

BAB III

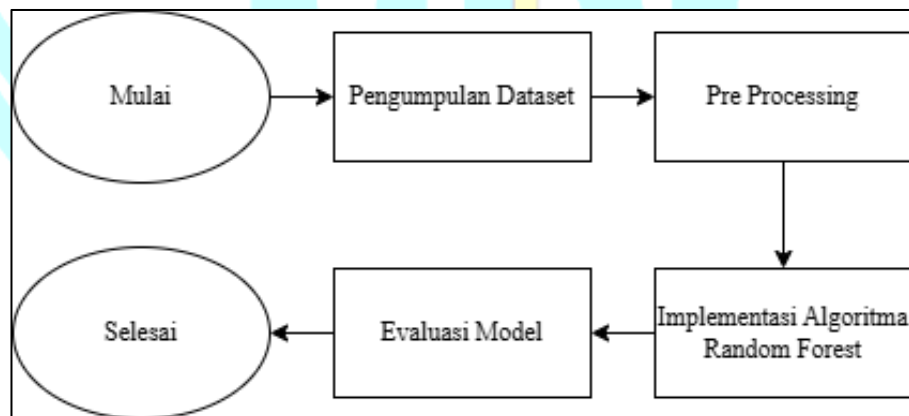
METODE PENELITIAN

3.1 Objek Penelitian

Objek Penelitian ini adalah identifikasi faktor penyebab tingkat stres pada manusia berdasarkan beberapa variabel pada *Dataset* yang ada menggunakan algoritma *Random Forest*. Penelitian ini bertujuan untuk identifikasi faktor penyebab stres dari beberapa faktor yang ada. *Dataset* yang digunakan bersumber dari *Website Kaggle* dengan nama *Dataset*-nya “*Stress Level Dataset*” didalam *Dataset* ini berjumlah 1100 data dengan beberapa variabel faktor yang mendukung untuk identifikasi terhadap faktor stres pada manusia.

3.2 Prosedur Penelitian

Untuk memperoleh hasil yang diharapkan dalam penelitian ini, diperlukan langkah-langkah penelitian yang terstruktur. Berikut prosedur yang dilakukan digambarkan pada Gambar 3.1.



Gambar 3. 1 Prosedur Penelitian

3.1.1 Pengumpulan Data

Mengumpulkan data yang akan digunakan dalam proses identifikasi faktor stres. *Dataset* yang di peroleh dari *Website Kaggle* dengan nama *Dataset*-nya “*Stress Level Dataset*” dan di terbitkan oleh Chhabii.

Didalam *Dataset* ini berjumlah 1100 data dan memiliki 21 atribut seperti Tabel 3.1 di bawah ini.

Tabel 2. 3 Fitur *Dataset*

Fitur	Keterangan
<i>anxiety_level (AL)</i>	Tingkat kecemasan individu dengan level (0 – 21)
<i>self_esteem (SE)</i>	Rasa percaya diri atau harga diri dengan level (0 – 30)
<i>mental_health_history (MHH)</i>	Riwayat masalah kesehatan mental label 0 adalah tidak dan adalah ya
<i>depression (DN)</i>	Tingkat depresi yang dialami dengan level (0 – 27)
<i>headache (HE)</i>	Frekuensi atau tingkat keparahan sakit kepala dengan level (0 – 5)
<i>blood_pressure (BP)</i>	Tingkat tekanan darah individu dengan level (1 – 3)
<i>sleep_quality (SQ)</i>	Kualitas tidur seseorang dengan level (0 – 5).
<i>breathing_problem (BPM)</i>	Masalah pernapasan yang dialami dengan level (0 – 5)
<i>noise_level (NL)</i>	Tingkat kebisingan di lingkungan dengan level (0 – 5).
<i>living_conditions (LC)</i>	Kondisi tempat tinggal dengan level (0 – 5)
<i>safety (SFT)</i>	Tingkat rasa aman individu dengan level (0 – 5)
<i>basic_needs (BN)</i>	Kemampuan memenuhi kebutuhan dasar dengan level (0 – 5)
<i>academic_performance (AP)</i>	Prestasi akademik atau nilai dengan level (0 – 5)

Fitur	Keterangan
<i>study_load (SL)</i>	Beban studi atau pekerjaan akademik dengan level (0 – 5)
<i>teacher_student_relationship (TSR)</i>	Hubungan antara guru dan siswa dengan level (0 – 5)
<i>future_career_concerns (FCC)</i>	Kekhawatiran tentang masa depan karier dengan level (0 – 5)
<i>social_support (SS)</i>	Dukungan sosial dari keluarga atau teman dengan level (0 – 3)
<i>peer_pressure (PP)</i>	Tekanan dari teman sebaya dengan level (0 – 5)
<i>extracurricular_activities (EA)</i>	Kegiatan ekstrakurikuler yang diikuti dengan level (0 – 5)
<i>Bullying (BLY)</i>	Pengalaman dibuli atau dilecehkan dengan level (0 – 5)
<i>stress_level (STL)</i>	Tingkat stres yang dialami dengan level (0 – 2)
	target yang akan digunakan

