

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Fenomena *triple burden of diseases* menjadi tantangan besar dalam pembangunan kesehatan global. Kondisi ini mencakup penyakit menular, penyakit tidak menular, dan kemunculan kembali penyakit menular yang sebelumnya terkendali, seperti Malaria, TBC, dan HIV/AIDS. Hipertensi menjadi salah satu masalah kesehatan utama di dunia di antara berbagai penyakit tidak menular. Hipertensi didefinisikan sebagai kondisi ketika tekanan darah sistolik mencapai  $\geq 140$  mmHg dan/atau tekanan darah diastolik mencapai  $\geq 90$  mmHg (Michael *et al.*, 2014).

Organisasi Kesehatan Dunia (WHO) mencatat bahwa hipertensi berperan sebagai faktor risiko utama penyakit kardiovaskular, termasuk serangan jantung, stroke, gagal jantung, dan penyakit ginjal kronis. Pada tahun 2016, jumlah penderita hipertensi global mencapai lebih dari 1,3 miliar orang, atau sekitar 31% populasi dewasa, dengan peningkatan sebesar 5,1% dibandingkan dekade sebelumnya (Bloch, 2016). Prevalensi hipertensi global mencapai 22%, sedangkan kawasan Asia Tenggara menempati posisi ketiga tertinggi dengan prevalensi 25% (Kemenkes RI, 2018).

Hipertensi juga menjadi masalah serius di Indonesia. Kementerian Kesehatan RI (2018) melaporkan bahwa prevalensi hipertensi meningkat dari 25,8% pada tahun 2013 menjadi 34,1% pada tahun 2018. Kelompok usia produktif

menjadi yang paling terdampak, dengan prevalensi hipertensi sebesar 13,2% pada usia 18-24 tahun dan meningkat menjadi 55,2% pada usia 55-64 tahun (Kemenkes RI, 2013). Jawa Tengah mencatat prevalensi hipertensi sebesar 30,4% dari total populasi penduduk berusia  $\geq 15$  tahun di tingkat regional. Kabupaten Sragen menjadi salah satu daerah dengan angka prevalensi tertinggi, yaitu 40,6% atau setara dengan 111.819 kasus pada tahun 2022 (Dinas Kesehatan Jawa Tengah, 2021; Dinas Kesehatan Kota Sragen, 2022).

Gejala hipertensi yang sering samar menyebabkan banyak penderita tidak menyadari kondisinya. Gejala yang muncul meliputi sakit kepala, nyeri dada, sesak napas, hingga mual dan kebingungan (Cita *et al.*, 2022). Hipertensi dipengaruhi oleh faktor yang tidak dapat dikendalikan seperti usia, jenis kelamin, dan keturunan, serta faktor yang dapat dikendalikan seperti obesitas, pola makan tidak sehat, kurang aktivitas fisik, dan kebiasaan merokok (Puspita dan Rismawan, 2019). Pengendalian faktor risiko melalui pencegahan dini menjadi langkah penting untuk menekan angka prevalensi hipertensi (Sihotang *et al.*, 2020).

Analisis data kesehatan dapat membantu mengidentifikasi pola hipertensi lebih dini. Teknologi data *mining* menjadi salah satu solusi dalam menganalisis faktor risiko hipertensi secara komprehensif. Data *mining* adalah proses menemukan pola, hubungan, dan tren dari kumpulan data besar (Zaki & Meira, 2014). Larose (2006) menambahkan bahwa data *mining* menggunakan teknik statistik dan pengenalan pola untuk mengeksplorasi data secara mendalam. Klasifikasi adalah salah satu teknik yang umum digunakan dalam bidang kesehatan

untuk membangun model prediktif. Tujuan dari klasifikasi adalah untuk memetakan risiko hipertensi berdasarkan atribut tertentu (Farid *et al.*, 2014).

Algoritma *k-Nearest Neighbor* (*k*-NN) dan *Naïve Bayes Classifier* sering digunakan dalam metode klasifikasi. Algoritma *k*-NN membandingkan kemiripan data baru dengan data sebelumnya berdasarkan kedekatan tetangga terdekat (Mustafa dan Simpen, 2019; Jumadi, 2015). *Naïve Bayes Classifier*, berbeda dengan *k*-NN, menggunakan pendekatan probabilitas berbasis aturan Bayes dengan asumsi independensi antar atribut (Zheng dan Webb, 2000; Syarifah dan Muslim, 2015). Karakteristik ini membuat *Naïve Bayes* dikenal sebagai metode yang cepat dan akurat, terutama untuk dataset berskala besar.

Evaluasi performa model dilakukan dengan menggunakan metode *K-Fold Cross Validation*, yang membagi dataset menjadi beberapa bagian untuk menguji akurasi model secara lebih mendalam (Cahyanti *et al.*, 2020). Metode ini efektif dalam mengurangi risiko bias selama pengujian model, meskipun memerlukan komputasi yang lebih tinggi (Haltuf, 2014). *K-Fold Cross Validation* diterapkan pada model *k*-NN untuk mendapatkan hasil evaluasi yang lebih akurat.

