BAB I

PENDAHULUAN

1.1 Latar Belakang

Permasalahan kesehatan mental pada remaja dan dewasa muda dalam beberapa dekade terakhir menunjukkan kecenderungan peningkatan yang signifikan. Depresi dan kecemasan menjadi dua gangguan psikologis yang paling sering dilaporkan serta berkontribusi besar terhadap penurunan kualitas hidup, gangguan fungsi sosial, serta rendahnya produktivitas akademik [1]. Banyak penelitian menyatakan bahwa sebagian besar gangguan mental mulai berkembang sebelum individu memasuki usia dewasa, sehingga periode remaja hingga mahasiswa merupakan fase krusial untuk deteksi dini dan intervensi yang tepat. Keterlambatan penanganan dapat memperburuk kondisi psikologis, meningkatkan risiko komorbiditas, bahkan berujung pada perilaku menyakiti diri[2].

Kebutuhan terhadap sistem deteksi dini semakin mendesak seiring meningkatnya angka kejadian depresi. Namun demikian, proses identifikasi masih menghadapi berbagai hambatan. Ketersediaan tenaga profesional kesehatan mental yang terbatas, durasi konsultasi yang singkat, serta stigma sosial menyebabkan banyak individu yang membutuhkan bantuan tidak memperoleh layanan yang memadai [1]. Situasi ini menuntut adanya pendekatan alternatif yang mampu membantu proses skrining secara lebih cepat, objektif, dan berbasis data. Keterbatasan dalam sistem layanan kesehatan mental tersebut menunjukkan bahwa pendekatan konvensional saja belum cukup untuk menjawab kompleksitas permasalahan depresi di masyarakat. Oleh karena itu, diperlukan inovasi yang mampu mendukung proses identifikasi risiko secara lebih sistematis dan efisien. Pemanfaatan data yang telah tersedia melalui pendekatan komputasional menjadi salah satu solusi yang relevan untuk menjawab tantangan tersebut [2].

Perkembangan teknologi komputasi membuka peluang pemanfaatan *machine learning (ML)* dalam bidang kesehatan mental. *ML* memiliki kemampuan untuk menganalisis data dalam jumlah besar dan menemukan pola kompleks yang sulit diidentifikasi melalui metode statistik tradisional. Dalam konteks prediksi depresi, algoritma *ML* dapat mempelajari hubungan antara berbagai faktor risiko sehingga mampu mengklasifikasikan individu berdasarkan kemungkinan mengalami gangguan tersebut [2].

Selain penelitian yang menunjukkan potensi *machine learning* dalam mendukung deteksi dini depresi dengan nilai *AUC* sekitar 0,80, sejumlah penelitian lain juga memperkuat efektivitas pendekatan berbasis *ensemble* dalam prediksi gangguan mental [1]. Khan mengembangkan model prediksi depresi pada mahasiswa menggunakan teknik *stacking ensemble* dengan metode *oversampling* untuk menangani ketidakseimbangan data [3]. Hasil penelitian tersebut menunjukkan bahwa kombinasi beberapa algoritma mampu meningkatkan performa klasifikasi dibandingkan model tunggal, sehingga menegaskan pentingnya strategi pemodelan dan pengolahan data dalam membangun sistem prediksi depresi yang akurat.

Selanjutnya, Masud mengintegrasikan *machine learning*, *deep learning*, dan pendekatan *explainable artificial intelligence* dalam deteksi depresi mahasiswa [12]. Dalam penelitian tersebut, algoritma *Random Forest* mampu mencapai akurasi 91,1% dengan F1-score 91,6%, serta memberikan interpretasi model melalui metode SHAP dan LIME [12]. Temuan ini menunjukkan bahwa *Random Forest* tidak hanya unggul dari sisi performa, tetapi juga memiliki kelebihan dalam aspek interpretabilitas, yang sangat penting dalam konteks kesehatan mental. Selain itu, penelitian oleh Akter dalam prediksi dini Alzheimer berbasis data *electronic health record* juga menunjukkan bahwa algoritma berbasis pohon keputusan seperti *Random Forest* memiliki performa yang kompetitif dengan nilai *AUC* di atas 0,80 [11]. Meskipun objek penelitiannya berbeda, hasil tersebut memperlihatkan konsistensi kemampuan *Random Forest* dalam menangani data kesehatan yang kompleks dan multivariat. Berdasarkan beberapa penelitian

