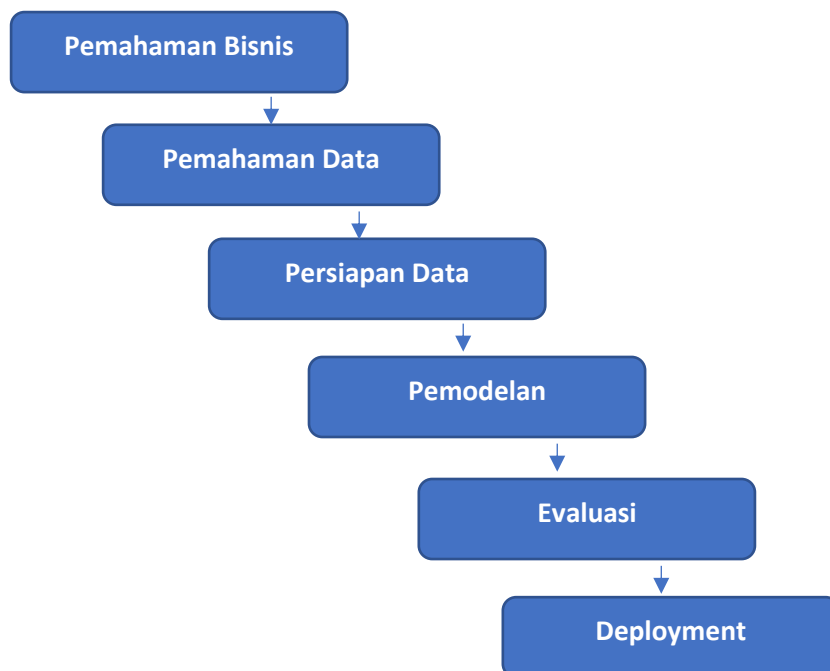


## BAB III METODOLOGI PENELITIAN

Bab ini membahas langkah-langkah dari proses penelitian yang akan dilaksanakan. Metode yang digunakan dalam penelitian ini mengikuti tahapan *model Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Tahapan penelitian dapat dilihat pada Gambar 3.1 di bawah ini:



Gambar 3.1 Alur Penelitian

Berikut ini merupakan penjelasan dari gambar 3.1 yang merupakan alur dari penelitian ini

1. Pemahaman bisnis (Business understanding)

Pada tahap pertama berfokus pada kebutuhan berdasarkan penilaian bisnis, tujuan bisnis berdasarkan pengolahan data untuk mengambil keputusan mengenai diagnosis penyakit jantung. Saat ini metode yang digunakan oleh pihak rumah sakit untuk memberi keputusan klinis dibuat berdasarkan ilmu dan pengetahuan dari pasien-pasien penyakit jantung.

## 2. Pemahaman data (Data understanding)

Data penelitian terdiri dua jenis yaitu data primer dan data sekunder. Data primer merupakan data yang diperoleh atau dikumpulkan oleh peneliti secara langsung dari sumber datanya. Data sekunder merupakan data yang diperoleh atau dikumpulkan peneliti dari berbagai sumber yang telah ada. Data sekunder dalam penelitian ini adalah *dataset public* yang di peroleh dari <https://www.kaggle.com/johnsmith88/heart-disease-dataset>. Kumpulan data ini berasal terdiri dari empat basis data: Cleveland, Hungaria, Swiss, dan Long Beach V. Ini berisi 76 atribut, termasuk atribut yang diprediksi, tetapi semua eksperimen yang dipublikasikan mengacu pada penggunaan subset dari 14 atribut tersebut. Atribut-atribut yang digunakan antara lain usia, jenis kelamin, jenis nyeri dada, tekanan darah istirahat, kolestrol, gula darah, hasil elektrokardiografi istirahat, detak jantung, angina, depresi ST yang diinduksi oleh olahraga relatif terhadap istirahat, kemiringan segmen ST latihan puncak, jumlah pembuluh utama, Bidang "target" mengacu pada adanya penyakit jantung pada pasien. Ini adalah bilangan bulat bernilai 0 = tidak ada penyakit dan 1 = penyakit.

## 3. Persiapan data (Data preparation)

Pengolahan dataset penyakit jantung yaitu menentukan atribut apa yang akan digunakan untuk proses perhitungan yang akan digunakan untuk menemukan hasil akurasi terbaik. Atribut data penelitian dapat dilihat pada table 3.1 dibawah ini.

Table 3.1 Atribut Dataset

No.	Atribut	Skala Pengukuran
1.	<i>Age</i> (Usia)	[29,0;77,0]
2.	<i>Sex</i> (Jenis Kelamin)	[0,0;1,0] 0,0 = wanita 1,0 =pria
3.	<i>Chest Pain real</i> (Nyeri dada)	[0,1,2,3] 0 = Tipikal angina 1 = Angina atipikal 2 = Nyeri non angina 3 = Tanpa gejala

Table 3.1 Atribut Dataset

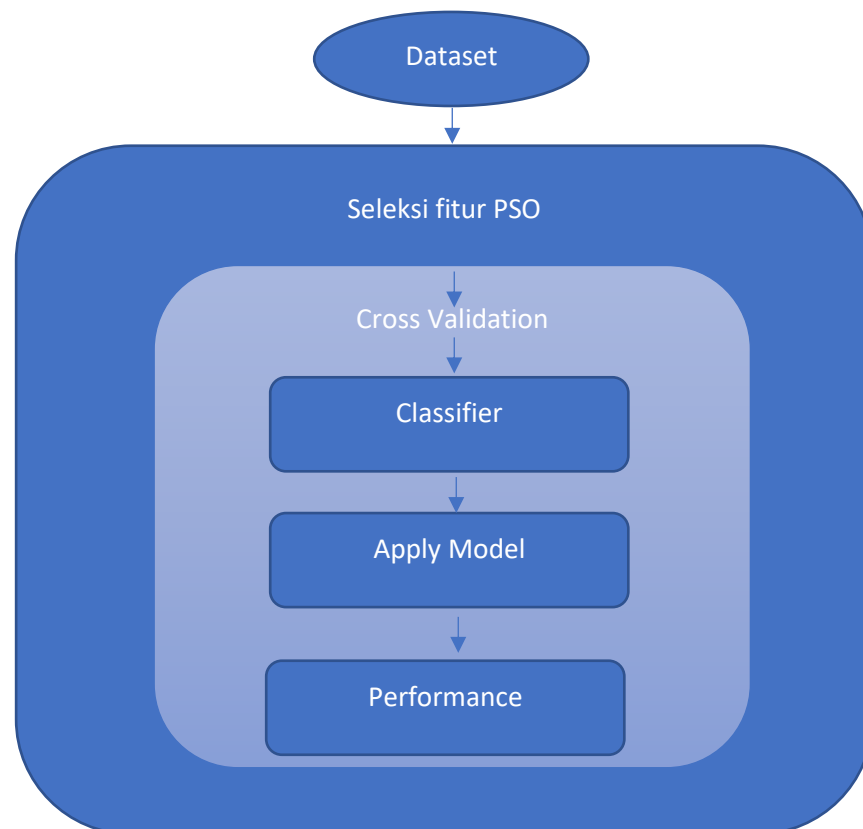
No.	Atribut	Skala Pengukuran
4.	<i>Trestbps</i> ( <i>Resting blood pressure</i> /Tekanan Darah)	[94,0;200,0]
5.	<i>Chol</i> (serum kolestoral)	[126,0;564,0]
6.	<i>Fbs</i> ( <i>fasting blood sugar</i> )/Gula Darah	[0,1] 0 = <i>false</i> 1 = <i>true</i>
7.	<i>Restecg</i> (hasil electrocardiografi)	[0,1,2] - Nilai 0 = normal. - Nilai 1 = memiliki kelainan gelombang ST-T (inversi gelombang T dan / atau elevasi atau depresi ST > 0,05 mV). - Nilai 2 = menunjukkan hipertrofi ventrikel kiri yang mungkin atau pasti berdasarkan kriteria Estes.
8.	<i>Thalach</i> (denyut jantung maksimum tercapai)	[71,0;202,0]
9.	<i>Exang</i> (latihan yang diinduksi angina)	[0;1] 0 = <i>No</i> 1 = <i>Yes</i>
10.	<i>Oldpeak</i> (Depresi ST disebabkan oleh olahraga relatif terhadap istirahat)	[0,0;6,2]
11.	<i>Slope</i> (kemiringan segmen latihan puncak ST)	[1,2,3] - Nilai 1 = Menanjak - Nilai 2 = Datar - Nilai 3 = Downsloping
12.	<i>Ca real</i> ( <i>number of major vessels (0-3) colored by flourosopy</i> )	[0,1,2,3]
13.	<i>Thal real</i>	[0,1,2] 0 = <i>Normal</i> ; 1 = <i>Fixed defect</i> ; 2 = <i>Reversable defect</i>

Table 3.1 Atribut Dataset

No.	Atribut	Skala Pengukuran
14.	Target (diagnosis penyakit jantung)	{0,1} -Value 0: No Disease -Value 1: Disease

#### 4. Pemodelan (modelling)

Pada tahapan pemodelan ini, terdapat dua metode yang akan digunakan yaitu algoritma KNN yang selanjutnya dikombinasikan dengan algoritma *Particle Swarm Optimization*. Kombinasi dilakukan dengan mencari bobot dari setiap atribut. Untuk membandingkan dan pengujiannya, akan menggunakan tools *RapidMiner* sehingga akan terlihat peningkatan akurasi. Model penelitian dapat dilihat pada gambar 3.2 dibawah ini.



Gambar 3.2 Pemodelan penelitian

#### 5. Validasi

Dalam tahapan validasi dan evaluasi, akan menggunakan *tools RapidMiner*. Dengan tools tersebut akan dilakukan pengujian keakurasian dengan beberapa teknik yaitu confusion matrix dan kurva ROC.

#### 6. Deployment

Setelah tahap evaluasi dimana menilai secara detail hasil dari sebuah model maka dilakukan pengimplementasian dari keseluruhan model yang telah dibangun. Selain itu juga dilakukan penyesuaian terhadap model sehingga dapat menghasilkan suatu hasil yang sesuai dengan target awal tahap CRISP-DM ini.

### 3.1 Studi Literatur

Studi literatur adalah serangkaian kegiatan yang berkenaan dengan metode pengumpulan data pustaka, membaca dan mencatat, serta mengelolah bahan penelitian. Studi Literatur adalah merupakan penelitian yang dilakukan oleh peneliti dengan mengumpulkan sejumlah buku buku, majalah yang berkaitan dengan masalah dan tujuan penelitian (*Alber et al. 2012*).

Teknik ini dilakukan dengan tujuan untuk mengungkapkan berbagai teori-teori yang relevan dengan permasalahan yang sedang dihadapi/diteliti sebagai bahan rujukan dalam pembahasan hasil penelitian. Pengertian Lain tentang Studi literatur adalah mencari referensi teori yang relefan dengan kasus atau permasalahan yang ditemukan.

Secara Umum Studi Literatur adalah cara untuk menyelesaikan persoalan dengan menelusuri sumber-sumber tulisan yang pernah dibuat sebelumnya. Dengan kata lain, istilah Studi Literatur ini juga sangat familier dengan sebutan studi pustaka. Dalam sebuah penelitian yang hendak dijalankan, tentu saja seorang peneliti harus memiliki wawasan yang luas terkait objek yang akan diteliti. Jika tidak, maka dapat dipastikan dalam persentasi yang besar bahwa penelitian tersebut akan gagal.

### **3.2 Alat dan bahan penelitian**

Dalam penelitian ini, untuk melakukan penambangan data, dibutuhkan perangkat keras dan lunak untuk menunjang kelancaran dalam proses penambangan.

Perangkat tersebut diantaranya:

1. Perangkat keras yang digunakan untuk melakukan penambangan data adalah sebagai berikut:
  - Processor core i3
  - RAM 4 Gb
  - SSD 256
2. Perangkat lunak yang digunakan untuk melakukan penambangan data adalah sebagai berikut:
  - Sistem operasi windows 10
  - Software RapidMiner