Beberapa penelitian terdahulu telah membandingkan performa algoritma *k-Nearest Neighbor* (*k*-NN) dan *Naïve Bayes Classifier*. Penelitian yang dilakukan oleh Tempola *et al.* (2018) membandingkan *k*-NN dan *Naïve Bayes Classifier* pada aktivitas status gunung api di Indonesia. Hasil penelitian menggunakan *3-Fold Cross Validation* menunjukkan bahwa metode *k*-NN memiliki rata-rata akurasi 63,68% dengan standar deviasi 7,47%, sedangkan metode *Naïve Bayes* mencapai

rata-rata akurasi 79,71% dengan standar deviasi 3,55%. Akurasi *Naïve Bayes* terbukti lebih konsisten dibandingkan *k*-NN.

Penelitian Delvika *et al.* (2022) membandingkan *k*-NN dan *Naïve Bayes Classifier* untuk prediksi risiko diabetes pada ibu hamil. Data diolah menggunakan teknik *K-Fold Cross Validation* dengan  $K=10$  untuk *Naïve Bayes* dan  $K=25$  untuk *k*-NN. Hasil penelitian menunjukkan akurasi *Naïve Bayes* sebesar 75,78%, sedangkan *k*-NN mencapai akurasi 74,48%. *Naïve Bayes* memberikan performa yang lebih baik dibandingkan *k*-NN.

Penelitian Annisa *et al.* (2023) membandingkan *k*-NN dan *Naïve Bayes Classifier* pada sentimen review aplikasi mobile JKN. Data sebanyak 2.847 sampel diperoleh melalui teknik *scraping* dan dibagi menggunakan *10-Fold Cross Validation*. Hasil penelitian menunjukkan akurasi *Naïve Bayes* sebesar 61,15%, sedangkan *k*-NN memiliki akurasi lebih tinggi, yaitu 87,59%.

Penelitian ini bertujuan untuk membandingkan akurasi antara algoritma *Naïve Bayes Classifier* dan *k-Nearest Neighbor (k-NN)* dalam memprediksi risiko hipertensi pada pasien. Penelitian ini juga akan mengembangkan *dashboard* interaktif untuk memvisualisasikan hasil analisis data. *Dashboard* ini diharapkan dapat membantu tenaga kesehatan dan pemangku kebijakan dalam memahami pola risiko hipertensi serta mengambil keputusan yang lebih tepat. Penelitian ini juga akan melakukan *deployment* aplikasi prediksi risiko hipertensi menggunakan *Streamlit* dengan kedua metode tersebut untuk memungkinkan prediksi yang lebih interaktif dan *real-time*.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini merumuskan masalah sebagai berikut:

1. Bagaimana perbandingan akurasi antara *Naïve Bayes Classifier* dan *k-Nearest Neighbor* dalam memprediksi risiko hipertensi pada pasien?
2. Bagaimana implementasi *dashboard* interaktif untuk menyajikan analisis data risiko hipertensi?
3. Bagaimana proses *deployment* aplikasi prediksi risiko hipertensi dengan metode *Naïve Bayes Classifier* dan *k-Nearest Neighbor* menggunakan *Streamlit*?

## 1.3. Batasan Masalah

Batasan masalah yang digunakan penelitian dalam tugas akhir ini adalah sebagai berikut:

1. Analisis klasifikasi menggunakan metode *Naïve Bayes Classifier* dan *k-Nearest Neighbor* untuk analisis klasifikasi.
2. Data yang digunakan adalah data hipertensi di Puskesmas Gesi, Sragen periode 1 Agustus – 30 September 2024.
3. Variabel penelitian meliputi: jenis kelamin, berat badan, tinggi badan, lingkar perut, tekanan darah sistolik, tekanan darah diastolik, serta faktor risiko (merokok, kurang aktivitas fisik, pola makan gula berlebihan, pola makan garam berlebihan, pola makan lemak berlebihan, kurang makan buah & sayur, konsumsi alkohol), dan status hipertensi.
4. Rasio pembagian data *training* dan *testing* adalah 90:10, 80:20, dan 70:30.

5. Pengolahan data dilakukan menggunakan *Python*, *Microsoft Excel*, dan *SPSS*.
6. Visualisasi data dilakukan melalui *Looker Studio* dalam bentuk *dashboard* interaktif.
7. *Deployment* aplikasi prediksi risiko hipertensi dilakukan menggunakan *Streamlit*.

#### **1.4. Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah dipaparkan, berikut tujuan dari penelitian tugas akhir ini sebagai berikut:

1. Membandingkan akurasi antara *Naïve Bayes Classifier* dan *k-Nearest Neighbor* dalam memprediksi risiko hipertensi pada pasien.
2. Mengembangkan *dashboard* interaktif untuk menyajikan analisis deskriptif data risiko hipertensi.
3. Melakukan *deployment* aplikasi prediksi risiko hipertensi menggunakan *Streamlit* dengan metode *Naïve Bayes Classifier* dan *k-Nearest Neighbor*.