terdahulu tersebut, dapat disimpulkan bahwa pendekatan *ensemble*, khususnya *Random Forest*, memiliki dasar empiris yang kuat untuk diterapkan dalam penelitian ini guna membangun model prediksi depresi mahasiswa yang akurat dan andal [2].

Hasil penelitian tersebut juga mengidentifikasi bahwa jenis kelamin, ras, tingkat pendidikan, serta sejumlah indikator medis merupakan prediktor penting terhadap munculnya depresi dan kecemasan [5]. Temuan ini memperlihatkan bahwa kondisi kesehatan mental bersifat multifaktorial, tidak hanya dipengaruhi oleh karakteristik individu tetapi juga oleh lingkungan sosialnya. Meskipun demikian, peneliti menyoroti bahwa penggunaan data pada tingkat komunitas belum sepenuhnya mampu merepresentasikan variasi risiko pada masing-masing individu, sehingga dibutuhkan eksplorasi lebih lanjut menggunakan data yang lebih spesifik.

Sebagian besar penelitian terdahulu memanfaatkan catatan medis elektronik atau data rumah sakit. Pendekatan tersebut memang memberikan gambaran klinis yang kuat, tetapi belum tentu mencerminkan kondisi unik yang dihadapi oleh mahasiswa. Lingkungan pendidikan tinggi memiliki dinamika tersendiri, seperti tekanan akademik, persaingan prestasi, tuntutan penyelesaian studi, perubahan pola hidup, kurangnya waktu istirahat, serta adaptasi sosial. Faktor-faktor tersebut berpotensi menjadi pemicu stres berkepanjangan yang meningkatkan kerentanan terhadap depresi.

Depresi pada mahasiswa tidak hanya berdampak pada aspek psikologis, tetapi juga berkaitan erat dengan performa akademik. Individu yang mengalami depresi cenderung mengalami kesulitan berkonsentrasi, kehilangan motivasi belajar, penurunan kehadiran, hingga risiko putus studi [3]. Oleh sebab itu, institusi pendidikan tinggi memerlukan sistem pendukung berbasis data yang mampu membantu mengidentifikasi mahasiswa berisiko sehingga layanan konseling dapat diberikan secara lebih tepat sasaran.

Dataset yang digunakan dalam penelitian ini menyediakan informasi yang relevan dengan konteks tersebut, meliputi variabel demografis, tekanan akademik, kepuasan belajar, durasi tidur, kebiasaan makan, hingga riwayat keluarga terkait gangguan mental [6]. Keberagaman fitur memungkinkan analisis hubungan yang komprehensif antara faktor personal, akademik, dan gaya hidup. Dengan ukuran data yang besar, pendekatan ML sangat potensial untuk menemukan pola laten yang tidak terlihat melalui pengamatan biasa.

Dalam pengembangan model prediksi depresi, pemilihan algoritma menjadi aspek yang sangat penting karena berpengaruh langsung terhadap akurasi dan kemampuan generalisasi model. Berbagai penelitian terdahulu telah menggunakan beragam algoritma *machine learning*, mulai dari metode klasik seperti *logistic regression*, *support vector machine* (SVM), dan *decision tree*, hingga metode berbasis *ensemble* seperti *Random Forest*, *Gradient Boosting*, XGBoost, dan teknik *stacking ensemble* [6]. Khan menunjukkan bahwa pendekatan *stacking ensemble* yang dikombinasikan dengan teknik *oversampling* mampu meningkatkan performa dibandingkan model tunggal dalam prediksi depresi mahasiswa [3]. Sementara itu, Masud membandingkan beberapa pendekatan termasuk *Random Forest* dan model berbasis *deep learning*, dan menemukan bahwa *Random Forest* tetap memberikan performa yang kompetitif dengan tingkat akurasi di atas 90% [12].

