

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Hipertensi adalah kondisi medis serius yang sering kali tidak menunjukkan gejala, namun dapat menyebabkan komplikasi kesehatan yang berbahaya jika tidak diobati. Kondisi ini ditandai dengan tekanan darah yang konsisten melebihi batas normal pada dinding arteri. Kondisi ini menyebabkan jantung harus bekerja lebih keras untuk mengalirkan darah ke seluruh tubuh melalui pembuluh darah. Jika tekanan darah tinggi tidak dikendalikan dalam waktu yang lama, hal ini dapat menimbulkan berbagai komplikasi serius, termasuk terganggunya aliran darah, kerusakan pada pembuluh darah, serta meningkatnya risiko terkena penyakit degeneratif. Kondisi ini bisa menimbulkan dampak serius apabila tidak ditangani dengan benar [1].

Hipertensi, merupakan salah satu penyakit yang sering dialami oleh individu dalam rentang usia 50 hingga 60 tahun, dengan prevalensi mencapai 55,2% [2]. Hipertensi dapat juga diartikan sebagai kondisi dengan tekanan darah secara konsisten melebihi batas normal. Tekanan darah sistolik dikatakan berada pada ambang batas atas jika mencapai 130 mmHg, sedangkan untuk tekanan darah diastolik batas atasnya adalah 80 mmHg [3] Hipertensi bukan hanya masalah kesehatan fisik, tetapi juga menjadi beban psikologis dan jasmani bagi penderita. Gejalanya yang sering kali tidak terlihat secara langsung dapat menyebabkan kecemasan dan ketegangan yang berkelanjutan. Tekanan darah tinggi yang tidak terkontrol dapat memberikan dampak langsung dan berbahaya bagi organ vital seperti ginjal, jantung, dan otak [4]. Oleh karena itu, kondisi ini dapat diminimalisir dengan menerapkan gaya hidup sehat, salah satunya melalui konsumsi makanan bergizi dan rendah garam, menurunkan berat badan, berolahraga secara teratur, dan berhenti merokok [5].

Data dari *World Health Organization* (WHO) menunjukkan bahwa di tahun 2015, terdapat sekitar 1,13 miliar individu di seluruh dunia yang mengalami kondisi hipertensi. Jumlah ini mengalami kenaikan secara berkelanjutan, menunjukkan

bahwa hipertensi telah menjadi permasalahan kesehatan global yang perlu diwaspadai. Jika situasi dan kesadaran individu yang tidak mengalami perbaikan, maka pada tahun 2025 *World Health Organization* mengungkapkan bahwa akan ada sebanyak 1,5 miliar orang yang akan menjadi korban dari penyakit ini [2].

Pada data Riset Kesehatan Dasar (Riskesdas) menyatakan bahwa pada tahun 2018, terdapat 658.201 orang berusia di atas 18 tahun di Indonesia yang menderita hipertensi. Provinsi Jawa Barat memiliki jumlah penderita terdiagnosis tertinggi, yaitu 131.153 orang, sedangkan Provinsi Kalimantan Utara memiliki jumlah penderita terdiagnosis terendah, yaitu 1.675 orang [6].

Dengan melakukan analisa lebih cepat dengan hasil yang akurat dan tepat dalam dunia medis sangatlah membantu tenaga medis dalam menangani atau memberikan instruksi yang lebih tepat terhadap pasien. Oleh sebab itu dibutuhkan pemanfaatan teknologi kecerdasan buatan atau *machine learning* yang membantu untuk memprediksi penyakit hipertensi secara tepat dan akurat untuk kepentingan dalam memastikan kesehatan masyarakat [7].

Penelitian ini membandingkan dua algoritma pembelajaran mesin yang sering digunakan, adalah *Support Vector Machine* dan *Random Forest*, pada konteks prediksi hipertensi. *Support Vector Machine* dikenal karena keandalannya dalam pengelolaan data yang kompleks dan memiliki dimensi tinggi dengan menentukan *hyperplane* terbaik untuk memisahkan antara dua kelas [8]. Sebaliknya, *Random Forest* adalah metode *ensemble learning* dengan menggabungkan prediksi dari berbagai pohon keputusan secara acak, sehingga mampu mengatasi masalah *overfitting* dan meningkatkan akurasi prediksi [9].

