

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hipertensi, atau tekanan darah tinggi, merupakan kondisi medis yang menjadi salah satu penyebab utama kematian di seluruh dunia. Menurut data yang dipublikasikan oleh WHO, diperkirakan lebih dari 1,28 miliar orang dewasa di seluruh dunia mengalami hipertensi, dan angka ini terus meningkat setiap tahun (WHO, 2021). Survei terbaru dari Riskesdas (Riset Kesehatan Dasar) menunjukkan bahwa prevalensi hipertensi pada populasi dewasa mencapai 34,1%, dengan tren peningkatan yang signifikan dari tahun-tahun sebelumnya (Kementerian Kesehatan Republik Indonesia, 2018). Hipertensi, atau tekanan darah tinggi, merupakan kondisi medis yang ditandai dengan peningkatan tekanan darah di atas batas normal, yaitu $\geq 140/90$ mmHg. Menurut definisi dari WHO (2021), hipertensi adalah keadaan tekanan darah sistolik (angka atas) mencapai 140 mmHg atau lebih, atau tekanan darah diastolik (angka bawah) mencapai 90 mmHg atau lebih. Beberapa faktor yang dapat memicu hipertensi meliputi faktor jenis kelamin, usia, penyakit penyerta, BMI, dan genetik. Jenis kelamin menjadi pertimbangan penting karena perbedaan fisiologis antara laki-laki dan perempuan, terutama pengaruh hormonal pada wanita pascamenopause. Usia termasuk sebagai variabel karena risiko hipertensi meningkat seiring penuaan akibat perubahan struktur pembuluh darah. Indeks Massa Tubuh (BMI) dipilih sebagai indikator obesitas yang berhubungan langsung dengan peningkatan tekanan darah. Penyakit penyerta seperti diabetes dan gangguan ginjal turut diperhitungkan karena dapat

memperparah kondisi hipertensi melalui berbagai mekanisme patofisiologis. Faktor genetik dimasukkan mengingat pengaruh riwayat keluarga terhadap kecenderungan hipertensi. Upaya klasifikasi individu berdasarkan status hipertensi diperlukan untuk mendukung strategi pencegahan dan penanganan yang lebih tepat sasaran, mengingat tingginya prevalensi hipertensi serta kompleksitas faktor-faktor penyebabnya. Dalam hal ini, klasifikasi menjadi suatu teknik statistik yang digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu ke dalam kategori yang telah ditentukan. Dalam konteks data mining, klasifikasi termasuk dalam metode *supervised learning* atau pembelajaran terarah yang memanfaatkan data pelatihan. Selama proses pelatihan, model dilatih dengan menggunakan dataset yang sudah diberi label (contohnya, status hipertensi: ya/tidak). Setelah model dilatih, langkah selanjutnya adalah pengujian, yang dilakukan dengan menggunakan data yang belum pernah dilihat sebelumnya untuk menilai tingkat efektifitas model dalam memprediksi kategori yang benar. Data pelatihan ini berfungsi untuk membangun pola atau model tertentu yang akan digunakan untuk meramalkan kelas dari setiap data pengujian. Keakuratan prediksi yang dihasilkan dari pola atau model yang dibuat akan mencerminkan kinerja masing-masing metode klasifikasi yang digunakan. Metode statistika yang dapat diterapkan dalam *supervised learning* antara lain adalah Regresi Logistik Biner dan Algoritma Naïve Bayes.

Pemilihan metode klasifikasi yang tepat sangat bergantung pada karakteristik data yang tersedia. Metode seperti Naive Bayes dan Regresi Logistik sering digunakan untuk menangani kumpulan data yang besar dan kompleks, meskipun metode tersebut memiliki keterbatasan dalam hal akurasi jika tidak

diterapkan dengan benar (Han dan Kamber, 2011). Regresi logistik biner adalah teknik statistik untuk memodelkan hubungan antara satu atau lebih variabel independen dan variabel dependen biner, dengan variabel dependen hanya mengambil dua nilai, yakni 0 atau 1 (Hosmer dan Lemeshow, 2000). Metode ini berfungsi untuk memperkirakan probabilitas terjadinya suatu peristiwa berdasarkan nilai dari variabel independen. Dalam penerapannya, Regresi Logistik Biner memiliki beberapa asumsi yang perlu diperhatikan. Variabel dependen yang digunakan dalam analisis ini harus bersifat dikotomi, sehingga tidak sesuai untuk variabel dengan lebih dari dua kategori. Hubungan yang diasumsikan dalam model ini bukanlah hubungan linear langsung antara variabel independen dan variabel dependen, melainkan hubungan antara variabel independen dengan logaritma *odds* atau perbandingan antara peluang terjadinya dan tidak terjadinya suatu peristiwa. Selain itu, metode ini tidak mengharuskan variabel independen berdistribusi normal maupun memiliki varian yang seragam (*homoskedastisitas*). Di samping itu, data yang digunakan sebaiknya terdiri atas pengamatan yang bersifat independen satu sama lain, guna memastikan bahwa tidak terjadi pengaruh antarobservasi yang dapat mengganggu validitas hasil analisis (Field, 2013).

