

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Stroke merupakan salah satu penyebab kematian dan kecacatan tertinggi di dunia. Berdasarkan data World Health Organization (WHO), stroke menempati peringkat kedua penyebab kematian global setelah penyakit jantung [1]. Di Indonesia, prevalensi penyakit stroke terus meningkat setiap tahunnya, terutama akibat gaya hidup tidak sehat, pola makan tidak seimbang, kurangnya aktivitas fisik, serta meningkatnya jumlah penderita hipertensi dan diabetes [2]. Kondisi ini menunjukkan bahwa deteksi dini risiko stroke sangat penting untuk mencegah dampak yang lebih serius pada kesehatan masyarakat.

Upaya pencegahan stroke sangat bergantung pada kemampuan untuk mengidentifikasi faktor risiko secara lebih cepat dan akurat. Selama ini, penilaian risiko pada pasien umumnya bergantung pada pemeriksaan klinis dan evaluasi manual oleh tenaga medis. Namun, pendekatan tersebut memiliki keterbatasan, terutama dalam menangani data dalam jumlah besar dan kompleks. Oleh karena itu, dibutuhkan metode pendukung berbasis teknologi yang mampu memberikan analisis data kesehatan secara lebih efektif dan efisien.

Dalam era digital saat ini, teknologi *Machine Learning* telah banyak dimanfaatkan dalam bidang kesehatan, termasuk untuk prediksi penyakit. *Machine Learning* memungkinkan sistem komputer mempelajari pola dari data historis untuk menghasilkan model prediksi yang dapat digunakan dalam mendukung pengambilan keputusan medis. Pendekatan ini memungkinkan sistem untuk menganalisis data medis pasien seperti tekanan darah, kadar gula darah, indeks masa tubuh, serta riwayat kesehatan untuk mengidentifikasi pola yang tidak mudah terlihat oleh analisis konvensional [3]. Penggunaan *Machine Learning* memungkinkan proses prediksi menjadi lebih cepat, akurat, serta berpotensi meningkatkan kualitas pengambilan keputusan medis.

Salah satu algoritma *Machine Learning* yang banyak digunakan untuk

analisis data tabular adalah Light Gradient Boosting Machine (LightGBM). LightGBM dikenal memiliki performa komputasi yang tinggi, mampu menangani dataset besar, serta menghasilkan akurasi yang kompetitif dibandingkan algoritma *boosting* lainnya [4]. Dengan struktur model berbasis *decision tree* yang dioptimalkan, LightGBM sangat cocok digunakan untuk memproses data kesehatan yang bersifat heterogen dan memiliki kombinasi fitur numerik maupun kategorikal.

Meskipun LightGBM memiliki performa tinggi, model prediksi saja belum cukup dalam konteks medis. Tenaga kesehatan membutuhkan model yang tidak hanya akurat, tetapi juga memberikan penjelasan mengenai alasan di balik prediksi tersebut. Oleh karena itu, *SHapley Additive exPlanations* (SHAP) digunakan untuk memberikan interpretasi terhadap kontribusi masing-masing fitur dalam model [5]. Dengan SHAP, analisis risiko menjadi lebih transparan sehingga membantu tenaga medis memahami faktor-faktor yang memengaruhi hasil prediksi.

Meskipun model Machine Learning mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi, penerapannya dalam bidang kesehatan menghadapi tantangan lain yaitu interpretabilitas model. Pendekatan Explainable Artificial Intelligence (XAI) dapat digunakan untuk meningkatkan interpretabilitas model. Salah satu metode yang banyak digunakan dalam XAI adalah SHapley Additive exPlanations (SHAP). Penggunaan SHAP dalam model prediksi kesehatan dapat meningkatkan tingkat kepercayaan pengguna terhadap teknologi berbasis kecerdasan buatan [6]. Interpretasi yang jelas dan dapat dijelaskan memungkinkan pasien serta profesional kesehatan menerima hasil prediksi dengan lebih baik. Hal ini penting mengingat bahwa keputusan klinis membutuhkan dasar yang kuat dan dapat dipertanggungjawabkan, bukan hanya angka prediksi tanpa penjelasan.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk menerapkan algoritma LightGBM dan analisis SHAP dalam memprediksi risiko penyakit stroke menggunakan dataset kesehatan individu. Melalui kombinasi metode ini, diharapkan diperoleh model prediksi yang akurat, cepat, dan mudah dijelaskan, sehingga dapat memberikan kontribusi nyata dalam mendukung upaya deteksi dini serta pencegahan stroke pada masyarakat.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma LightGBM dalam membangun model prediksi penyakit stroke berdasarkan data Kesehatan individu?
2. Seberapa akurat model LightGBM dalam memprediksi risiko stroke dibandingkan dengan data aktual yang tersedia?
3. Faktor-faktor apa saja yang paling berpengaruh terhadap prediksi risiko stroke berdasarkan hasil analisis menggunakan metode SHAP?

## 1.3 Batasan Masalah

Agar penelitian lebih terarah dan fokus, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan berasal dari Stroke Prediction Dataset yang tersedia di website Kaggle, dengan variabel sebagai berikut: id, gender, age, hypertension, heart\_disease, ever\_married, work\_type, residence\_type, avg\_glucose\_level, bmi, smoking\_status, stroke.
2. Penelitian ini hanya berfokus pada penerapan algoritma LightGBM dan analisis menggunakan SHAP secara global dalam memprediksi risiko stroke tanpa membandingkan dengan algoritma lain.
3. Penelitian ini tidak membahas aspek medis secara mendalam, hanya difokuskan pada analisis data dan permodelan prediktif.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma LightGBM untuk membangun model prediksi risiko stroke berdasarkan data kesehatan individu.
2. Menganalisis tingkat akurasi dan kinerja model dalam mengidentifikasi potensi risiko stroke menggunakan algoritma LightGBM.

3. Mengidentifikasi variabel yang memiliki pengaruh paling signifikan terhadap hasil prediksi risiko stroke menggunakan analisis SHAP.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Penelitian ini dapat menjadi referensi dalam penerapan algoritma LightGBM dan analisis SHAP untuk prediksi penyakit berbasis data medis, khususnya penyakit stroke.
2. Hasil penelitian ini diharapkan dapat digunakan sebagai dasar dalam pengembangan sistem deteksi dini risiko stroke serta menambah literatur dalam bidang penerapan *machine learning* sehingga dapat membantu tenaga medis maupun masyarakat dalam upaya pencegahan penyakit stroke.

### 1.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

1. BAB I PENDAHULUAN, berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
2. BAB II TINJAUAN PUSTAKA, berisi studi literatur dan dasar teori yang berkaitan dengan penelitian, termasuk konsep penyakit stroke, *machine learning*, algoritma LightGBM, metode SHAP dan evaluasi model.
3. BAB III METODE PENELITIAN, berisi uraian mengenai objek penelitian, sumber data, tahapan penelitian, *preprocessing*, penerapan algoritma LightGBM, serta analisis interpretabilitas menggunakan SHAP.
4. BAB IV HASIL DAN PEMBAHASAN, berisi hasil implementasi algoritma LightGBM, visualisasi dan interpretasi SHAP, analisis performa model, serta pembahasan hasil pengujian terhadap dataset stroke.
5. BAB V PENUTUP, berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian di masa mendatang.