BAB I PENDAHULUAN

1.1 Latar Belakang

Kesehatan mental mahasiswa merupakan salah satu aspek penting yang perlu mendapat perhatian karena dapat berpengaruh terhadap proses belajar, prestasi akademik, dan kualitas hidup mahasiswa. Selama menempuh pendidikan di perguruan tinggi, mahasiswa menghadapi berbagai tuntutan, seperti beban tugas, ujian, persaingan akademik, keterbatasan waktu, tekanan ekonomi, serta permasalahan sosial. Berbagai tuntutan tersebut dapat menjadi sumber stres apabila tidak diimbangi dengan kemampuan mengelola tekanan dan dukungan yang memadai. Penelitian (Albab et al., 2022) terhadap 311 mahasiswa Fakultas Ilmu Pendidikan Universitas Negeri Malang menunjukkan bahwa 13,5% responden mengalami stres akademik kategori rendah, 59,2% kategori sedang, dan 27,3% kategori tinggi. Temuan tersebut memperlihatkan bahwa stres akademik merupakan permasalahan yang perlu diperhatikan dalam lingkungan pendidikan tinggi.

Tingkat stres mahasiswa dapat berkaitan dengan berbagai aspek, antara lain kecemasan, penghargaan terhadap diri sendiri, kondisi psikologis, kualitas tidur, beban akademik, dukungan sosial, dan kekhawatiran terhadap karier pada masa depan. Hubungan berbagai aspek tersebut membuat klasifikasi tingkat stres menjadi permasalahan yang tidak selalu dapat dijelaskan melalui satu faktor saja. Pengolahan data dengan pendekatan *machine learning* dapat digunakan untuk mempelajari pola yang terdapat pada sejumlah atribut dan menghasilkan prediksi kategori tingkat stres berdasarkan data yang tersedia. Meskipun demikian, hasil klasifikasi bergantung pada karakteristik data, algoritma yang digunakan, konfigurasi model, dan metode evaluasi. Oleh karena itu, penerapan metode klasifikasi perlu disertai evaluasi yang terstruktur agar performa model dapat dinilai secara lebih objektif.

Salah satu algoritma yang dapat digunakan untuk klasifikasi adalah *Random Forest*. Algoritma ini membentuk sejumlah pohon keputusan dan menggabungkan hasil prediksi pohon-pohon tersebut untuk menghasilkan prediksi akhir. Penggabungan beberapa pohon dapat membantu model mempelajari pola yang beragam pada data. Namun, performa *Random Forest* juga dapat dipengaruhi oleh

pengaturan *hyperparameter*, seperti jumlah pohon, kedalaman pohon, jumlah minimum sampel untuk membagi simpul, jumlah minimum sampel pada daun, kriteria pemisahan, dan pengaturan pengambilan sampel. Konfigurasi yang berbeda dapat menghasilkan model dengan karakteristik dan performa yang berbeda. Oleh sebab itu, optimasi *hyperparameter* perlu diuji untuk mengetahui apakah konfigurasi yang dipilih melalui proses pencarian benar-benar menghasilkan peningkatan performa pada data yang tidak digunakan selama pelatihan.

Optimasi *hyperparameter* pada penelitian ini direncanakan menggunakan *GridSearchCV* dengan *Stratified K-Fold Cross Validation*. *GridSearchCV* mengevaluasi kombinasi parameter yang telah ditentukan, sedangkan *Stratified K-Fold Cross Validation* membagi data menjadi beberapa lipatan dengan mempertahankan proporsi kelas pada setiap lipatan. Penggunaan pembagian berstrata penting karena variabel target dalam penelitian ini terdiri atas tiga kategori tingkat stres, yaitu rendah, sedang, dan tinggi. Dengan mempertahankan proporsi kelas pada proses validasi, evaluasi setiap konfigurasi diharapkan lebih representatif terhadap distribusi kelas dalam data.

Meskipun demikian, konfigurasi *hyperparameter* terbaik berdasarkan proses validasi belum tentu memberikan peningkatan performa ketika diuji pada data yang terpisah. Hal ini dapat terjadi karena konfigurasi terpilih merupakan hasil pemilihan berdasarkan data validasi tertentu, sehingga nilai evaluasinya dapat dipengaruhi oleh variasi pembagian data. Jika data yang digunakan untuk memilih konfigurasi sekaligus digunakan untuk melaporkan performa akhir, hasil evaluasi berisiko memberikan gambaran yang terlalu optimistis. Oleh karena itu, proses pemilihan model dan pengukuran performa perlu dipisahkan agar penilaian terhadap manfaat optimasi dapat dilakukan secara lebih objektif.

