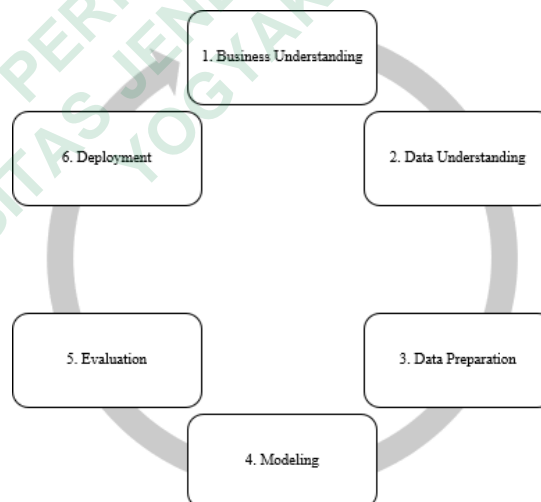


## BAB III METODE PENELITIAN

### A. Desain Karya Tulis Ilmiah

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan *data mining*. Penelitian ini menggunakan data sekunder berupa data pemeriksaan terakhir pasien yang tercatat pada periode Januari-Desember 2024. Teknik analisis yang digunakan adalah klasifikasi dengan Algoritma *Decision Tree* untuk mengelompokkan pasien berdasarkan tingkat resiko hipertensi. Menurut penelitian Risdiana., (2022) Penelitian kuantitatif adalah suatu pendekatan penelitian yang berfokus pada penggunaan angka, mulai dari tahap pengumpulan data, estimasi data hingga penyajian hasil. Penelitian kuantitatif memiliki karakteristik khas yang berfokus pada perhitungan, angka, atau kuantitas. Pada penelitian ini menggunakan alur *data mining life cycle* sebagai berikut (Hidayati et al., 2021):



Gambar 3. 1 Tahapan *Cross-Industry Standard Process for Data Mining*

Pendekatan CRISP-DM terdiri dari enam tahapan utama sebagaimana ditampilkan pada gambar diatas. Keenam tahapan tersebut meliputi: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* dan *Deployment*.

### 1. *Business Understanding*

Proses untuk memahami permasalahan bisnis, menetapkan pendekatan yang akan digunakan untuk menyelesaikan masalah tersebut, mengidentifikasi kebutuhan data serta merancang rencana manajemen proyek (Wilandini., 2022) Penelitian ini bertujuan untuk membantu mengelompokkan data penyakit hipertensi serta mengidentifikasi risiko hipertensi dengan memanfaatkan algoritma *Decision Tree*.

### 2. *Data Understanding*

Proses ini dilakukan dengan mengidentifikasi atribut di lembar data. Atribut dilakukan dalam proses penyortiran awal, yang melibatkan menentukan atau memilih atribut yang tidak memiliki koneksi untuk membuat model (Sholeh et al., 2023). Penelitian ini menggunakan data kasus hipertensi pasien yang diperoleh dari rekam medis di Puskesmas Godean 1

### 3. *Data Preparation*

Fase pengolahan data melibatkan proses perhitungan yang optimal. Pada tahap ini dilakukan pembersihan data (*data cleaning*) untuk menghapus informasi atau data yang tidak relevan agar tidak mengganggu proses pengolahan data (Rahmadayanti et al., 2023). Tahap pengolahan data mencakup proses *data cleaning* pada rekam medis kasus hipertensi guna menghilangkan elemen-elemen yang tidak diperlukan dalam pengolahan data

### 4. *Modeling*

Menerapkan teknik pemodelan yang tepat serta mengenali dan menampilkan pola-pola yang ditemukan (Yudiana et al., 2023). Penelitian ini menggunakan teknik *data mining* berupa *Clustering* dan algoritma *Decision Tree* untuk proses analisis data.

### 5. *Evaluation*

Pada tahap ini, dilakukan evaluasi terhadap semua langkah yang telah dilaksanakan berdasarkan tujuan awal yang telah ditetapkan (Fitrianti et al., 2023).

