

# BAB I

## PENDAHULUAN

### 1.1 Latar belakang masalah

Penyakit yang dikenal sebagai hipertensi atau tekanan darah tinggi adalah suatu kondisi yang bukan berasal dari infeksi, ini menjadi isu besar dalam kesehatan banyak negara, termasuk di Indonesia[1]. Hipertensi ini sering dikenal juga sebagai *silent killer* karena pada gejala awal biasanya tidak menimbulkan tanda-tanda yang cukup jelas, tetapi kondisi seperti ini dapat meningkatkan komplikasi serius seperti penyakit kardiovaskular, stroke, gagal jantung dan jantung koroner[2]. Tingginya angka kejadian serta risiko komplikasi yang ditimbulkan menjadikan hipertensi sebagai masalah kesehatan utama yang memerlukan perhatian khusus[3]. Dalam upaya pencegahan, deteksi dini dan pengendalian secara berkelanjutan. Oleh karena itu, deteksi dini dan klasifikasi yang akurat terhadap status kesehatan pasien menjadi sangat krusial untuk mencegah fatalitas.

Dengan kemajuan teknologi dan informasi dalam bidang kesehatan, penggunaan data kesehatan yang cukup luas sekarang dapat dianalisis melalui pendekatan *machine learning*. Model *machine learning* dapat dimanfaatkan untuk mengenali pola yang rumit dalam data yang sulit dianalisis dengan menggunakan teknik pendekatan statistik konvensional[4]. Sementara itu, data mining merupakan proses interaktif yang melibatkan penggunaan data numerik dan kecerdasan buatan untuk menemukan informasi yang relevan dari sekumpulan data berukuran besar[5]. Namun, dalam pengolahan data kesehatan masih sering dilakukan secara konvensional sehingga potensi hubungan antar berbagai variabel risiko, seperti indikator klinis dan karakteristik individu belum dimanfaatkan secara optimal untuk melakukan prediksi atau klasifikasi secara otomatis dan efisien. Pemanfaatan dataset dari Kaggle memberikan kemudahan bagi peneliti dalam melakukan analisis prediktif serta memungkinkan penelitian dilakukan secara efisien dan dapat direplikasi, khususnya pada kajian prediksi risiko penyakit.

Sejumlah penelitian terdahulu telah menerapkan berbagai metode dalam analisis dan klasifikasi penyakit hipertensi. Metode *Support Machine* (SVM) digunakan karena kemampuannya dalam mengelompokkan data berdasarkan faktor risiko klinis dan gaya hidup, namun penerapannya memerlukan penentuan parameter dan kernel yang kompleks sehingga menyulitkan interpretasi hasil. [6] Selain itu, pendekatan Naive Bayes menawarkan proses komputasi yang sederhana, tetapi asumsi bahwa setiap faktor risiko sering kali tidak sesuai

dengan kondisi data kesehatan yang memiliki hubungan yang kompleks.[7] Di sisi lain, regresi logistik memiliki keunggulan dalam hal interpretabilitas, namun kurang optimal dalam menangani pola nonlinier pada dataset berskala besar[8]. Oleh sebab itu, sangat dibutuhkan pendekatan metode lain yang dapat menangani kompleksitas data kesehatan serta memberikan penjelasan yang jelas tentang faktor apa saja yang mempengaruhi risiko penyakit hipertensi.

Berdasarkan keterbatasan yang ada pada penelitian terdahulu, maka dari itu penelitian ini mengusulkan penerapan model hybrid yang mengombinasikan algoritma Xgboost dengan metode SHAP (*Shapley Additive exPlanations*) dalam analisis dan prediksi risiko penyakit hipertensi. *Extreme Gradient Boosting (XGBoost)* adalah merupakan sebuah algoritma pembelajaran mesin yang didasari oleh pohon keputusan dengan memanfaatkan teknik ensemble seperti hutan acak untuk memperbaiki kemampuan model dalam melakukan prediksi[9]. Pendekatan hybrid ini memadukan keunggulan XGBoost sebagai model prediksi dengan performa yang baik dan kemampuan SHAP dalam memberikan interpretasi terhadap kontribusi setiap variabel[10]. SHAP merupakan metode interpretasi model yang berbasis pada teori permainan kooperatif untuk menilai peran masing-masing fitur dalam hasil akhir prediksi dengan cara yang adil[11]. Metode SHAP menghitung nilai shapley mirip dengan teori permainan koalisi[12].

## **1.2 Rumusan Masalah**

Dengan konteks yang telah diuraikan dalam kajian ini, maka dapat disimpulkan beberapa isu yang relevan sebagai berikut:

1. Bagaimana penerapan model hybrid XGBoost dan SHAP dalam melakukan analisis dan prediksi risiko penyakit hipertensi berdasarkan data kesehatan pasien?
2. Faktor-faktor risiko apa saja yang mendapatkan pengaruh paling dominan terhadap peningkatan risiko hipertensi berdasarkan hasil interpretasi model SHAP?
3. Apakah model hybrid XGBoost-SHAP mampu menghasilkan model prediksi yang akurat sekaligus interpretatif dalam konteks analisis risiko penyakit hipertensi?

## **1.3 Batasan Masalah**

Berdasarkan pertanyaan penelitian yang telah disusun, maka studi ini juga menetapkan batasan-batasan masalah, antara lain:

1. Penelitian ini hanya menggunakan dataset hipertensi yang diperoleh dari data publik kaggle.
2. Variabel yang digunakan hanya pada variabel risiko yang tersedia di dalam dataset tersebut dan tidak melibatkan data tambahan dari luar dataset tersebut.

3. Penelitian ini berfokus pada analisis dan prediksi risiko penyakit hipertensi tanpa membahas aspek pengobatan atau intervensi medis.
4. Metode prediksi yang digunakan dalam penelitian ini dibatasi pada algoritma XGBoost, sedangkan metode interpretasi model menggunakan SHAP.

#### **1.4 Tujuan Penelitian**

1. Menentukan elemen-elemen yang paling berpengaruh terhadap timbulnya penyakit tekanan darah tinggi.
2. Mengetahui kinerja algoritma XGBoost dalam memprediksi risiko penyakit hipertensi berdasarkan data yang digunakan.
3. Menerapkan model hybrid XGBoost dan SHAP dalam analisis dan prediksi risiko penyakit hipertensi menggunakan dataset kesehatan dari Kaggle.
4. Menginterpretasikan hasil prediksi model XGBoost dengan menggunakan metode SHAP untuk mengetahui kontribusi masing-masing variabel risiko.

#### **1.5 Manfaat Penelitian**

Adapun manfaat yang dibuat dalam penelitian ini, diantara lain:

1. Menjadi referensi bagi penelitian selanjutnya yang mengkaji penerapan model hybrid dalam prediksi dan interpretasi risiko penyakit.
2. Membantu memberikan gambaran mengenai faktor-faktor risiko yang berpengaruh terhadap hipertensi berdasarkan analisis data kesehatan.
3. Berkontribusi pada penelitian akademis tentang pengembangan penerapan *machine learning* pada bidang kesehatan, khususnya dalam analisis risiko penyakit hipertensi.