Selain itu, penelitian pada bidang kesehatan lainnya seperti yang dilakukan oleh Akter juga memperlihatkan bahwa metode berbasis *ensemble* secara umum menghasilkan nilai *AUC* yang lebih tinggi dibandingkan metode tradisional [11]. Meskipun dalam beberapa kasus algoritma *boosting* memperoleh nilai sedikit lebih tinggi, *Random Forest* menunjukkan performa yang stabil, tidak mudah mengalami *overfitting*, serta memiliki keunggulan dalam interpretabilitas melalui analisis *feature importance*. Berdasarkan berbagai temuan tersebut, dapat disimpulkan bahwa algoritma berbasis *ensemble*, khususnya *Random Forest*, merupakan metode yang tepat untuk diterapkan dalam penelitian ini guna membangun model prediksi depresi mahasiswa yang akurat dan andal [14].

Random Forest membangun sejumlah besar pohon keputusan dari subset data yang berbeda, kemudian menggabungkan hasilnya untuk memperoleh prediksi akhir [4]. Pendekatan ini efektif dalam meningkatkan akurasi sekaligus mengurangi risiko *overfitting*. Selain itu, *Random Forest* mampu menangani data dengan banyak variabel, tidak mensyaratkan asumsi distribusi tertentu, serta menyediakan informasi mengenai tingkat kepentingan fitur. Kemampuan tersebut menjadikannya sesuai untuk karakteristik dataset mahasiswa yang memiliki kombinasi faktor heterogen dan kemungkinan hubungan nonlinier. Dengan memanfaatkan keunggulan tersebut, *Random Forest* diharapkan tidak hanya menghasilkan model prediksi dengan performa yang baik, tetapi juga memberikan wawasan mengenai faktor-faktor dominan yang mempengaruhi depresi mahasiswa [9]. Informasi ini penting bagi pengambil kebijakan di lingkungan kampus dalam merancang program pencegahan yang lebih efektif.

Berdasarkan uraian di atas, terlihat adanya kebutuhan nyata untuk mengembangkan model prediksi depresi yang kontekstual terhadap populasi mahasiswa. Penelitian terdahulu telah membuktikan efektivitas *ML* pada data remaja secara umum. Namun, kajian berbasis variabel kehidupan akademik sehari-hari masih terbatas. Oleh karena itu, penelitian ini berupaya mengisi kesenjangan tersebut dengan membangun model klasifikasi depresi mahasiswa menggunakan algoritma *Random Forest*. Adapun tujuan dari penelitian ini adalah untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap depresi mahasiswa serta mengevaluasi kemampuan *Random Forest* dalam memprediksi kondisi tersebut. Diharapkan hasil penelitian dapat memberikan kontribusi ilmiah dalam pengembangan penerapan *data science* di bidang kesehatan mental sekaligus memberikan manfaat praktis bagi institusi pendidikan dalam menciptakan sistem deteksi dini yang lebih efektif.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah dari penelitian ini adalah:

1. Bagaimana performa model prediksi depresi mahasiswa menggunakan algoritma *Random Forest* dengan penerapan metode *Synthetic Minority Oversampling Technique (SMOTE)* berdasarkan dataset yang digunakan dalam penelitian?
2. Bagaimana perbandingan performa model Random Forest sebelum dan sesudah penerapan metode SMOTE dalam memprediksi depresi mahasiswa?
3. Faktor atau variabel apa saja yang paling berpengaruh terhadap prediksi depresi mahasiswa berdasarkan hasil pemodelan menggunakan algoritma *Random Forest*?

