

**KLASIFIKASI PENYAKIT SERANGAN  
JANTUNG MENGGUNAKAN  
METODE NAÏVE BAYES**

**PROPOSAL TUGAS AKHIR**



Diajukan oleh :

Billi

Elgiaridi

8020190100

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir  
Sebagai Akhir Proses Studi Strata 1

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS DINAMIKA BANGSA  
2022**

## **IDENTITAS PROPOSAL PENELITIAN**

Judul proposal : KLASIFIKASI PENYAKIT SERANGAN JANTUNG  
MENGUNAKAN METODE NAÏVE BAYES

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

a. Nama Lengkap : Billi Elgiaridi

b. NIM : 8020190100

c. Jenis Kelamin : Laki-Laki

d. Tempat/Tgl. Lahir : Sungai Pulai, 21 Juli 2001

e. Alamat : Kec. Mersam, Kab. Batanghari,  
Prov Jambi

f. No Telepon : 082278121066

g. Gmail : billielgi019@gmail.com

## PERNYATAAN HASIL EVALUASI

Nama / Nim : Billi Elgiaridi / 8020190100

PRODI : ~~SI/TI/SK~~

JUDUL : KLASIFIKASI PENYAKIT SERANGAN JANTUNG  
MENGUNAKAN METODE NAÏVE BAYES

Hasil Evaluasi : Disetujui / Disetujui dengan perbaikan / Ditolak \*)

1. Catatan
2. Alasan penolakan tugas akhir
  - Tugas Akhir tidak relevan dengan program studi
  - Pernah ada topik sejenis
  - Metode yang dipakai tidak jelas
  - Masalah terlalu sempit
  - .....

Mengetahui,  
Ketua Program Studi TI

**BENY, S.KOM, M.SC**  
**YDB.07.84.055**

\*) Coret yang tidak perlu

## 1. LATAR BELAKANG

Serangan jantung merupakan salah satu penyakit paling mematikan di dunia.[1] Penyakit ini adalah salah satu penyakit yang paling sering terjadi kasusnya pada kalangan masyarakat, dimana penyakit jantung ini dapat menimpa dan menyerang siapapun tanpa memandang usia, jenis kelamin dan gaya hidup.[2] Serangan jantung dapat terjadi di mana saja dan kapan saja. Seringkali masyarakat terlambat dalam melakukan penanganan awal, mencari informasi dan pertolongan. Karena pada dasarnya serangan jantung dapat menyerang secara tiba-tiba dan dapat mengancam nyawa.[3] Berdasarkan data kejadian penyakit di Indonesia dari GLOBOCAN (IARC) tahun 2012, terdapat 43,30 kasus baru infark miokard dan 12,90% kematian. Jumlah penderita penyakit jantung pada tahun 2013 sebanyak 61.682 orang, dengan kejadian tertinggi per provinsi yaitu Jawa Tengah sebesar 11.511 orang. Beberapa faktor yang dapat menyebabkan penyakit jantung umumnya adalah kebiasaan hidup tidak sehat, merokok, begadang dan pola makan yang tidak baik.[4]

Dari data diatas dapat diketahui bahwa banyak orang yang belum menanggapi penyebab penyakit ini dengan serius, banyak alternatif cara untuk mencegah bahkan menyembuhkan penyakit-penyakit tersebut, namun kurangnya akses informasi/media menjadi alasan penderita terlambat untuk memeriksakan diri ke dokter.[1]

Terdapat hubungan antara kurangnya akses informasi/media dengan keterlambatan pemeriksaan awal Penyakit Jantung.[1] Kurangnya akses informasi tentang serangan jantung meningkatkan angka kematian setiap tahun. Oleh karena itu, diperlukan suatu sistem klasifikasi yang dapat memberikan informasi tentang penyakit serangan jantung dan mengkonfirmasi klasifikasi penyakit serangan jantung seseorang secara dini. Untuk menjalankan sistem klasifikasi membutuhkan pengelolaan pengetahuan yang tepat yang diwarisi dari para ahli untuk mencapai hasil yang akurat. Salah satu metode yang dapat digunakan dalam pekerjaan ini adalah penerapan *classifier naive Bayes*, dimana metode *naive bayes* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naive bayes* memiliki akurasi dan kecepatan yang sangat

tinggi saat diaplikasi ke dalam database dengan data yang besar.[5] Data yang digunakan pada penelitian ini di ambil secara online dari website [www.kaggle.com](http://www.kaggle.com) berupa data campuran, yaitu data *numeric* dan kategoris. Maka penelitian ini memanfaatkan metode naïve bayes sebagai pengoptimasian.