Penelitian ini menggunakan pendekatan *Nested Cross Validation* untuk memisahkan proses pemilihan *hyperparameter* dari proses evaluasi model. Pada pendekatan ini, *outer loop* digunakan untuk mengevaluasi performa model pada data pengujian yang tidak terlibat dalam proses pemilihan parameter, sedangkan *inner loop* digunakan untuk memilih konfigurasi *hyperparameter* melalui *GridSearchCV*. Kedua model yang dibandingkan, yaitu *Random Forest* dengan konfigurasi bawaan (*baseline*) dan *Random Forest* dengan konfigurasi hasil

optimasi (*optimized*), dievaluasi menggunakan pembagian data luar yang sama. Dengan rancangan tersebut, perbandingan dilakukan pada data pengujian yang sepadan dan proses optimasi tidak menggunakan data pengujian pada *outer fold* yang bersangkutan.

Penelitian ini menggunakan *dataset Student Stress Factors: A Comprehensive Analysis* yang diperoleh dari platform Kaggle. *Dataset* tersebut terdiri atas 1.100 baris data, 20 atribut masukan, dan satu atribut target, yaitu *stress_level*, yang memuat tiga kategori tingkat stres. Atribut masukan mencakup aspek yang berkaitan dengan kondisi psikologis, kebiasaan, lingkungan, dan kehidupan akademik. Data tersebut digunakan untuk membangun dan mengevaluasi model klasifikasi tingkat stres berdasarkan pola yang tersedia di dalam *dataset*. Hasil klasifikasi dalam penelitian ini tidak dimaksudkan sebagai diagnosis klinis terhadap kondisi kesehatan mental mahasiswa.

Untuk mengurangi ketergantungan hasil evaluasi pada satu pembagian data, eksperimen dilakukan menggunakan lima *random seed*, yaitu 1, 7, 24, 42, dan 99. Pada setiap *random seed*, digunakan *Stratified 5-Fold Cross Validation* pada bagian luar dan bagian dalam. Lima *outer fold* untuk setiap *random seed* menghasilkan 25 eksperimen luar secara keseluruhan. Prediksi dari data pengujian pada setiap *outer fold* dikumpulkan sebagai *out-of-fold prediction*. Karena setiap eksperimen luar menghasilkan prediksi terhadap 1.100 data pada bagian pengujiannya, pengumpulan hasil dari 25 eksperimen menghasilkan 5.500 prediksi secara keseluruhan. Prediksi tersebut berasal dari pengulangan evaluasi pada beberapa pembagian data; dengan demikian, 5.500 prediksi tidak dapat diperlakukan sebagai 5.500 pengamatan independen yang berbeda.

Performa model dibandingkan menggunakan *Accuracy*, *Precision Macro*, *Recall Macro*, dan *F1-Score Macro*. Penggunaan beberapa metrik tersebut bertujuan memberikan gambaran yang lebih menyeluruh mengenai ketepatan prediksi dan performa klasifikasi pada ketiga kelas. Selain perbandingan metrik, hasil prediksi kedua model dianalisis menggunakan *McNemar Exact Test* pada pasangan prediksi yang sesuai. Pengujian ini digunakan untuk menilai apakah terdapat bukti perbedaan proporsi kesalahan prediksi antara kedua model pada data pengujian

yang sama. Hasil pengujian statistik ditafsirkan bersama hasil evaluasi deskriptif, bukan sebagai satu-satunya dasar untuk menentukan model yang lebih baik.

Penelitian terdahulu telah membahas penerapan *machine learning* untuk klasifikasi atau prediksi tingkat stres, termasuk penggunaan *Random Forest* dan perbandingan metode klasifikasi. (Oktaviani et al., 2024) , misalnya, membahas perbandingan kinerja *Random Forest* dan *SMOTE Random Forest* untuk mendeteksi dan mengukur tingkat stres pada mahasiswa tingkat akhir (Anissa & Qoiriah, 2025) membahas prediksi tingkat stres berdasarkan pola hidup menggunakan *machine learning*. Penelitian-penelitian tersebut menjadi landasan bahwa pendekatan komputasional dapat digunakan untuk mempelajari pola yang berkaitan dengan tingkat stres. Namun, perbedaan *dataset*, metode, rancangan validasi, dan metrik evaluasi membuat hasil penelitian terdahulu tidak dapat langsung digunakan untuk menyimpulkan bahwa optimasi *hyperparameter* selalu meningkatkan performa model pada setiap *dataset*.

