

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesehatan adalah keadaan sejahtera fisik, mental, dan social secara menyeluruh, bukan hanya ketiadaan penyakit atau kelemahan. Organisasi Kesehatan Dunia (WHO) mendefinisikan kesehatan sebagai keadaan kesejahteraan fisik, mental, dan sosial yang lengkap, menekankan bahwa hal itu melampaui tidak adanya penyakit belaka [1]. Kesehatan yang baik memungkinkan individu untuk beraktivitas dengan produktif, berpikir jernih, serta berpartisipasi aktif dalam kehidupan sosial dan ekonomi. Oleh karena itu, kesehatan merupakan elemen kunci yang mendasari keberlangsungan hidup manusia dalam berbagai aspek kehidupan.

Kesehatan yang baik dapat dipengaruhi oleh beberapa faktor baik internal maupun eksternal. Faktor internal sering mencakup atribut individu seperti ketahanan psikologis, efikasi diri, dan kontrol emosional, yang secara signifikan mempengaruhi kesehatan [2]. Sebaliknya, faktor eksternal termasuk dukungan sosial, kondisi ekonomi, dan determinan politik, yang membentuk konteks yang lebih luas di mana individu hidup dan membuat pilihan terkait kesehatan [3], [4]. Salah satu faktor yang sering diabaikan namun memiliki pengaruh signifikan adalah kondisi kesehatan mental, yang dapat dipengaruhi oleh tekanan sosial, lingkungan kerja, dan kualitas relasi interpersonal.

Kesehatan mental mencakup kesejahteraan emosional, psikologis, dan sosial, mempengaruhi bagaimana individu berpikir, merasa, dan bertindak. Ini merupakan bagian integral dari kesehatan secara keseluruhan, mempengaruhi fungsi sehari-hari dan kualitas hidup. Organisasi Kesehatan Dunia (WHO) mendefinisikan kesehatan mental sebagai keadaan sejahtera yang memungkinkan seseorang mengenali kemampuannya, mengatasi tekanan hidup, bekerja secara produktif, dan berkontribusi pada komunitasnya [5]. Saat ini, kesehatan mental mulai dianggap sebagai komponen penting dari kesehatan umum.

Untuk bisa mengatasi masalah kesehatan mental diperlukan sebuah solusi terbaru untuk bisa mengatasi hal tersebut yaitu menggunakan pembelajaran mesin (*machine learning*). *Machine learning* yang merupakan cabang dari *artificial intelligence* memungkinkan sistem belajar dari data tanpa perlu diprogram secara eksplisit dengan cara mempelajari pola-pola yang tersembunyi di dalam data tersebut [6]. Sayangnya, data yang tersedia untuk public pada kumpulan data kesehatan mental merupakan data yang tidak seimbang dan mempunyai nilai yang hilang. Hal ini dapat menyebabkan hasil akurasi klasifikasi *machine learning* yang buruk.

Beberapa algoritma yang populer dalam bidang klasifikasi kesehatan mental seperti: *K-Neighbors Nearest* (KNN), *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Naïve Bayes* (NB), dan *Random Forest* (RF) telah banyak dilakukan dalam bidang ini. Algoritma KNN terkenal dengan kesederhanaan, efektivitas dalam dimensi rendah, dan kemampuan beradaptasi terhadap berbagai ukuran jarak dan distribusi data. Namun, masih memiliki sensitivitas terhadap pilihan K, masalah dimensi tinggi, dan noise dalam data, yang menyebabkan klasifikasi yang salah [7]. Algoritma SVM sangat baik dalam klasifikasi biner yang efektif dan pendekatan penghalusan konvolusi yang diusulkan, yang meningkatkan efisiensi komputasi dan mempertahankan kinerja dalam dimensi tinggi. Tetapi, memiliki sensitivitas terhadap variabel redundan, yang memerlukan pemilihan variabel untuk akurasi [8].

Algoritma LR memiliki kesederhanaan dan efisiensi dalam klasifikasi biner, tetapi efek marjinal rata-rata menawarkan interpretasi yang lebih jelas dan perbandingan yang valid. Namun, algoritma ini masih lemah terhadap interpretasi koefisien yang sulit dan non-kolaps, yang mengarah ke perbandingan model yang tidak valid [9]. Algoritma NB sangat baik dalam hal kecepatan dan kesederhanaan karena asumsi independensi kondisional, meskipun asumsi ini sering salah. Akan tetapi, masih dapat memungkinkan terjadinya peningkatan bias dan potensi ketidakakuratan, meskipun masih dapat bekerja dengan sangat baik, terutama dengan teknik penghalusan [10]. Algoritma RF menawarkan beberapa solusi seperti

pengurangan kompleksitas, waktu pemrosesan yang lebih cepat, dan ketahanan terhadap outlier. Namun, mungkin tidak bekerja sebaik metode yang lebih intensif data dalam skenario kompleks tertentu [11].