3.1.2 Pre-processing

Sebelum data dimasukkan ke dalam model, beberapa proses preprocessing dilakukan untuk memastikan bahwa data siap untuk di analisis:

1. Penanganan Nilai yang Hilang: Dalam kasus di mana *Dataset* yang memiliki nilai yang hilang akan dilakukan imputasi dengan menggunakan metode yang sesuai (misalnya, median imputasi untuk fitur numerik).
2. Penghapusan Kolom: Dalam kasus ini kolom yang tidak relevan pada *Dataset* yang digunakan akan dihapus.
3. Normalisasi/Skalasi: Jika diperlukan, beberapa fitur dapat dinormalisasi atau diskalasi menggunakan metode *Min-Max* untuk membuat model bekerja lebih baik. normalisasi yang merupakan proses penskalaan nilai *variable* dari data

sehingga memiliki rentang nilai tertentu atau sama (Dhea et al. 2023). Rumus normalisasi min max dapat dilihat Rumus (6).

$$X^1 = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

Keterangan :

X : nilai asli

X_{min} : nilai minimum dari fitur

X_{max} : nilai maksimum dari fitur

4. Pembagian Data: Data Uji (*Test Data*), Data yang digunakan untuk menguji model dan mengevaluasi performanya sekitar 20% dari total *Dataset*. Data Latih Data yang digunakan untuk melatih model sekitar 80% dari total *Dataset*.

3.1.3 Implementasi Algoritma *Random Forest*

Metode utama dari proses klasifikasi adalah menggunakan algoritma dari pengklasifikasi *Random Forest*. *Random Forest* bekerja dengan menyatukan berbagai pohon keputusan agar hasil prediksinya jadi lebih tepat dan konsisten. Tiap pohon di dalam hutan ini dilatih menggunakan sampel data yang dipilih secara acak dari data latihan. Pendekatan ini memungkinkan setiap pohon untuk mempelajari aspek data yang beragam. Hal ini membuat model menjadi lebih tangguh dan meminimalisir potensi terjadinya *overfitting*. Keputusan akhir diambil berdasarkan pendapat mayoritas dari seluruh pohon keputusan yang telah dibangun. Untuk melihat alur kerja *Random Forest* dilihat pada ilustrasi *Random Forest* terdapat pada Gambar 2.1 dan untuk perhitungan *Random Forest* pada Rumus (1).

3.1.4 Evaluasi Model

Untuk mengetahui seberapa baik model memprediksi tingkat stres, digunakan *Confusion Matrix*. *Confusion Matrix* adalah suatu matriks yang berisi informasi tentang keakuratan prediksi dan kesesuaian dengan keadaan sebenarnya, menunjukkan apakah prediksi tersebut benar atau salah (Abdi et al. 2024).

Pada penelitian ini menggunakan *Confusion Matrix multiclass* karena memiliki 3 kelas Tabel 3.1 menunjukkan bentuk dari *Confusion Matrix multiclass*.

Tabel 3. 1 *Confusion Matrix*

Actual	Prediction		
	Normal (0)	Stres (1)	Stres Tinggi (2)
Normal (0)	TP0	FN	FN
Stres (1)	FN	TP1	FN
Stres Tinggi (2)	FN	FN	TP2

Dalam riset ini, penilaian terhadap model dilakukan dengan memakai metode *Confusion Matrix multiclass*, yang menyuguhkan keluaran evaluasi model dalam wujud tabel matriks. Andaikata *Dataset* terdiri atas 3 kelas, kelas pertama dinamakan Normal (0) kelas kedua dinamakan Stres (1) dan kelas ketiga dinamakan Stres Tinggi (2). *Confusion Matrix* ini memakai tabel matriks yang merekam komparasi luaran klasifikasi data uji berdasarkan data pelatihan dengan data yang senyatanya. Data uji dikerjakan untuk menganalisis unjuk kerja model memakai metrik *Accuracy*, *precision*, *recall*, serta *F1-score*. Perihal formula kalkulasi dari akurasi diperlihatkan pada Rumus (2), lalu formula untuk menghitung *precision* diperlihatkan pada Rumus (3), kemudian formula untuk menghitung *recall* diperlihatkan pada Rumus (4), serta yang terakhir formula buat menghitung *F1-score* diperlihatkan pada Rumus (5).