

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Depresi merupakan salah satu permasalahan Kesehatan mental yang banyak dialami oleh mahasiswa dan berdampak langsung pada aspek akademik, emosional, serta kualitas hidup individu. Berbagai penelitian menunjukkan bahwa mahasiswa berada pada fase transisi yang rentan, di mana tekanan akademik, tuntutan prestasi, serta masalah personal dan sosial dapat memicu munculnya gejala depresi [1]. Kondisi ini tidak hanya menurunkan motivasi belajar dan konsentrasi, tetapi juga berpotensi mengganggu keberlanjutan studi apabila tidak terdeteksi sejak dini [2].

Seiring dengan perkembangan teknologi, pendekatan berbasis kecerdasan buatan dan machine learning mulai banyak dimanfaatkan untuk mendukung deteksi dini gangguan kesehatan mental. Beberapa penelitian menerapkan pendekatan deep learning untuk mendeteksi depresi menggunakan data tidak terstruktur, seperti teks media sosial. Model hybrid CNN-BiGRU dengan attention mechanism dan FastText terbukti mampu meningkatkan akurasi dalam mendeteksi depresi dari unggahan media sosial berbahasa Indonesia [3]. Pendekatan serupa juga dilakukan dengan membandingkan berbagai arsitektur deep learning seperti LSTM, BiLSTM, GRU, CNN, dan RNN, di mana BiLSTM menunjukkan performa terbaik dalam memahami konteks teks secara dua arah [4]. Selain teks, data visual seperti ekspresi wajah dan arah pandangan mata juga dimanfaatkan untuk deteksi depresi menggunakan arsitektur CNN multi-input tanpa pooling layer, yang menghasilkan akurasi sangat tinggi [5].

Meskipun pendekatan berbasis deep learning menunjukkan performa yang tinggi, metode tersebut umumnya memerlukan data besar, proses prapemrosesan yang kompleks, serta sumber daya komputasi yang tidak sedikit. Selain itu, penggunaan data media sosial atau data visual juga memiliki keterbatasan dalam hal privasi serta penerapan praktis di lingkungan pendidikan. Oleh karena itu, beberapa penelitian mulai mengarah pada penggunaan data survei terstruktur dan algoritma machine learning klasik yang lebih sederhana dan mudah

diinterpretasikan. Random Forest, misalnya, telah digunakan untuk memprediksi kondisi kesehatan mental remaja berdasarkan faktor lingkungan sekolah dan menunjukkan performa yang cukup baik [6].

Berdasarkan berbagai penelitian terdahulu tersebut, dapat disimpulkan bahwa meskipun studi terkait deteksi dan prediksi depresi telah banyak dilakukan, sebagian besar penelitian masih berfokus pada data tidak terstruktur atau menggunakan pendekatan deep learning yang kompleks. Penelitian yang secara khusus membandingkan algoritma machine learning klasik berbasis data survei terstruktur untuk memprediksi risiko depresi pada mahasiswa masih relatif terbatas. Padahal, data survei mahasiswa dengan instrumen yang valid seperti DASS memiliki potensi besar untuk dikembangkan menjadi sistem prediksi yang praktis, efisien, dan mudah diterapkan di lingkungan perguruan tinggi [7].

Oleh karena itu, penelitian ini dilakukan untuk menerapkan dan membandingkan algoritma Logistic Regression dan Random Forest dalam memprediksi risiko depresi pada mahasiswa. Logistic Regression dipilih karena sifatnya yang sederhana, stabil, dan mudah diinterpretasikan, sehingga memungkinkan identifikasi faktor-faktor yang berkontribusi terhadap risiko depresi. Sementara itu, Random Forest dipilih karena kemampuannya dalam menangani hubungan variabel yang kompleks dan nonlinier serta ketahanannya terhadap overfitting [8]. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem deteksi dini depresi mahasiswa yang berbasis data survei, serta menjadi dasar pengambilan keputusan bagi pihak kampus dan layanan konseling dalam merancang strategi pencegahan dan intervensi kesehatan mental yang lebih tepat sasaran.