Tahapan evaluasi dalam penelitian ini adalah :

- a. *Accuary* (Akurasi)
  - b. *Precision* (Presisi)
  - c. *Recall* (*Sensitivitas* atau *True Postive Rate*)
  - d. *FI-Score*
6. *Deployment*

Tahapan di mana aplikasi yang telah selesai dibuat dan diuji akan secara otomatis ditempatkan di server produksi setelah repositorinya terhubung dengan server (Irvan et al., 2024).

## **B. Lokasi dan Waktu Kegiatan**

### **1. Lokasi Penelitian**

Penelitian ini dilaksanakan di Puskesmas Godean 1 yang berlokasi di Jl. Pandean 7, Sidoluhur, Godean, Pandean, Sidoluhur, Sleman, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55264.

### **2. Waktu Penelitian**

Penelitian ini dilaksanakan pada bulan Februari sampai Juli 2025 di Puskesmas Godean 1.

## **C. Populasi dan Sampel**

### **1. Populasi**

Populasi adalah seluruh objek atau subjek dalam suatu wilayah yang memenuhi kriteria tertentu yang relevan dengan penelitian (Suriani et al., 2023). Populasi dalam penelitian ini adalah seluruh data rekam medis pasien rawat jalan hipertensi di Puskesmas Godean 1 pada periode 1 Tahun pada bulan Januari - Desember 2024. Berdasarkan data yang diperoleh dari bagian rekam medis terdapat sebanyak 8.058 kasus hipertensi selama periode tersebut.

### **2. Sampel**

Berdasarkan (Sugiyono, 2021) Pada penelitian kuantitatif sampel merupakan sebagian kecil yang mewakili jumlah dan sifat-sifat yang terdapat dalam suatu populasi. Hal ini didukung oleh penelitian Asrulla et al., (2023)

Sampel merupakan bagian atau perwakilan dari populasi yang diteliti dan digunakan sebagai sumber data. Teknik pengambilan sampel yang digunakan adalah *Total Sampling*. Penelitian oleh Melda et al., (2020) Total sampling adalah teknik pengambilan sampel dengan menggunakan seluruh anggota populasi sebagai sampel penelitian. Sampel dalam penelitian ini adalah data rekam medis pasien hipertensi yang memenuhi kriteria inklusi sebagai berikut:

- a. Data rekam medis rawat jalan kasus hipertensi
- b. Data pemeriksaan terakhir pasien yang tercatat pada periode Januari hingga Desember 2024
- c. Memiliki catatan lengkap mengenai tekanan darah sistolik dan diastolik
- d. Memiliki informasi lengkap mengenai umur, jenis kelamin dan IMT
- e. Tidak terdapat duplikat data
- f. Tidak terdapat data yang hilang (*Missing Data*)

Dalam penelitian ini, terdapat 2.087 sampel kasus hipertensi yang terdiri atas 774 pasien laki-laki dan 1.313 pasien perempuan.

#### **D. Variabel Penelitian**

Berdasarkan (Sugiyono, 2021) Secara teori variabel dapat diartikan sebagai karakteristik dari seseorang atau suatu objek yang menunjukkan perbedaan antara individu satu dengan yang lainnya atau antara satu objek dengan objek lainnya. Hal ini juga didukung oleh penelitian (Loliyana et al., 2023) Variabel penelitian mencakup segala aspek yang ditetapkan oleh peneliti untuk dikaji guna memperoleh informasi relevan yang kemudian dianalisis untuk menarik kesimpulan. Dalam penelitian ini, terdapat dua variabel, yaitu:

##### **1. Variabel Independen**

Variabel independen merupakan variabel yang memberikan pengaruh atau menjadi faktor penyebab terjadinya perubahan pada variabel dependen (Dekanawati et al., 2023). Variabel ini diperoleh dari data rekam medis dan digunakan sebagai *input* dalam algoritma *Decision Tree* sebagai berikut:

- a. Jenis Kelamin (Laki-laki /Perempuan)
- b. Usia (dalam tahun)

- c. Sistolik (mmHg)
- d. Diastolik (mmHg)
- e. Indeks Masa Tubuh (IMT)

