sleep Science and Practice*, 9(9).
- Kim, H., Kim, S. H., Jang, S.-I., & Park, E.-C. (2022). Association between sleep quality and anxiety in Korean adolescents. *Journal of Preventive Medicine & Public Health*, 55(3), 173–181.
- Kiss, O., Yuksel, D., Prouty, D. E., Baker, F. C., & de Zambotti, M. (2022). A dataset reflecting the multidimensionality of insomnia symptomatology in adolescence using standardized questionnaires. *Data in Brief*, 44, 108523.
- Liu, J. (2023). Prediction of stress levels in sleep patterns based on random forest. *Proceedings of the 2023 International Conference on Machine Learning and Automation* (pp. 69–74).

- Maçãs, C., Campos, J. R., Lourenço, N., & Machado, P. (2024). Visualisation of Random Forest classification. *Information Visualization*, 23(4), 312–327.
- Simon, S. M., Glaum, P., & Valdovinos, F. S. (2023). Interpreting random forest analysis of ecological models to move from prediction to explanation. *Scientific Reports*, 13(3881).
- Stromájer, G. P., Csimá, M., Iváncsik, R., Varga, B., Takács, K., & Stromájer-Rácz, T. (2023). Stress and Anxiety among High School Adolescents: Correlations between Physiological and Psychological Indicators in a Longitudinal Follow-Up Study. *Children*, 10(9), 1548.
- ten Brink, M., Lee, H. Y., Manber, R., Yeager, D. S., & Gross, J. J. (2021). Stress, Sleep, and Coping Self-Efficacy in Adolescents. *Journal of Youth and Adolescence*, 50(3), 485–505.
- Uccella, S., Cordani, R., Salfi, F., Gorgoni, M., Scarpelli, S., Gemignani, A., ... Nobili, L. (2023). Sleep deprivation and insomnia in adolescence: Implications for mental health. *Brain Sciences*, 13(4), 569.
- Wahyudi, A. P., Mahesa, A. L. P., Apsari, A. N., Dalimunthe, D. W., Widitya, S., & Afifah, S. (2025). Faktor yang Mempengaruhi Kesehatan Remaja di Indonesia: Studi Literatur. *JUKEJ: Jurnal Kesehatan Jompa*, 4(1), 503–509.
- Wang, H., & Fan, X. (2023). Academic Stress and Sleep Quality among Chinese Adolescents: Chain Mediating Effects of Anxiety and School Burnout. *International Journal of Environmental Research and Public Health*, 20(3), 2219.
- Xiao, X., Zou, Y., Huang, J., Luo, X., Yang, L., Li, M., ... Li, Y. (2024). An interpretable model for landslide susceptibility assessment based on Optuna hyperparameter optimization and Random Forest. *Geomatics, Natural Hazards and Risk*, 15(1), 2347421.