Metode lain yang digunakan adalah Metode *Naive Bayes*. Metode *Naive Bayes* adalah algoritma klasifikasi probabilistik yang sering digunakan dalam statistik dan pembelajaran mesin untuk mengklasifikasikan data berdasarkan probabilitas. Algoritma ini disebut "naif" karena mengasumsikan bahwa setiap variabel (fitur) yang digunakan sebagai prediktor bersifat independen satu sama lain, padahal pada kenyataannya, hubungan antar-variabel seringkali terjadi dalam banyak kasus dunia nyata. Algoritma *Naive Bayes* memanfaatkan probabilitas

statistika sederhana yaitu dengan menghitung probabilitas kelas dan probabilitas bersyarat menggunakan data *training* untuk kemudian digunakan dalam mengklasifikasikan pengamatan baru. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang kuat (naif) akan independensi dari masing-masing kondisi atau kejadian.

Penelitian terdahulu tentang Metode Regresi Logistik Biner dan *Naïve Bayes* dilakukan oleh Handoko dan Hendry (2023) yang membandingkan Metode *K-Nearest Neighbor*, *Decision Tree*, *SVM*, *Random Forest*, *Naive Bayes* dan *Logistic Regresion* dalam membandingkan ketepatan prediksi dari diabetes Gestasional. Kesimpulan yang dapat diambil adalah Regresi Logistik Biner diketahui memiliki akurasi terbaik dengan nilai sebesar 0,772. Selain itu, Penelitian Fanny, et.al (2023) mengklasifikasikan penerima bantuan sosial Program Keluarga Harapan (PKH) menggunakan metode *Naive Bayes* dan Regresi Logistik Biner. Kesimpulan yang didapat adalah ketepatan klasifikasi dengan metode Regresi Logistik Biner sebesar 75%, sedangkan ketepatan klasifikasi dengan metode *Naive Bayes* sebesar 70%. Dengan demikian, metode Regresi Logistik Biner memberikan hasil klasifikasi yang lebih baik dibandingkan dengan metode *Naïve Bayes*.

Penelitian Tugas Akhir ini membahas analisis perbandingan klasifikasi terhadap pasien Puskesmas Padangsari, Kota Semarang, Jawa Tengah pada bulan Januari – Oktober tahun 2024. Metode yang digunakan adalah Regresi Logistik Biner dan algoritma *Naive Bayes*, dengan menghitung ketepatan klasifikasinya menggunakan *Confusion Matrix*. Penelitian ini diharapkan dapat mengklasifikasikan status pasien penyakit hipertensi dengan nilai sensitivitas tertinggi.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang, maka rumusan masalah yang akan dibahas adalah sebagai berikut:

1. Bagaimana penerapan metode Regresi Logistik Biner dan metode Naïve Bayes dalam mengklasifikasikan status pasien penderita hipertensi?
2. Faktor risiko apa saja yang secara signifikan memengaruhi status hipertensi berdasarkan metode Regresi Logistik Biner dan metode Naïve Bayes?
3. Bagaimana ketepatan klasifikasi dan perbandingan nilai *Area Under Curve* (AUC) pada grafik *Receiver Operating Characteristic* (ROC) untuk mengevaluasi performa metode Regresi Logistik Biner dan Naïve Bayes dalam mengklasifikasikan status pasien penderita hipertensi?

1.3 Batasan Masalah

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

1. Variabel yang digunakan pada penelitian ini yaitu usia, jenis kelamin, BMI (*body mass index*), penyakit penyerta, genetik (keturunan), serta klasifikasi (normal dan Hipertensi).
2. Membagi dan membandingkan setiap data menjadi data latih dan data uji dengan perbandingan data latih dan data uji dipilih, yaitu 70% : 30%; 80% : 20%.; dan 90% : 10%.

1.4 Tujuan

Berdasarkan perumusan masalah yang telah ditentukan mengenai tujuan pada penelitian Tugas Akhir ini adalah sebagai berikut:

1. Menganalisis penerapan metode Regresi Logistik Biner dan metode Naïve Bayes dalam mengklasifikasikan status pasien penderita hipertensi.
2. Mengidentifikasi faktor-faktor risiko yang berpengaruh signifikan terhadap status hipertensi berdasarkan analisis metode Regresi Logistik Biner dan metode Naïve Bayes. Mengidentifikasi faktor-faktor risiko yang berpengaruh signifikan terhadap status hipertensi pasien dalam model Regresi Logistik Biner dan Naive Bayes.
3. Mengevaluasi ketepatan klasifikasi serta membandingkan nilai *Area Under Curve* (AUC) pada grafik *Receiver Operating Characteristic* (ROC) sebagai indikator performa kedua metode dalam mengklasifikasikan status pasien penderita hipertensi.