

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Kasus bunuh diri telah menjadi permasalahan kesehatan serius di masyarakat. Menurut Organisasi Kesehatan Dunia (Frijanto, 2022), setiap tahunnya sekitar 800.000 orang meninggal karena bunuh diri, menjadikannya salah satu penyebab utama kematian di seluruh dunia [1]. Menurut data dari WHO, lebih dari 720.000 orang meninggal akibat bunuh diri setiap tahunnya, dan diperkirakan ada 20 percobaan bunuh diri untuk setiap kasus yang berhasil. Pada tahun 2021, bunuh diri tercatat sebagai penyebab kematian ketiga tertinggi di kalangan individu berusia 15-29 tahun secara global, dengan 73% dari seluruh kasus bunuh diri terjadi di negara-negara berpenghasilan rendah dan menengah [2]. Penelitian lain juga menyimpulkan bahwa terdapat peningkatan risiko bunuh diri yang signifikan yang dipengaruhi oleh sejumlah faktor, di antaranya adalah riwayat percobaan bunuh diri sebelumnya, adanya gangguan kesehatan mental, serta kondisi sosial ekonomi yang kurang menguntungkan [3]. Faktor-faktor ini berkontribusi secara signifikan terhadap kecenderungan individu untuk melakukan tindakan bunuh diri. Selain itu, ada juga berbagai elemen lain yang dapat mempengaruhi keputusan seseorang untuk mengambil langkah ekstrem tersebut, menunjukkan bahwa bunuh diri merupakan masalah kompleks yang melibatkan interaksi antara berbagai aspek psikologis, sosial, dan ekonomi [4].

Penelitian sebelumnya menunjukkan bahwa masalah bunuh diri menjadi perhatian serius di seluruh dunia, dengan tren kasus yang terus meningkat, terutama di kawasan Asia [5]. Dengan menggunakan pendekatan Exploratory Data Analysis (EDA) telah mengungkapkan bahwa kelompok usia tertentu, seperti lansia di atas 75 tahun dan anak-anak usia 5 hingga 14 tahun, memiliki resiko bunuh diri yang meningkat secara signifikan. Namun, penelitian ini memiliki beberapa keterbatasan, seperti kustomisasi data yang terbatas dari WHO, rentang waktu data yang tidak mutakhir hingga 2015 [6]. Model Algoritma memiliki berbagai macam jenis, tetapi pada penelitian ini dipilih satu model yang akan digunakan untuk mengukur pentingnya kinerja setiap variabel yang memiliki hubungan signifikan terhadap regresi. Variabel yang signifikan inilah yang nantinya digunakan sebagai fitur untuk membangun model [7].

*Extreme Gradient Boosting (XGBOOST)* adalah metode yang diperkenalkan oleh Chen dan Guestrin pada tahun 2016, yang mengimplementasikan konsep *Gradient Boosting (GB)* dengan cara yang efisien, cepat, dan terukur. *XGBOOST* mengadopsi pendekatan level-wise dalam proses pembangunannya, di mana serangkaian pohon keputusan dibentuk, dan setiap model bergantung pada model sebelumnya. Model awal dalam *XGBOOST* sering kali memiliki kelemahan dalam hal inisialisasi prediksi, tetapi kelemahan ini diperbaiki melalui pembaruan bobot pada setiap model yang dihasilkan. Dengan cara ini, *XGBOOST* tidak hanya mampu memberikan prediksi yang lebih akurat, tetapi juga membantu mencegah model dari *overfitting*, sehingga hasilnya lebih dapat diandalkan dalam berbagai situasi [8].

Dalam penelitian sebelumnya, yang dilakukan oleh (Setiawan, 2021), yang dapat disimpulkan bahwa penelitian tersebut hanya menggunakan pendekatan eksplorasi dengan EDA untuk analisis dan visualisasi data, tanpa mengimplementasikan model machine learning seperti *XGBOOST* [9]. Untuk menghasilkan model perkiraan besaran kasus bunuh diri dengan akurasi yang lebih baik dan dapat diinterpretasikan, penting untuk menggunakan metode machine learning yang tepat, seperti *XGBOOST*. Selain itu, analisis faktor risiko dan variabel yang relevan juga harus dilakukan untuk meningkatkan pemahaman dan interpretasi model [10].

Penelitian lain yang dilakukan pada tahun 2023 menunjukkan bahwa *XGBOOST* memiliki keunggulan dalam memprediksi penyakit *kardiovaskular (CVD)* dibandingkan dengan beberapa algoritma lain seperti *Decision Tree*, *K-Nearest Neighbors*, *Naïve Bayes*, dan *Random Forest*, dengan mencapai akurasi tertinggi sebesar 92,34% [11]. Penelitian lain pada tahun 2024 juga menerapkan metode *XGBOOST* untuk mendeteksi dan menganalisis kanker payudara, dengan hasil akurasi mencapai 94,74% [12]. Pemilihan hyperparameter yang tepat sangat penting untuk meningkatkan kinerja dan akurasi model klasifikasi. *Hyperparameter tuning* memungkinkan peningkatan kinerja model menuju hasil yang optimal dengan mencoba berbagai kombinasi parameter dan nilai secara berulang [13].

Berdasarkan penelitian-penelitian yang telah dibahas sebelumnya, studi ini bertujuan untuk mengoptimalkan kinerja algoritma *XGBOOST* dengan penerapan hyperparameter menggunakan *Randomsearchcv*, guna menemukan pendekatan terbaik dalam memprediksi jumlah kasus bunuh diri dengan metrik MAE, MSE, dan R-squared.

Model yang dikembangkan akan memprediksi jumlah kasus bunuh diri pada tahun berikutnya, dengan mempertimbangkan fitur-fitur seperti negara, tahun, jenis kelamin, usia, jumlah bunuh diri, populasi, dan generasi. Prediksi ini didasarkan pada pola historis dari tahun-tahun sebelumnya untuk masing-masing negara yang dianalisis. Diharapkan, model ini dapat memberikan estimasi yang valid terhadap angka kejadian bunuh diri yang akan datang, sehingga hasilnya dapat digunakan sebagai acuan dalam pengambilan keputusan preventif oleh pihak terkait. Dengan penelitian ini, diharapkan dapat memberikan kontribusi dalam pengembangan model machine learning yang akurat dan efisien untuk memprediksi jumlah kasus bunuh diri, serta memberikan wawasan tentang bagaimana *hyperparameter tuning* dan teknik resampling dapat mempengaruhi kinerja model.

## 1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah diuraikan sebelumnya, permasalahan yang akan menjadi fokus dalam penelitian ini adalah sebagai berikut :

1. Bagaimana kinerja algoritma *XGBOOST* dalam memprediksi jumlah kasus bunuh diri?
2. Bagaimana pengaruh teknik *hyperparameter tuning* terhadap performa model *XGBOOST*?

## 1.3 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Menganalisis performa model *XGBOOST* dalam memprediksi jumlah kasus bunuh diri berdasarkan data dari berbagai benua.
2. Mengoptimalkan performa model dengan menerapkan teknik *hyperparameter tuning* menggunakan *Randomsearchcv*.

#### 1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan, yaitu:

1. Penelitian ini hanya akan membangun model prediksi tingkat bunuh diri menggunakan metode *XGBOOST (XGBRegressor)*. Tidak akan dilakukan perbandingan dengan metode prediksi lainnya dalam penelitian ini.
2. Data yang digunakan berasal dari laporan WHO dalam rentang tahun 1985–2016, sehingga hasil prediksi tidak mencakup tren kasus bunuh diri setelah tahun tersebut.
3. Negara yang dianalisis dipilih berdasarkan ketersediaan data dan tingkat kasus yang signifikan.
4. Dataset terdiri dari 12 kolom dengan besaran dataset sebanyak 27.820.