## 2. Variabel Dependen

Variabel dependen adalah variabel yang terpengaruh atau menjadi hasil dari adanya variabel bebas. Dengan kata lain variabel dependen merupakan variabel yang dipengaruhi oleh variabel independen (Dekanawati et al., 2023). Variabel dependen merupakan hasil klasifikasi yang menjadi target keluaran dari algoritma *Decision Tree*, yaitu status hipertensi pasien sebagai berikut:

1. Normal
2. Prahipertensi
3. Hipertensi Tingkat 1
4. Hipertensi Tingkat 2

## E. Definisi Operasional

Tabel 3. 1 Definisi Operasional

No	Variabel	Definisi	Alat Ukur	Hasil Ukur	Skala Ukur
1	Jenis Kelamin	Kategori jenis kelamin pasien yang tercatat dalam rekam medis (Laki-laki / Perempuan)	Data Rekam Medis Pasien	1. Laki-laki 2. Perempuan	Nominal
2	Usia	Umur pasien dalam tahun yang dihitung dari tanggal lahir yang tercatat dalam rekam medis	Data Rekam Medis Pasien	Angka dalam tahun	Rasio
3	Tekanan Sistolik	Tekanan darah bagian atas saat jantung berkontruksi, diukur dalam mmHg	Data Rekam Medis Pasien	Angka dalam satuan mmHg	Rasio
4	Tekanan Diastolik	Tekanan darah bagian bawah saat jantung, diukur dalam	Data Rekam Medis Pasien	Angka dalam satuan mmHg	Rasio

No	Variabel	Definisi	Alat Ukur	Hasil Ukur	Skala Ukur
		rekam medis mmHg			
5	Indeks Masa Tubuh (IMT)	Berat badan pasien dalam kilogram	Data Rekam Medis Pasien	1. Underweight 2. Normal 3. Overweight 4. Obesitas I 5. Obesitas II	Nominal
6	Status Hipertensi	Klasifikasi tekanan darah pasien sesuai standar medis	Data <i>Excel</i> berdasarkan hasil pengukuran tekanan darah	1. Normal 2. Prehipertensi 3. Hipertensi tingkat 1 4. Hipertensi tingkat 2	Nominal

## F. Alat dan Metode Pengumpulan Data

### 1. Alat Pengumpulan Data

Alat pengumpulan data yang digunakan pada penelitian ini sebagai berikut:

- a. Dokumen Rekam Medis: Berkas *Excel* yang berisi data riwayat kesehatan pasien di Puskesmas Godean 1, khususnya data pasien yang telah didiagnosis hipertensi
- b. Formulir Ekstraksi Data: Lembar kerja *Excel* untuk mencatat variabel-variabel yang diperlukan data data rekam medis pasien (usia, jenis kelamin, tekanan darah sistolik tekanan darah diastolik dan IMT)
- c. Laptop: Perangkat yang digunakan untuk proses digitalisasi data (*preprocessing*) serta penerapan algoritma *Decision Tree* menggunakan *software*
- d. *Software* pengolahan data: Aplikasi seperti *Microsoft Excel 2019* dan *RapidMiner Version 10,3*

### 2. Metode Pengumpulan Data

Penelitian ini menggunakan metode dokumentasi dan ekstraksi datasekunder yang mencakup beberapa tahapan sebagai berikut:

a. Identifikasi Data

Mengidentifikasi seluruh pasien dengan diagnosis hipertensi berdasarkan data rekam medis rawat jalan di Puskesmas Godean 1 pada periode Januari - Desember 2024.

b. *Ekstraksi* Data

Mengambil informasi yang relevan dari rekam medis menggunakan formulir ekstraksi yang telah disesuaikan dengan variabel penelitian seperti Jenis Kelamin (Laki-laki /Perempuan), Usia (dalam tahun), Sistolik (mmHg), Diastolik (mmHg) dan IMT.

c. Perhitungan Indeks Massa Tubuh (IMT)

Menghitung nilai IMT berdasarkan data berat badan dan tinggi badan pasien sebagai salah satu variabel klasifikasi.