Beberapa penelitian yang menerapkan *Naïve Bayes Classifier* sudah banyak diteliti. Seperti pada penelitian yang diimplementasikan oleh Muhamad, yang mengemukakan *Naïve Bayes* mempunyai banyak keunggulan, diantaranya memiliki akurasi yang tinggi dengan bentuk algoritma yang sederhana namun cepat menyelesaikan masalah, namun validasi pengujian yang dilakukan dalam penelitian tersebut hanya satu kali. Mekanisme pengujian dilakukan dengan menggunakan 123 dataset dengan parameter akurasi saja, peneliti akan melakukan pengujian dengan mekanisme membagi 303 data kedalam 5 subset yang akan divalidasi dengan *5-fold cross validation* dengan parameter pengujian akurasi, presisi dan recall. Penelitian selanjutnya adalah percobaan yang diimplementasikan oleh Li dkk yang menjelaskan bahwa algoritma *naïve bayes* dapat digunakan untuk mengelola dataset dengan jumlah besar, algoritma *naïve bayes* mampu menghasilkan akurasi lebih dari 78%, hasil tersebut didapatkan melalui seleksi fitur terlebih dahulu sebelum melakukan klasifikasi sehingga dapat meningkatkan performa dari algoritma naïve bayes. Sedangkan pada penelitian yang dilakukan nuraini lebih baik diimplementasikan pada data dengan jumlah yang besar serta dapat melakukan penganganan data yang bersifat kurang lengkap dan lumayan baik pada suatu atribut yang memiliki noise dan tidak relevannya pada sebuah data.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya maka pada penelitian ini akan dilakukan penelitian pembuatan sistem klasifikasi penyakit jantung menggunakan *Naïve Bayes Classifier*. Dengan tujuan memberikan kontribusi penelitian melalui konfigurasi penerapan model algoritma dengan mekanisme validasi pengujian yang berbeda dari penelitian sebelumnya untuk hasil tingkat akurasi, presisi, dan recall yang lebih akurat, dengan mengangkat judul **“KLASIFIKASI PENYAKIT SERANGAN JANTUNG MENGGUNAKAN METODE NAÏVE BAYES”**

## **2. RUMUSAN MASALAH**

Berdasarkan uraian diatas, maka dapat dirumuskan masalah yang akan dikaji dalam penelitian ini adalah Bagaimana cara kerja sistem yang berjalan dengan memanfaatkan algoritma menggunakan metode naïve bayes dalam mendiagnosa penyakit jantung?

### **1. BATASAN MASALAH**

Beberapa masalah yang dijadikan fokus dalam penelitian ini antara lain:

1. Penelitian hanya terfokus pada penyakit jantung yang umum dimasyarakat.
2. Input berupa nilai kriteria dari hasil pemeriksaan pasien penyakit jantung.
3. Output berupa hasil, apakah pasien menderita penyakit jantung atau tidak.
4. Kriteria yang digunakan adalah usia, jenis kelamin, tipe sakit dada, tekanan darah, kolesterol, gula darah, hasil elektrokardiogram, denyut per menit, induksi angina, depresi, kemiringan segmen, pembuluh darah, dan terdiagnosa

### **2. TUJUAN DAN MANFAAT PENELITIAN**

#### **1.4.1 Tujuan penelitian**

Adapun tujuan dari penelitian ini adalah untuk:

1. Mendeskripsikan usia, jenis kelamin, tipe nyeri dada, tekanan darah, kolestrol, gula darah, hasil elektrokardiogram, denyut per menit, induksi angina, depresi, kemiringan segmen, dan terdiagnosa.
2. Menganalisis hubungan usia, jenis kelamin, nyeri dada, tekanan darah, kolestrol, gula darah, hasil elektrokardiogram, denyut per menit, angina, depresi, kemiringan segmen, dan terdiagnosa dengan penyakit serangan jantung

#### **1.4.2 Manfaat penelitian**

Manfaat yang diperoleh dari penelitian ini adalah:

1. Membantu masyarakat untuk dapat mengetahui gejala dini penyakit jantung.

2. Membantu dokter maupun praktisi kesehatan yang dapat membaca hasil pemeriksaan laboratorium untuk mendiagnosa penyakit jantung dengan layak.