Berdasarkan permasalahan tersebut, hal yang perlu dikaji bukan hanya apakah proses optimasi dapat menemukan konfigurasi *hyperparameter* dengan skor validasi terbaik, tetapi juga apakah konfigurasi tersebut menghasilkan peningkatan performa ketika dievaluasi secara terpisah dari proses pemilihannya. Penelitian ini berfokus pada pengujian perbandingan *Random Forest baseline* dan *Random Forest optimized* menggunakan *Nested Cross Validation*, lima *random seed*, serta evaluasi metrik klasifikasi dan pengujian statistik terhadap pasangan prediksi. Rancangan tersebut digunakan untuk mengurangi risiko penilaian performa yang terlalu optimistis akibat penggunaan data yang sama untuk pemilihan model dan evaluasi.

Dengan demikian, penelitian ini bertujuan menganalisis penerapan optimasi *hyperparameter Random Forest* untuk klasifikasi tingkat stres mahasiswa pada *dataset* yang digunakan, membandingkan performa model sebelum dan sesudah optimasi, serta mengevaluasi bukti perbedaan hasil prediksi kedua model. Penelitian ini juga menggunakan *Permutation Importance* sebagai analisis tambahan untuk mengamati kontribusi relatif atribut terhadap prediksi model. Hasil penelitian diharapkan dapat memberikan gambaran yang lebih terukur mengenai

manfaat optimasi *hyperparameter* dalam kondisi eksperimen yang ditetapkan, sekaligus menjelaskan batasan interpretasi hasil klasifikasi pada *dataset* tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Seberapa baik performa algoritma Random Forest dalam mengklasifikasikan tingkat stres mahasiswa berdasarkan *dataset Student Stress Factors: A Comprehensive Analysis* berdasarkan nilai Accuracy, Precision Macro, Recall Macro, dan F1-Score Macro?
2. Bagaimana proses optimasi *hyperparameter Random Forest* menggunakan *GridSearchCV* dengan *Stratified 5-Fold Cross Validation* pada bagian validasi dalam pendekatan *Nested Cross Validation*?
3. Bagaimana perbandingan performa *Random Forest baseline* dan *Random Forest optimized* berdasarkan metrik *Accuracy*, *Precision Macro*, *Recall Macro*, dan *F1-Score Macro* pada rangkaian eksperimen yang dilakukan?
4. Apakah terdapat bukti perbedaan proporsi kesalahan prediksi antara *Random Forest baseline* dan *Random Forest optimized* berdasarkan *McNemar Exact Test* pada pasangan prediksi yang sesuai?
5. Bagaimana kontribusi relatif atribut terhadap prediksi model *Random Forest optimized* berdasarkan analisis *Permutation Importance*, dengan mempertimbangkan kestabilan dan keterbatasan interpretasinya?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, tujuan penelitian ini adalah:

1. Mengevaluasi performa algoritma Random Forest dalam mengklasifikasikan tingkat stres mahasiswa berdasarkan *dataset Student Stress Factors: A Comprehensive Analysis* menggunakan metrik Accuracy, Precision Macro, Recall Macro, dan F1-Score Macro.
2. Melakukan optimasi *hyperparameter Random Forest* menggunakan *GridSearchCV* dengan *Stratified 5-Fold Cross Validation* pada bagian validasi dalam pendekatan *Nested Cross Validation*.
3. Membandingkan performa *Random Forest baseline* dan *Random Forest optimized* berdasarkan metrik *Accuracy*, *Precision Macro*, *Recall Macro*,

dan *F1-Score Macro* pada rangkaian eksperimen dengan lima *random seed* dan lima *outer fold*.