1.2 Rumusan Masalah

Bagaimana perbandingan kinerja algoritma Logistic Regression dan Random Forest dalam memprediksi risiko depresi pada mahasiswa berdasarkan metrik evaluasi accuracy, precision, recall, F1-Score dan ROC-AUC?

1.3 Batasan Masalah

1. Penelitian ini menggunakan dataset sekunder yang diperoleh dari platform

Kaggle, dengan variabel yang merepresentasikan faktor gaya hidup, kebiasaan belajar, dan kondisi akademik mahasiswa sebagai prediktor depresi.

2. Algoritma machine learning yang digunakan dibatasi pada Logistic Regression dan Random Forest, tanpa membandingkan dengan metode lain seperti Support Vector Machine atau *deep learning*.
3. Evaluasi performa model hanya difokuskan pada metrik metrik evaluasi accuracy, precision, recall, F1-Score dan ROC-AUC, sedangkan metrik evaluasi lainnya tidak dibahas secara mendalam.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Mengetahui kinerja algoritma Logistic Regression dan Random Forest dalam memprediksi risiko depresi pada mahasiswa berdasarkan data survei dari Kaggle.
2. Menganalisis dan membandingkan performa Logistic Regression dan Random Forest dengan menggunakan metrik akurasi, precision, recall, f1-score dan ROC-AUC.

1.5 Manfaat Penelitian

1. Manfaat Teoritis
 - a. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *machine learning* terapan untuk analisis kesehatan mental, khususnya prediksi depresi mahasiswa.
 - b. Menambah literatur mengenai perbandingan kinerja algoritma Logistic Regression dan Random Forest dalam memprediksi risiko depresi.
 - c. Menjadi acuan bagi penelitian selanjutnya untuk mengembangkan model prediksi yang lebih kompleks, misalnya dengan algoritma *deep learning* atau pendekatan optimasi, guna meningkatkan akurasi serta interpretabilitas.
2. Manfaat Praktis
 - a. Memberikan informasi kepada pihak perguruan tinggi mengenai potensi risiko depresi mahasiswa sehingga dapat digunakan sebagai dasar

penyusunan program pencegahan dan intervensi dini.

- b. Menjadi referensi bagi konselor atau praktisi pendidikan dalam mengidentifikasi faktor-faktor yang paling berpengaruh terhadap depresi mahasiswa berdasarkan hasil prediksi model *machine learning*.
- c. Memberikan gambaran kepada mahasiswa mengenai potensi risiko depresi yang dialami, sehingga dapat meningkatkan kesadaran dan mendorong perilaku preventif untuk menjaga kesehatan mental.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun menjadi lima bab utama, dengan uraian sebagai berikut :

BAB I PENDAHULUAN, berisi Latar Belakang Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan. Bab ini menjelaskan dasar pemikiran dilakukannya penelitian, perumusan permasalahan, ruang lingkup penelitian, tujuan, serta manfaat penelitian yang diharapkan.

BAB II TINJAUAN PUSTAKA, berisi tinjauan Pustaka yang menguraikan penelitian-penelitian terdahulu yang relevan, Dasar Teori yang mendukung penelitian, termasuk teori tentang depresi, konsep *machine learning*, algoritma *Logistic Regression* dan *Random Forest*, serta metrik evaluasi model.

BAB III METODE PENELITIAN, didalamnya terdapat penjelasan metodologi penelitian yang dilakukan, meliputi objek penelitian, alur penelitian yang divisualisasikan dalam bentuk bagan alir (*flowchart*), serta alat dan bahan yang digunakan.

BAB IV HASIL DAN PEMBAHASAN, bab ini menyajikan hasil dari implementasi metode yang telah dirancang. Isinya meliputi hasil preprocessing data, hasil EDA, performa model *Logistic Regression* dan *Random Forest*, evaluasi model, serta analisis perbandingan kedua algoritma.

BAB V PENUTUP, berisi kesimpulan dari penelitian yang menjawab rumusan masalah, serta saran yang ditujukan bagi peneliti selanjutnya.