### **3. LANDASAN TEORI**

#### **5.1 Serangan Jantung**

Serangan jantung merupakan manifestasi sindrom koroner akut (acute coronary syndrome), disebut juga infark miokardial akut, karena terjadi kematian otot jantung yang disebabkan oleh berhentinya aliran darah karena penyumbatan secara tiba-tiba arteri koroner oleh trombus[6]. Arteri koroner adalah pembuluh darah yang memasok darah dan oksigen ke sel-sel otot jantung. Penyumbatan pada arteri koroner mengurangi pasokan darah dan oksigen menuju otot jantung (iskemia) mengakibatkan injuri pada otot jantung dan menghasilkan manifestasi klinis nyeri dada (angina pectoris).

Jika aliran darah ke jantung tidak pulih dalam waktu 20 - 40 menit, bisa mengakibatkan kematian otot jantung yang tidak dapat pulih (irreversible). Kematian otot jantung secara komplit terjadi antara 6-8 jam. Sel otot jantung yang mati akan digantikan oleh jaringan parut. Jadi serangan jantung merupakan manifestasi klinis akut penyakit kardiovaskuler yang dapat ditandai dengan angina pektoris (nyeri dada), infark miokardial akut atau bahkan kematian mendadak

## **5.2 DATA MINING**

Data Mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan Data Mining adalah kenyataan bahwa Data Mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, Data Mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani[7]:

1. Jumlah data yang sangat besar
2. Dimensi data yang tinggi
3. Data yang heterogen dan berbeda sifat

Dalam perkembangannya data mining memiliki banyak definisi yang cukup beragam sehingga data mining dapat menambah ilmu pengetahuan. Berikut ini adalah beberapa definisi data mining pada umumnya[8]:

Menurut Turban, dkk (dalam Kusri, 2009:3) Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Menurut Turban, dkk (dalam Kusri, 2009:4) Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar.

Dari beberapa pendapat diatas, maka dapat disimpulkan bahwa data mining merupakan sebuah teknik dari beberapa bidang ilmu untuk menemukan hubungan yang tidak diketahui sebelumnya di dalam gudang data sehingga menjadi suatu informasi yang dapat digunakan kemudian.

### **5.2.1 Pengolahan Data Mining**

Pengolahan Data Mining Pengolahan data mining terdiri dari beberapa metode pengolahan, yaitu [9]:

- 1). Predictive modelling yang merupakan pengolahan data mining dengan melakukan prediksi/ peramalan. Tujuan metode ini untuk membangun model

prediksi suatu nilai yang mempunyai ciri-ciri tertentu. Contoh algoritmanya Linear Regression, Neural Network, Support Vector Machine, dan lainlain.

2). Association (Asosiasi) merupakan teknik dalam data mining yang mempelajari hubungan antar data. Contoh penggunaannya seperti untuk menganalisis perilaku mahasiswa yang datang terlambat. Contohnya jika mahasiswa memiliki jadwal dengan dosen A dan B, maka mahasiswa akan datang terlambat. Contoh algoritmanya FP-Growth, A Priori, dan lain-lain.

3). Clustering (Klastering) atau pengelompokan merupakan teknik untuk mengelompokkan data ke dalam suatu kelompok tertentu. Contoh algoritmanya K-Means, K-Medoids, Self-Organisation Map (SOM), Fuzzy C-Means, dan lain-lain. Contoh untuk clustering: Terdapat lima pulau di Indonesia: Sumatera, Kalimantan, Jawa, Sulawesi dan Papua. Maka lima pulau tersebut dijadikan tiga klaster berdasarkan waktunya: Waktu Indonesia Barat (Sumatera, Kalimantan dan Jawa), Waktu Indonesia Tengah (Sulawesi) dan Waktu Indonesia Timur (Papua).

4). Classification merupakan teknik mengklasifikasikan data. Perbedaannya dengan metode clustering terletak pada data, dimana pada clustering variabel dependen tidak ada, sedangkan pada classification diharuskan ada variabel dependen. Contoh algoritma yang menggunakan metode ini ID3 dan K Nearest Neighbors.

5). Pohon Keputusan (Decision Tree) Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, pohon keputusan sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

### 5.3 KLASIFIKASI

Klasifikasi adalah salah satu pembelajaran yang paling umum di data mining. Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas[10].

Klasifikasi adalah penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau standar yang ditetapkan. Beberapa algoritma yang dapat digunakan dalam klasifikasi *data mining* adalah sebagai berikut[11] :

#### A. *Neural Network*

*Neural Network* (Jaringan Saraf Tiruan) adalah prosesor tersebar paralel yang sangat besar dan memiliki kecenderungan untuk menyimpan pengetahuan yang bersifat pengalaman dan membuatnya siap untuk digunakan.