Di bidang kesehatan, penting untuk memiliki model prediksi mampu menghasilkan hasil yang akurat serta mampu diandalkan, sehingga dapat membantu dalam diagnosis dini, pengobatan yang lebih efektif, dan pencegahan penyakit. Dengan membandingkan kinerja kedua algoritma ini secara langsung, dapat menentukan algoritma yang paling baik dalam memprediksi hipertensi. Pendekatan ini diharapkan mampu memberikan kontribusi signifikan dalam upaya pengembangan metode prediksi hipertensi yang lebih modern dan presisi, yang

pada akhirnya dapat meningkatkan pencegahan dan manajemen kondisi kesehatan masyarakat secara umum.

Pada penelitian ini, peneliti akan mengatasi permasalahan ketidakseimbangan kelas. Metode *resampling* yang diterapkan dalam upaya prediksi hipertensi adalah *oversampling*. Metode tersebut akan diuji menggunakan dua algoritma pembelajaran mesin, adalah *Support Vector Machine* (SVM) serta *Random Forest*. Pada penelitian sebelumnya yaitu berjudul “Pengaruh *Random Undersampling*, *Oversampling*, serta SMOTE terhadap Kinerja Model Prediksi Penyakit Kardiovaskular”. Penelitian ini mengevaluasi dampak teknik *resampling* seperti *Random Undersampling*, *Oversampling*, serta SMOTE pada model prediksi untuk penyakit kardiovaskular menggunakan algoritma SVM, *Random Forest*, dan *LASSO Logistic Regression*. Hasilnya membuktikan bahwa penerapan teknik SMOTE secara nyata memperbaiki performa model, terutama dalam algoritma *Random Forest* dalam hal akurasi, sensitivitas, dan spesifisitas [10] Kemudian penelitian lainya yaitu [11] mengenai pengembangan model klasifikasi kualitas udara berdasarkan data Indeks Standar Pencemaran Udara (ISPU) DKI Jakarta tahun 2022. Menggunakan penerapan teknik *Oversampling* SMOTE, akurasi model *Random Forest* meningkat dari 98% menjadi 99%, dan akurasi SVM meningkat dari 91% menjadi 95% [11]

Pada Penelitian [12] membandingkan efektivitas tiga teknik penyeimbangan kelas adalah *Random Oversampling* (ROS), *Random Undersampling* (RUS), dan SMOTE-NC (*Synthetic Minority Over-sampling Technique* for Nominal and Continuous) pada model klasifikasi *Random Forest* (RF) serta *Support Vector Machine* (SVM). Hasilnya memperlihatkan bahwa *Random Undersampling* memberikan performa terbaik pada model *Random Forest* serta *Support Vector Machine*, dengan peningkatan nilai *Accuracy*, *Sensitivity*, dan *Specificity* dibandingkan dengan teknik lainnya [12] Kemudian penelitian [13] mengevaluasi efektivitas teknik *resampling* (ROS, RUS, dan SMOTE) pada klasifikasi data pendidikan yang tidak seimbang menggunakan algoritma RF. Hasil menunjukkan bahwa SMOTE meningkatkan kinerja model dari segi akurasi serta nilai *F1-Score* apabila dibandingkan dengan teknik *resampling* lainnya [13]

Berdasarkan penjelasan latar belakang diatas penelitian ini bertujuan untuk mengevaluasi efektivitas kinerja algoritma *Support Vector Machine* serta *Random Forest* dalam memprediksi hipertensi. Selain itu, dapat menentukan kinerja algoritma yang paling efektif dalam prediksi hipertensi.