4. Menganalisis bukti perbedaan proporsi kesalahan prediksi antara *Random Forest baseline* dan *Random Forest optimized* menggunakan *McNemar Exact Test* pada pasangan prediksi yang sesuai, dengan tingkat signifikansi sebesar 0,05.
5. Menganalisis kontribusi relatif atribut terhadap prediksi model *Random Forest optimized* menggunakan *Permutation Importance* serta menjelaskan keterbatasan interpretasi hasilnya.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan informasi mengenai penerapan algoritma *Random Forest* untuk klasifikasi tingkat stres mahasiswa berdasarkan *dataset Student Stress Factors: A Comprehensive Analysis*.
2. Menambah pemahaman mengenai penggunaan *GridSearchCV, Stratified K-Fold Cross Validation,* dan *Nested Cross Validation* dalam proses optimasi *hyperparameter* serta evaluasi model klasifikasi.
3. Memberikan informasi mengenai perbandingan performa model *Random Forest baseline* dan *Random Forest optimized* berdasarkan metrik *Accuracy, Precision Macro*, *Recall Macro*, dan *F1-Score Macro*.
4. Memberikan gambaran mengenai kestabilan performa model melalui pengujian dengan beberapa *random seed* dan pembagian data yang berbeda.
5. Memberikan informasi mengenai hasil pengujian statistik terhadap perbedaan proporsi kesalahan prediksi antara model *Random Forest baseline* dan *Random Forest optimized* menggunakan *McNemar Exact Test*.
6. Memberikan informasi mengenai kontribusi relatif atribut terhadap prediksi model berdasarkan analisis *Permutation Importance*, dengan tetap memperhatikan keterbatasan interpretasinya.
7. Menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan klasifikasi tingkat stres mahasiswa, optimasi *hyperparameter*, evaluasi model, pengujian statistik, dan interpretasi model *machine learning*.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan *dataset Student Stress Factors: A Comprehensive Analysis* yang diperoleh dari platform Kaggle, dengan jumlah 1.100 data, 20 fitur masukan, dan satu atribut target, yaitu *stress_level*. Identitas dan karakteristik responden dalam *dataset* tidak terverifikasi secara spesifik sebagai mahasiswa Indonesia. Selain itu, keberadaan fitur *teacher_student_relationship* mengindikasikan bahwa *dataset* dapat mencakup konteks pendidikan sekolah. Oleh karena itu, hasil penelitian ini terbatas pada klasifikasi berdasarkan karakteristik data yang digunakan dan tidak dapat digeneralisasikan secara langsung untuk menggambarkan tingkat stres mahasiswa Universitas Muhammadiyah Jember (UM Jember) maupun mahasiswa di Indonesia secara keseluruhan.
2. Variabel target yang digunakan adalah stress_level dengan tiga kategori tingkat stres, yaitu rendah, sedang, dan tinggi.
3. Algoritma klasifikasi yang digunakan dalam penelitian ini adalah *Random Forest*.
4. Penelitian membandingkan dua model, yaitu *Random Forest baseline* yang menggunakan konfigurasi bawaan (*default*) dari *Scikit-learn* dan *Random Forest optimized* yang menggunakan konfigurasi *hyperparameter* hasil pencarian *GridSearchCV*.
5. Optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV* dengan *Stratified K-Fold Cross Validation* sebanyak lima *fold* pada proses validasi internal (*inner loop*).
6. Optimasi *hyperparameter* dilakukan terhadap *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *criterion*, dan *bootstrap*. Sementara itu, *max_features* ditetapkan pada nilai sqrt dan tidak divariasikan dalam proses optimasi. Oleh karena itu, pengaruh variasi nilai *max_features* terhadap performa model *Random Forest* belum dievaluasi dalam penelitian ini.
7. Ruang pencarian *GridSearchCV* terdiri atas 576 kombinasi *hyperparameter* berdasarkan nilai-nilai yang ditentukan dalam penelitian.

8. Evaluasi model dilakukan menggunakan *Nested Cross Validation*, dengan *Stratified 5-Fold Cross Validation* pada *outer loop* untuk mengevaluasi performa model dan *inner loop* untuk memilih konfigurasi *hyperparameter* terbaik.
9. Eksperimen dilakukan menggunakan lima nilai *random seed*, yaitu 1, 7, 24, 42, dan 99.
10. Setiap nilai *random seed* menjalankan lima *outer fold*, sehingga diperoleh 25 eksperimen pada *outer loop* (*outer experiment*).
11. Evaluasi performa model dilakukan menggunakan metrik *Accuracy*, *Precision Macro*, *Recall Macro*, dan *F1-Score Macro*.
12. Perbandingan hasil prediksi model *Random Forest baseline* dan *Random Forest optimized* dilakukan menggunakan *McNemar Exact Test* dengan tingkat signifikansi 0,05.
13. Analisis kontribusi fitur terhadap prediksi model dilakukan menggunakan *Permutation Importance*. Hasil analisis digunakan untuk menggambarkan kontribusi relatif fitur dalam model yang dievaluasi, dengan mempertimbangkan keterbatasan interpretasi hasilnya.