#### B. *Decision Tree*

*Decision tree* sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak di minati. Dalam *decision tree* ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami.

#### C. *Naïve Bayes*

*Naïve Bayes* merupakan sebuah model klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. *Naïve Bayes* didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*.

#### D. *K-Nearest Neighbor*

Algoritma *k-nearest neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi

Jadi dapat disimpulkan bahwa klasifikasi merupakan kegiatan yang digunakan untuk mengelompokkan data kedalam kelas-kelas tertentu dalam pengerjaan klasifikasi dapat menggunakan beberapa metode algoritma seperti *Neural Network*, *Decision Tree*, *Naive Bayes*, *K-Nearest Neighbor*

## 5.4 ALGORITMA NAÏVE BAYES

*Naive Bayes classifier* (NBC) merupakan salah satu metoda pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

*Naive Bayes* merupakan perhitungan *teorema bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas. Selain itu, algoritma *Naive Bayes* juga mampu menangani set data yang memiliki banyak atribut. Persamaan dari *Naive Bayes* sebagai berikut[10]:

$$P(C_i | X) = \frac{P(X | C_i) \cdot P(C_i)}{P(X)} \quad (2.1)$$

Penjelasan dari formula tersebut adalah sebagai berikut :

$X$  = Kriteria suatu kasus berdasarkan masukan

$C_i$  = Kelas solusi pola ke- $i$ , dimana  $i$  adalah jumlah label kelas.

$P(C_i | X)$  = Probabilitas kemunculan label kelas  $C_i$  dengan kriteria masukkan  $X$ .

$P(X | C_i)$  = Probabilitas kriteria masukan  $X$  dengan label kelas  $C_i$ .

$P(X)$  = Probabilitas labek kelas  $C_i$ .

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C) \cdot P(F_1 \dots F_n)}{P(F_1 \dots F_n)}$$

Dimana Variabel  $C$  merepresentasikan kelas, sementara variable  $F_1 \dots F_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas  $C$  (Posterior) adalah peluang munculnya kelas  $C$  (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$\text{Posterior} = \frac{\text{pior} \times \text{likelihood}}{\text{evidence}}$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $(C|F1, \dots, Fn)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(C|F1, \dots, Fn) &= P(C)P(F1, \dots, Fn|C) \\ &= P(C)P(F1|C)P(F2, \dots, Fn|C, F1) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3, \dots, Fn|C, F1, F2) \\ &= (C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2)P(F4, \dots, Fn|C, F1, F2, F3) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2) \dots P(Fn|C, F1, F2, F3, \dots, Fn-1) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya factor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk  $(F1, F2 \dots Fn)$  saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(Fi|Fj) = \frac{P(Fi \cap Fj)}{P(Fj)} = \frac{.P(Fi)P(Fj)}{P(Fj)} = P(Fi)$$

Untuk  $i \neq j$ , sehingga :

$$P(Fn|C, Fn) = P(Fn|C)$$

Persamaan di atas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss :

$$P(Xi = xi)|(Y = yj) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Di mana:

P : Peluang

$X_i$  : Atribut ke  $i$

$x_i$  : Nilai atribut ke  $i$

Y : Kelas yang dicari

$y_i$  : Sub kelas Y yang dicari

$\mu$  : Mean, menyatakan rata-rata dari seluruh atribut

$\sigma$  : Deviasi standar, menyatakan varian dari seluruh atribut

Berdasarkan pengertian diatas dapat disimpulkan bahwa algoritma *naïve bayes* merupakan salah satu metode algoritma yang memiliki independensi yang kuat (naif) yang digunakan untuk memprediksi kemungkinan dimasa yang akan datang melalui pengalaman dari masa lampau.

## 5.5 WEKA

WEKA merupakan sebuah perangkat lunak yang menerapkan berbagai algoritma *machine learning* untuk melakukan beberapa proses yang berkaitan dengan sistem temu kembali informasi atau *data mining*.