Salah satu algoritma yang banyak digunakan dalam klasifikasi masalah kesehatan mental adalah *Extreme Gradient Boost* (XGBoost) dikarenakan algoritma ini memiliki performa yang cepat dan efisien dibandingkan algoritma yang lain. Selain itu juga, Untuk meningkatkan kinerja model dan memastikan hasil yang akurat diperlukan suatu Teknik yang bernama *hyperparameter tuning*. *Hyperparameter tuning* adalah proses memilih nilai optimal dari parameter eksternal model yang tidak dapat dipelajari dari data, seperti jumlah pohon dalam algoritma boosting atau tingkat pembelajaran [12]. Oleh karena itu, penggunaan *hyperparameter tuning* untuk model menjadi suatu solusi yang efektif dalam meningkatkan kinerja model tersebut.

Berdasarkan uraian latar belakang masalah yang telah dipaparkan di atas maka penulis mengatasi masalah tersebut dengan melakukan penelitian yang berjudul “Analisis Penggunaan *Hyperparameter Tuning* Untuk Klasifikasi Data Kesehatan Mental Menggunakan Algoritma *Extreme Gradient Boost*” untuk melakukan proses optimasi menggunakan *hyperparameter tuning* dari algoritma *Extreme Gradient Boost* untuk klasifikasi data kesehatan mental. Peneliti berharap agar hasil penelitian ini bisa menjadi sebuah acuan untuk proses klasifikasi data kesehatan mental yang lebih baik dengan tingkat akurasi yang tinggi sehingga tingkat kesehatan mental seseorang dapat diketahui lebih pasti.

1.2 Rumusan Masalah

Dalam penelitian ini terdapat beberapa rumusan masalah yang ditemukan oleh peneliti, antara lain :

1. Bagaimana pengaruh penggunaan pendekatan Teknik *resampling* dan imputasi data dapat meningkatkan kinerja klasifikasi *machine learning*?
2. Bagaimana pengaruh penggunaan *hyperparameter tuning* terhadap

peningkatan akurasi model XGBoost dalam klasifikasi data kesehatan mental?

1.3 Batasan Masalah

Dalam penelitian kali ini diterapkan beberapa batasan-batasan masalah oleh peneliti, antara lain :

1. Dataset yang digunakan adalah *Mental Health* Dataset yang didapatkan pada website Kaggle dengan rentang waktu dari tahun 2014 hingga 2015 yang berbentuk *commaseparated values (CSV)*.
2. Proses penanganan data hilang yang digunakan pada penelitian ini adalah *K-Nearest Neighbors (KNN)* Imputasi
3. Proses balancing data yang digunakan pada penelitian ini adalah SMOTE (*Synthetic Minority Over-sampling Technique*).
4. Proses normalisasi data yang digunakan pada penelitian ini adalah *MinMaxScalar*
5. Teknik evaluasi model yang digunakan pada penelitian ini adalah *K-Fold Cross Validation*.
6. Teknik evaluasi kinerja akhir yang digunakan pada penelitian ini adalah *confution matrix*.

1.4 Tujuan Penelitian

Dalam penelitian ini terdapat beberapa tujuan yang ingin dicapai oleh peneliti, antara lain :

1. Untuk mengetahui pengaruh penggunaan pendekatan Teknik resampling dan imputasi data dapat meningkatkan kinerja klasifikasi *machine learning*.
2. Untuk mengetahui pengaruh penggunaan *hyperparameter tuning* terhadap peningkatan akurasi model XGBoost dalam klasifikasi data kesehatan mental.

1.5 Manfaat Penelitian

Dalam penelitian ini, terdapat beberapa manfaat yang ingin dicapai oleh peneliti adalah untuk berkontribusi dalam meningkatkan akurasi model klasifikasi data kesehatan mental melalui optimasi hyperparameter tuning. Secara **praktis**, hasil penelitian dapat digunakan untuk mengembangkan sistem pendukung keputusan bagi tenaga medis, khususnya psikolog, dalam mengidentifikasi pola kesehatan mental pasien secara lebih akurat, mempercepat skrining awal, dan mendukung pengambilan keputusan berbasis data. Secara **teoritis**, penelitian ini memperkaya pemahaman tentang efektivitas hyperparameter tuning dalam meningkatkan performa model machine learning serta menjadi referensi bagi studi lanjutan di bidang kesehatan mental berbasis kecerdasan buatan.

1.6 Sistematika Penulisan

Dalam pembuatan naskah penelitian digunakan sistematika penulisan yang disusun sedemikian rupa pada contoh dibawah ini:

BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, tujuan penelitian, batasan masalah, dan manfaat penelitian.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka referensi terdahulu dan dasar teori yang mendasari penelitian.

BAB III METODE PENELITIAN, berisi gambaran metode penelitian yang digunakan serta alat dan bahan untuk mendukung penelitian.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan tahapan penulis menjabarkan hasil penelitian yang telah dilakukan dari rancangan yang telah dibuat.

BAB V PENUTUP, berisi kesimpulan dan saran yang peneliti rangkum selama proses penelitian.