## 1.2 Rumusan Masalah

Mengacu pada latar belakang yang telah dipaparkan, hipertensi termasuk dalam kategori penyakit tidak menular yang berisiko tinggi, karena kerap kali tidak menimbulkan gejala pada tahap awal, namun dapat menimbulkan komplikasi serius pada jantung, ginjal, otak, dan pembuluh darah. Tingginya jumlah penderita hipertensi, baik secara global maupun di Indonesia, menunjukkan bahwa penyakit ini telah menjadi masalah kesehatan masyarakat yang mendesak. Maka dari itu, dibutuhkan metode yang lebih efisien guna mendeteksi hipertensi secara dini agar penanganan dapat dilakukan lebih cepat dan tepat. Salah satu tantangan utama dalam membangun model prediksi hipertensi adalah ketidakseimbangan kelas pada data medis. Kondisi tersebut berpotensi menyebabkan model prediksi menjadi tidak seimbang dan menurunkan tingkat akurasi. Untuk mengatasi hal tersebut, pendekatan berbasis *machine learning*, khususnya algoritma seperti *Support Vector Machine* (SVM) dan *Random Forest* (RF), dapat dimanfaatkan karena keduanya memiliki kemampuan yang baik dalam klasifikasi. Merujuk pada permasalahan yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana kinerja algoritma *Support Vector Machine* (SVM) dalam memprediksi kondisi hipertensi?
2. Bagaimana kinerja algoritma *Random Forest* dalam melakukan prediksi hipertensi?
3. Dari kedua algoritma tersebut, algoritma pembelajaran mesin mana yang paling baik digunakan dalam memprediksi hipertensi?

### 1.3 Batasan Masalah

Berdasarkan rumusan masalah serta tujuan penelitian, ditetapkan ruang lingkup pembahasan guna memastikan bahwa penelitian tetap berada dalam konteks yang relevan, yaitu sebagai berikut:

1. Penelitian ini menggunakan dataset dari Puskesmas Purwokerto Timur I.
2. Variabel yang digunakan yaitu : usia, kategori, jenis kelamin, tinggi badan, berat badan, *Sistole*, *Diastole*, dan diagnosis.

### 1.4 Tujuan dan Manfaat

Penelitian ini bertujuan dan bermanfaat untuk:

1. Mengetahui kinerja algoritma *Support Vector Machine* (SVM) dalam memprediksi kondisi hipertensi.
2. Mengetahui kinerja algoritma *Random Forest* dalam melakukan prediksi hipertensi.
3. Mengetahui dari kedua algoritma tersebut mana yang lebih baik dalam memprediksi hipertensi.

### 1.5 Metode Penelitian

Penelitian ini menerapkan metode kuantitatif dengan pendekatan data *mining* dalam membangun model prediksi penyakit hipertensi menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest*. Data yang digunakan berasal dari rekam medis pasien di Puskesmas Purwokerto Timur I.

Tahapan-tahapan dalam pelaksanaan penelitian ini mencakup:

1. Studi Literatur : mempelajari lebih lanjut mengenai algoritma *Support Vector Machine*, *Random Forest*, dan *hipertensi*.
2. Pengumpulan Data: Dataset dikumpulkan dari bagian Rekam Medis di Puskesmas Purwokerto Timur I.
3. Pra-pemrosesan Data (*Preprocessing*): Dataset yang telah diperoleh akan dibersihkan dari data duplikat, nilai kosong, dan *outlier*.

4. Pemodelan: Algoritma *Support Vector Machine* (SVM) dan *Random Forest* digunakan sebagai metode utama dalam membangun model klasifikasi dibantu dengan *pycharm*.
5. Evaluasi Model: Model yang telah dikembangkan akan dinilai menggunakan sejumlah metrik performa, antara lain akurasi, presisi, *recall*, serta *F1-Score*.
6. Analisis Hasil: Hasil evaluasi akan dianalisis untuk mengetahui seberapa baik model mampu memprediksi risiko *hipertensi*.
7. Penyusunan Laporan : Hasil dari seluruh penelitian ditulis dalam bentuk laporan tugas akhir.

### 1.6 Jadwal Pelaksanaan

Penulisan Tugas Akhir ini disusun selama 6 Bulan dengan langkah-langkah yang mulai dari Studi Literatur, Pengumpulan Data, Pra-pemrosesan Data, Pemodelan, Evaluasi Model, Analisis Hasil, dan Penyusunan Laporan.

Tabel 1. 1 Rancangan Jadwal Kegiatan Penelitian

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan Data						
3	Pra-pemrosesan Data						
4	Pemodelan						
5	Evaluasi Model						
6	Analisis Hasil						
7	Penyusunan Laporan/Buku TA						