d. Pembersihan Data

Memastikan data yang digunakan bebas dari kesalahan dan ketidaksesuaian dengan menghapus atau memperbaiki data yang tidak lengkap atau tidak valid dan menghapus variabel yang tidak dibutuhkan seperti berikut :

- 1) Tanggal
- 2) No Rm
- 3) NIK
- 4) Nama (Disamarkan)
- 5) Nama Orangtua
- 6) Tanggal Lahir
- 7) Alamat
- 8) No Bpjs
- 9) Jenis Kunjungan
- 10) No Telephone
- 11) Jenis Pembayaran
- 12) Tekanan Nadi
- 13) Clinic
- 14) Obat

15) Dokter

16) Perawat

e. Pengkodean Data

Mengubah variabel kategorikal seperti jenis kelamin, IMT dan Status Hipertensi ke dalam bentuk numerik untuk mempermudah proses analisis menggunakan algoritma *Decision Tree*.

f. Penyimpanan Data

Menyimpan seluruh data yang telah diproses dalam format digital pada file *Excel* agar siap digunakan dalam tahap analisis lebih lanjut.

### G. Metode Pengolahan dan Analisis Data

Pada karya tulis ilmiah ini analisis data akan menggunakan *RapidMiner Version 10,3* yang mendukung penggunaan algoritma *Decision Tree* untuk mengklasifikasikan status hipertensi berdasarkan data rekam medis pasien. Adapun langkah-langkahnya sebagai berikut:

1. Pra-pemrosesan Data (*Data Preprocessing*)

Langkah awal dalam analisis data adalah melakukan pra-pemrosesan untuk memastikan data siap digunakan dalam model *Decision Tree*. Tahapan ini mencakup:

a. Pembersihan Data (*Data Cleaning*)

Menghapus atau memperbaiki data yang tidak lengkap atau tidak valid, seperti data yang hilang pada variabel berat badan, tinggi badan atau tekanan darah.

b. Transformasi Data

Melakukan transformasi data jika diperlukan, misalnya normalisasi atau standarisasi apabila skala data tidak seragam.

c. Pengkodean Variabel Kategorikal

Mengubah data variabel menjadi bentuk numerik agar bisa diproses oleh algoritma seperti berikut :

1) Jenis kelamin ( Laki-laki = 1, Perempuan = 2)

- 2) Indeks Masa Tubuh (Underweight = 1, Normal = 2, Overweight = 3, Obesitas I = 4 dan Obesitas II = 5)
- 3) Status hipertensi (Normal = 1, Prehipertensi = 2, Hipertensi tingkat 1 = 3 dan Hipertensi tingkat 2 = 4)

## 2. Pembagian Data (*Data Splitting*)

Setelah data siap, langkah selanjutnya adalah membagi dataset menjadi dua bagian utama:

### a. Data Latih (*Training Data*)

Digunakan untuk melatih model *Decision Tree* sejumlah 1251 kasus hipertensi

### b. Data Uji (*Testing Data*)

Digunakan untuk menguji performa model yang telah dibangun guna menilai keakuratan dan efektivitas klasifikasi dengan jumlah 835 kasus hipertensi.

Pembagian data dilakukan dengan rasio 60:40 di mana persentase terbesar digunakan untuk pelatihan model.

## 3. Penerapan Algoritma *Decision Tree*

Setelah proses pra-pemrosesan dan pembagian data selesai, algoritma *Decision Tree* digunakan untuk mengklasifikasikan pasien berdasarkan status hipertensi yaitu: Normal, Pra-Hipertensi, Hipertensi Tingkat 1 dan Hipertensi Tingkat 2. Hal ini guna memanfaatkan variabel-variabel seperti usia, jenis kelamin, tekanan darah sistolik dan diastolik serta IMT untuk melakukan prediksi klasifikasi.

## 4. Evaluasi Model

Tahap akhir dari proses ini adalah mengevaluasi kinerja model yang telah dibangun. Evaluasi dilakukan menggunakan beberapa metrik sebagai berikut:

a. *Confusion Matrix* : Menampilkan jumlah klasifikasi yang benar dan salah untuk masing-masing kategori.

### b. Akurasi (*Accuracy*)

Persentase klasifikasi yang tepat dari seluruh data.

- c. Presisi (*Precision*) : Seberapa akurat model dalam mengklasifikasikan kategori tertentu.
- d. Recall (*Sensitivity*) : Kemampuan model dalam mengenali semua data yang benar-benar termasuk dalam suatu kategori.
- e. F1-Score : Nilai gabungan dari precision dan recall, memberikan gambaran menyeluruh terhadap performa model.

## H. Etika Penelitian

### 1. *Ethical Clearance* (Kelayakan Etika)

Penelitian ini telah lulus layak etik No.Skep/240/KEP/VI/2025 yang menyatakan bahwa penelitian dapat dilaksanakan setelah memenuhi persyaratan yang ditentukan

### 2. Sukarela

Peneliti telah memperoleh izin resmi dari instansi terkait untuk menggunakan data rekam medis sebagai data sekunder tanpa informed consent, karena tidak ada interaksi langsung dengan subjek. Data yang digunakan bersifat otoritatif dan hanya dimanfaatkan untuk kepentingan penelitian.

### 3. Anonimitas (Tanpa Nama)

Seluruh data yang digunakan tidak mencantumkan identitas pribadi pasien. Data disajikan dalam bentuk anonim menggunakan kode atau simbol untuk menggantikan nama asli guna melindungi privasi pasien.

### 4. Kerahasiaan

Peneliti bertanggung jawab menjaga kerahasiaan seluruh data yang diperoleh. Informasi hanya digunakan untuk kepentingan penelitian dan disampaikan dalam laporan tanpa mengungkap identitas individu secara langsung.

### 5. *Beneficence* (Kebajikan)

Penelitian ini tidak memberikan manfaat secara langsung bagi responden karena data yang digunakan diperoleh melalui rekam medis pasien tanpa keterlibatan langsung subjek dalam proses pengambilan data.

6. *Maleficience* (Tidak merugikan)

Data yang digunakan telah dianonimkan untuk menjaga kerahasiaan identitas pasien dan mencegah dampak negatif terhadap individu yang datanya digunakan.

7. *Risk of harm* (Risiko kerugian)

Peneliti menyadari adanya potensi risiko kesalahan klasifikasi, sehingga evaluasi model dilakukan secara hati-hati untuk mengurangi dampak yang merugikan.

8. *Justice* (Keadilan)

Model klasifikasi dikembangkan secara adil tanpa memihak kelompok tertentu, agar seluruh pasien memperoleh manfaat yang setara dari hasil penelitian.

## I. Rencana Pelaksanaan Karya Ilmiah

### 1. Persiapan

Pada tahap awal peneliti mulai dengan mengidentifikasi masalah di lapangan, khususnya terkait pemanfaatan data rekam medis pasien hipertensi yang belum optimal di Puskesmas Godean 1. Setelah itu peneliti menyusun proposal yang berisi latar belakang, rumusan masalah, tujuan, manfaat dan metode penelitian. Setelah proposal disusun dan disetujui oleh dosen pembimbing, peneliti mengurus surat izin penelitian ke Puskesmas Godean 1 dan juga surat *Ethical Clearance* (EC) melalui tim etik dari kampus sebagai persyaratan untuk pengambilan data rekam medis pasien. EC ini bertujuan memastikan bahwa penelitian dilakukan sesuai dengan prinsip etika penelitian, terutama yang berkaitan dengan data kesehatan pasien.

### 2. Pelaksanaan

Setelah izin dari EC didapatkan, peneliti mulai melaksanakan penelitian yang mencakup beberapa langkah berikut :

a. Pengambilan Data

- 1) Melakukan identifikasi pasien hipertensi berdasarkan rekam medis yang ada di Puskesmas Godean 1

- 2) Melakukan ekstraksi data rekam medis termasuk variabel yang relevan untuk penelitian
- 3) Menyelesaikan proses pembersihan data dan pengkodean data yang akan digunakan untuk analisis lebih lanjut.

b. Pra-pemrosesan Data

Pra-pemrosesan data merupakan tahapan penting untuk memastikan kualitas data sebelum dilakukan analisis. Langkah-langkah pra-pemrosesan dalam penelitian ini meliputi:

1) *Data Cleaning* (Pembersihan Data)

Proses ini dilakukan untuk menghapus data duplikat, memperbaiki kesalahan entri data dan menghilangkan data yang tidak konsisten atau tidak relevan, seperti data pasien yang tidak memiliki nilai tekanan darah lengkap atau nilai IMT yang kosong.

2) *Handling Missing Values* (Penanganan Data Kosong)

Data yang tidak lengkap atau hilang (*missing value*), seperti data tekanan darah yang tidak terisi, dikeluarkan dari dataset agar tidak memengaruhi hasil klasifikasi.

3) *Feature Selection* (Seleksi Atribut)

Peneliti memilih atribut-atribut penting yang berpengaruh terhadap klasifikasi hipertensi. Atribut yang digunakan dalam penelitian ini antara lain:

- a) Tekanan Darah Systolik
- b) Tekanan Darah Diastolik
- c) Usia
- d) Jenis Kelamin
- e) IMT

4) *Labeling*

Label klasifikasi hipertensi ditentukan berdasarkan nilai tekanan darah sesuai pedoman JNC 7, yang mengelompokkan pasien ke dalam kategori: Normal, Prehipertensi, Hipertensi Tingkat 1 dan Hipertensi Tingkat 2.

c. Pembuatan Model Klasifikasi

Data dibagi menjadi 60% data latih dan 40% data uji, kemudian dianalisis menggunakan algoritma *Decision Tree* (C4.5) dengan bantuan aplikasi *RapidMiner 10.3*.

d. Evaluasi Model

Model yang dihasilkan diuji menggunakan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*. Hasil yang diperoleh menunjukkan akurasi sebesar 98,56% yang menandakan model bekerja dengan sangat baik.

### 3. Penyusunan Laporan

Setelah seluruh proses pengolahan dan analisis data selesai dilakukan, peneliti mulai menyusun laporan karya tulis ilmiah. Laporan ini disusun secara bertahap dan sesuai dengan format penulisan yang berlaku di kampus. Adapun langkah-langkahnya sebagai berikut:

a) Penyusunan Struktur Laporan

- 2) Bab 1 Pendahuluan (Latar belakang, Rumusan Masalah, Tujuan Penelitian dan Manfaat Penelitian)
- 3) Bab 2 Tinjauan Pustaka (Teori tentang hipertensi, puskesmas, rekam medis, *data mining*, algoritma *Decision Tree* dan evaluasi model *Decision Tree*)
- 4) Bab 3 Metode Penelitian (Desain penelitian, Populasi, Sampel, Teknik pengumpulan data dan Pengolahan data)
- 5) Bab 4 Hasil dan Pembahasan (Hasil analisis Evaluasi model dan Interpretasi)
- 6) Bab 5 Kesimpulan dan Saran (Kesimpulan penelitian dan Rekomendasi untuk penelitian selanjutnya)

b) Penulisan Laporan

- 1) Menyusun setiap bab laporan secara terstruktur berdasarkan hasil analisis yang telah dilakukan
- 2) Mencantumkan gambar atau tabel sebagai pendukung hasil visualisasi dari analisis yang dilakukan

- c) Revisi dan Finalisasi Laporan
  - 1) Melakukan revisi setelah mendapatkan umpan balik dari pembimbing dan pihak terkait lainnya
  - 2) Finalisasi laporan dengan memastikan bahwa semua bagian laporan sudah lengkap, jelas dan memenuhi standar akademik
- d) Presentasi Hasil Penelitian
  - 1) Menyiapkan presentasi hasil penelitian untuk disampaikan di depan dosen pembimbing atau penguji
  - 2) Memastikan bahwa presentasi mencakup poin-poin penting seperti latar belakang, tujuan, metode, hasil analisis dan kesimpulan.