1.3 Batasan Masalah

Berikut adalah batasan masalah yang diterapkan dalam penelitian ini:

1. Data yang digunakan dalam penelitian ini merupakan dataset mahasiswa yang diperoleh dari platform Kaggle dan bersifat sekunder. Dataset tersebut telah tersedia untuk keperluan penelitian dan analisis, serta tidak melibatkan pengumpulan data primer secara langsung oleh peneliti. Penelitian ini tidak mencakup penambahan data baru di luar dataset yang telah disediakan. Selain itu, data yang digunakan telah melalui proses anonimisasi sehingga tidak memuat informasi identitas pribadi responden, sehingga tetap memperhatikan aspek privasi dan etika penggunaan data.
2. Variabel yang dianalisis terbatas pada atribut demografis, akademik, serta gaya hidup yang terdapat di dalam dataset, seperti usia, jenis kelamin, tekanan akademik, kepuasan belajar, durasi tidur, kebiasaan makan, dan variabel lain yang relevan.
3. Algoritma machine learning yang digunakan untuk membangun model prediksi adalah *Random Forest* sebagai metode utama.
4. Evaluasi kinerja model dilakukan menggunakan metrik performa klasifikasi seperti *accuracy, precision, recall, F1-score,* atau *AUC* sesuai kebutuhan analisis.

5. Penelitian ini tidak membahas diagnosis klinis maupun pemberian rekomendasi terapi, melainkan hanya sebatas pengembangan model prediksi berbasis data.
6. Interpretasi hasil penelitian didasarkan pada keluaran model dan tidak menggantikan penilaian profesional di bidang kesehatan mental.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, maka tujuan dari penelitian ini adalah:

1. Mengetahui performa model prediksi depresi mahasiswa menggunakan algoritma *Random Forest* dengan penerapan metode *Synthetic Minority Oversampling Technique (SMOTE)* berdasarkan dataset yang digunakan dalam penelitian.
2. Menganalisis perbandingan performa model *Random Forest* sebelum dan sesudah penerapan metode *SMOTE* dalam memprediksi depresi mahasiswa.
3. Mengidentifikasi faktor atau variabel yang paling berpengaruh terhadap prediksi depresi mahasiswa berdasarkan hasil pemodelan menggunakan algoritma *Random Forest*.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi dalam pengembangan penerapan *machine learning* pada bidang kesehatan mental, khususnya dalam pemodelan prediksi depresi pada mahasiswa.

2. Menambah wawasan mengenai faktor-faktor yang berpotensi berkaitan dengan munculnya depresi sehingga dapat mendukung upaya deteksi dini secara berbasis data.
3. Menjadi referensi bagi pengembangan penelitian selanjutnya yang berkaitan dengan pemodelan prediksi gangguan mental menggunakan teknik *data mining* atau *machine learning*.
4. Memberikan gambaran mengenai kemampuan algoritma *Random Forest* dalam menyelesaikan permasalahan klasifikasi pada data kesehatan mental.

1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun untuk memberikan gambaran umum mengenai isi setiap bab yang terdapat dalam skripsi. Adapun sistematika penulisan adalah sebagai berikut:

- **Bab I : Pendahuluan**

  Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi.

- **Bab II : Tinjauan Pustaka**

  Bab ini memuat teori-teori yang mendukung penelitian, meliputi konsep depresi pada mahasiswa, pengenalan *machine learning*, metode klasifikasi, algoritma *Random Forest*, serta metrik evaluasi model. Selain itu, bab ini juga menyajikan hasil-hasil penelitian terdahulu yang relevan sebagai landasan dalam melakukan penelitian.

- **Bab III : Metodologi Penelitian**

  Bab ini menjelaskan tahapan pelaksanaan penelitian, mulai dari deskripsi dataset, proses prapengolahan data, pemilihan fitur, pembagian data latih dan data uji, pembangunan model menggunakan algoritma *Random Forest*, hingga metode evaluasi kinerja model.

- **Bab IV : Hasil dan Pembahasan**

Bab ini menyajikan hasil dari proses pemodelan yang telah dilakukan. Pada bab ini akan dibahas performa model berdasarkan metrik evaluasi, analisis variabel yang berpengaruh terhadap prediksi depresi, serta interpretasi hasil yang diperoleh.

- **Bab V : Kesimpulan dan Saran**

  Bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya.