

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Di era modern ini, gaya hidup yang kurang sehat menjadi hal yang lazim terutama dengan meningkatnya pola makan yang kurang sehat, junk food, kebiasaan merokok, dan kurangnya aktivitas fisik [1]. Dari berbagai faktor tersebut memicu munculnya penyakit tidak menular atau bisa disebut PTM. Jenis dari penyakit PTM seperti stroke yang terkait dengan pola hidup kurang sehat pada masyarakat di era modern ini [2].

Stroke menjadi penyebab kematian global terbesar ketiga pada 2021 menurut WHO dan utama di Indonesia dengan 19,42% dari total kematian [3]. Proyeksi 2030 memperkirakan stroke tetap dominan di negara berpenghasilan menengah (14,4%) dan ketiga di negara berpenghasilan rendah (8,2%). Seiring tingginya angka ini, teknologi seperti kecerdasan buatan (AI) mulai dimanfaatkan untuk menganalisis data kesehatan secara efisien dan AI kini dapat membantu mendeteksi dini penyakit stroke [4]. *Machine learning* telah menjadi salah satu bidang ilmu yang semakin dikenal luas, termasuk dalam penerapannya di bidang kesehatan sebagai cabang dari *artificial intelligence* (AI) [5].

Penelitian ini menerapkan dua algoritma machine learning, yaitu *Random Forest* dan *Gradient Boosting*, untuk mengklasifikasikan penyakit stroke. Berdasarkan dari berbagai penelitian, model klasifikasi terbukti dapat membantu tenaga medis dalam mendiagnosis dan menunjukkan kemampuan yang efektif dalam melakukan prediksi [6]. *Random Forest* dipilih karena kemampuannya dalam menghasilkan prediksi yang konsisten, sehingga mampu mengurangi risiko *overfitting* pada data yang kompleks [7]. Di sisi lain, *Gradient Boosting* bekerja dengan memperbaiki kesalahan prediksi secara bertahap, menjadikannya efektif dalam menangkap pola pada data yang beragam [8]. Meskipun demikian, tantangan utama yang dihadapi adalah memastikan ketepatan prediksi, yang sering kali

terkendala oleh ketidakseimbangan data dan kompleksitas berbagai faktor risiko stroke [9].

Penelitian ini bertujuan untuk menganalisis pengaruh penerapan teknik SMOTE dalam mengatasi ketidakseimbangan data dan mengoptimalkan algoritma *Random Forest* serta *Gradient Boosting*. Optimalisasi dilakukan melalui penyesuaian *hyperparameter* menggunakan metode *Random Search*. *Random Search* dipilih karena lebih efisien dalam tuning dan waktu komputasinya lebih cepat. Dibandingkan *Grid Search*, metode ini dapat mengeksplorasi ruang parameter lebih luas terutama saat parameter yang diuji besar [10]. Evaluasi kinerja model berdasarkan metrik akurasi, presisi, recall, dan F1 Score. Selain itu, penelitian ini juga memanfaatkan *K-Fold Cross Validation* dan *Confusion Matrix* untuk memastikan hasil evaluasi yang lebih komprehensif dan andal. Dengan kombinasi dari *Random Forest*, *Gradient Boosting*, dan SMOTE, diharapkan model ini mampu meningkatkan akurasi klasifikasi serta mendukung deteksi dini penyakit stroke secara akurat.

## 1.2 Rumusan Masalah

Berdasarkan uraian masalah yang telah dipaparkan, rumusan masalah yang mendasari penelitian ini adalah:

1. Seberapa optimal teknik SMOTE dalam mengatasi ketidakseimbangan data?
2. Bagaimana pengaruh penyesuaian *hyperparameter* pada algoritma *Random Forest* dan *Gradient Boosting* dalam meningkatkan performa klasifikasi penyakit stroke?

## 1.3 Batasan Masalah

Berdasarkan dari latar belakang yang sudah dipaparkan, Adapun Batasan masalah yang mendasari pada penelitian ini, diantaranya sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah *Stroke Prediction Dataset* yang tersedia di Kaggle tahun 2020, dengan total 5.110 data dan 12 kolom

atribut.

2. Penelitian ini menggunakan metrik akurasi, *presisi*, *recall*, dan *F1 Score*, serta metode *K-Fold Cross Validation* dan *Confusion Matrix*.
3. Variabel yang digunakan pada penelitian ini meliputi *gender*, *age*, *hypertensi*, *heart disease*, *ever married*, *work type*, *residence type*, *level glucose*, *bmi*, dan status *smoking*.
4. Penelitian ini berfokus pada teknik Oversampling SMOTE untuk algoritma *Random Forest* dan *Gradient Boost*.
5. Peneliti menggunakan metode *random Search* untuk *hyperparameter tuning*.

#### 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk memberikan solusi terhadap permasalahan dalam klasifikasi penyakit stroke dengan menerapkan algoritma *Random Forest* dan *Gradient Boosting*. Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis seberapa optimal teknik SMOTE dalam menangani ketidakseimbangan data pada klasifikasi penyakit stroke, serta mengevaluasi dampaknya terhadap performa model.
2. Mengevaluasi dan menentukan penyesuaian hyperparameter pada algoritma *Random Forest* dan *Gradient Boosting* untuk meningkatkan performa klasifikasi penyakit stroke.

#### 1.5 Manfaat Penelitian

Penelitian ini bertujuan untuk memberikan solusi terhadap permasalahan dalam klasifikasi penyakit stroke dengan menerapkan algoritma *Random Forest* dan *Gradient Boosting*. Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi di bidang machine learning.

khususnya dalam penerapan metode ensemble untuk bidang kesehatan. Hasil penelitian ini dapat menjadi referensi bagi penelitian lanjutan yang ingin mengeksplorasi dan mengembangkan metodologi baru dalam analisis prediktif di bidang kesehatan yang lebih luas, sehingga dapat mendukung penerapan teknologi berbasis AI di sektor kesehatan secara lebih efektif dan berkelanjutan.

## 2. Manfaat Praktis

Penelitian ini diharapkan dapat memberikan wawasan dalam pengembangan model berbasis machine learning yang lebih optimal melalui penerapan *hyperparameter tuning* dan teknik *SMOTE* dalam bidang kesehatan.

### 1.6 Sistematika Penulisan

Sistematika penulisan yang disusun pada skripsi ini terdapat 5 bab dan pokok-pokok permasalahan diuraikan sebagai berikut:

#### **BAB I : PENDAHULUAN**

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, tujuan dari penelitian, dan manfaat penelitian

#### **BAB II : TINJAUAN PUSTAKA**

Pada bab ini berisi studi literatur dan dasar teori yang digunakan pada penelitian.

#### **BAB III : METODE PENELITIAN**

pada bab ini beris tentang alur penelitian, metode yang digunakan, terakhir alat dan bahan.

#### **BAB IV: HASIL DAN PEMBAHASAN**

Pada bab ini berisi tentang hasil dan pembahasan dari penelitian.

#### **BAB V : PENUTUP**

pada bab ini memuat kesimpulan pada penelitian dan memuat saran guna untuk penelitian selanjutnya