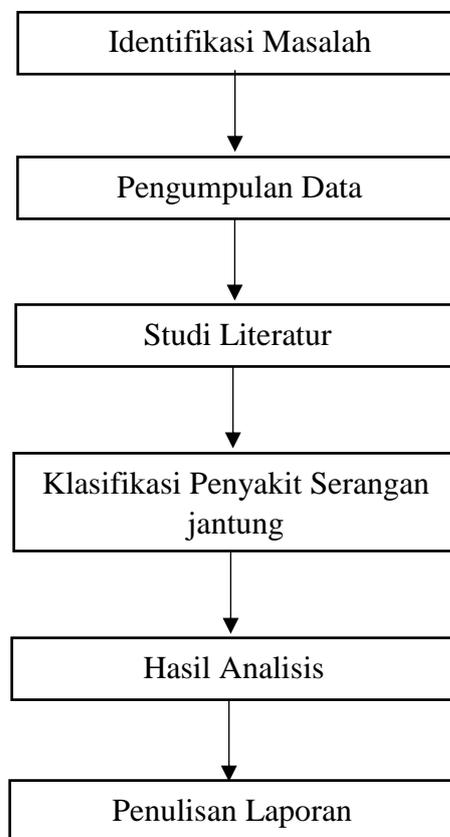
*The Waikato Environment for Knowledge Analysis* (Weka) adalah rangkaian lengkap perpustakaan kelas *Java* yang mengimplementasikan banyak *state-of-the-art* pembelajaran mesin dan algoritma *data mining*. Weka tersedia secara bebas di *World Wide Web* dan menyertai teks baru pada dokumen *data mining* dan sepenuhnya menjelaskan semua algoritma yang dikandungnya. Aplikasi yang ditulis menggunakan *library class* pada Weka yang dapat dijalankan pada komputer manapun dengan kemampuan *browsing Web*, ini memungkinkan pengguna untuk menerapkan teknik pembelajaran mesin untuk data mereka sendiri terlepas dari *platform* komputer.

WEKA memiliki implementasi semua teknik pembelajaran untuk klasifikasi dan regresi, yaitu decision trees, rules set, pengklasifikasian teorema bayes, Support Vektor Mechines (SVM), logistik dan linier, multi layers perceptrons dan metode nearest neighbour[10].

## 6. METODE PENELITIAN

### 6.1 Kerangka Kerja Penelitian

Kerangka kerja (*frame work*) merupakan tahapan-tahapan yang dilakukan selama mengerjakan penelitian. Kerangka kerja penelitian dibuat agar mempermudah pencapaian hasil penelitian, dapat menyelesaikan penelitian tepat waktu dan penelitian dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar berikut :



**Gambar 6.1 Kerangka Kerja Penelitian**

Berdasarkan gambar kerangka kerja penelitian diatas, dapat diuraikan sebagai berikut:

#### 1. Identifikasi Masalah

Pada tahap ini penulis mengidentifikasi hal apa saja yang menjadi permasalahan data mining mengenai penyakit serangan jantung, sehingga penulis

dapat menentukan rencana kerja serta menentukan data apa saja yang akan dibutuhkan dalam penelitian ini.

## 2. Pengumpulan data

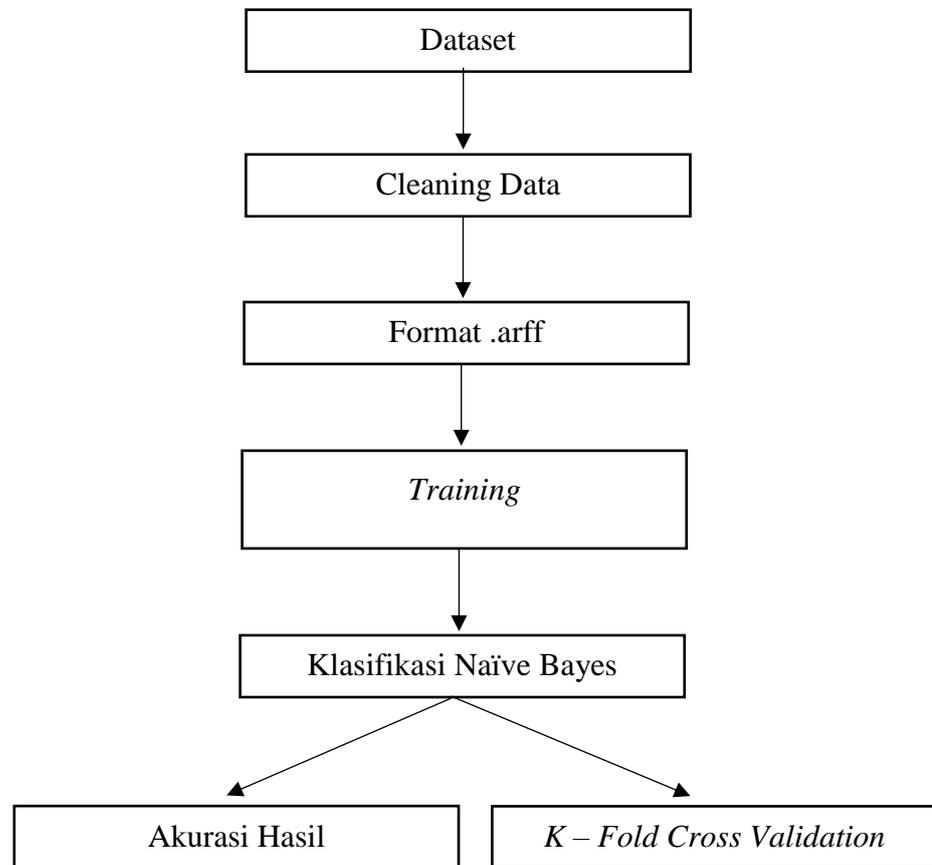
Pada tahapan ini penulis melakukan pengumpulan data dan informasi yang dibutuhkan dalam penelitian. Pengumpulan data ini dilakukan dengan beberapa metode yaitu:

## 3. Studi literatur

Pada tahapan ini penulis melakukan pencarian terhadap landasan-landasan teori yang diperoleh dari berbagai sumber seperti buku, jurnal ilmiah dan juga referensi lainya untuk melengkapi penelitian baik mengenai konsep dan teori sehingga memiliki acuan yang baik dan relevan.

## 4. Klasifikasi Penyakit Serangan Jantung

Pengklasifikasian data dilakukan dengan bererapa tahapan. Adapun alur proses pengklasifikasian dapat dilihat pada gambar 6.2



**Gambar 6.2 Klasifikasi Penyakit Serangan Jantung**

Berikut ini penjelasan mengenai tahapan-tahapan pada alur proses pengklasifikasian:

a. Dataset

Dataset ini diambil dari situs kaggle.com yang memiliki atribut seperti umur, jenis kelamin, nyeri dada, tekanan darah, kolestrol, gula darah, hasil elektrokardiogram, denyut per menit, angina, depresi, kemiringan segmen, dan terdiagnosa

b. Cleaning Data

Cleaning data merupakan preproses dimana pada tahap ini dilakukan pembersihan atau pemisahan data dari noise data dan data yang tidak konsisten seperti pasien yang memiliki atribut tidak lengkap.

Pengklasifikasikan umur dibagi menjadi 4 kategori yaitu dari 30-39 tahun, 40-49 tahun, 50-59 tahun, dan > 60 tahun.

c. Format arff

WEKA mendukung beberapa format file dalam input-nya diantaranya adalah format arff. Attribute-Relation File Format (ARFF) merupakan tipe file text yang berisi berbagai instance data yang berhubungan dengan suatu set atribut data. Kemudian data-data ini disusun dengan menggunakan tanda koma (,) sebagai pemisah data per-atributnya.

d. Training

Pada tahapan ini dilakukan proses training oleh algoritma klasifikasi naïve bayes untuk membentuk sebuah model classifier pada dataset yang sudah ada. Model ini merupakan representasi pengetahuan yang akan digunakan prediksi kelas data baru yang belum pernah ada.

e. Naïve Bayes Klasifikasi

Pada tahap ini, dilakukan proses pengklasifikasian data menggunakan tools WEKA terhadap data mahasiswa yang sudah ditesting menggunakan model classifier yang sudah dibentuk.

f. Akurasi hasil

Akurasi hasil menunjukkan kedekatan hasil pengukuran dengan nilai sesungguhnya.

g. *K-Fold Cross Validation*

Cross Validation merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Salah satu metode cross-validation yang populer adalah K-Fold Cross Validation.

5. Hasil Analisis

Hasil dari analisis akan dinilai keakuratan data model yang sudah di dapatkan dengan perhitungan naïve bayes menggunakan teknik *Cross Validation*. Salah satu metode *Cross-Validation* yang populer dan akan digunakan adalah *K-Fold Cross Validation*. Dalam teknik ini dataset dibagi menjadi sejumlah K-buah pertisi secara acak. Kemudian dilakukan

sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data pertisi ke-K sebagai data testing dan memanfaatkan sisa pertisi lainnya sebagai data training. Dimana hasil dengan persentasi terbesar yang akan di ambil. Selain itu, bentuk visualisasi dari setiap atribut dapat dilihat pada tools WEKA.

#### 6. Seleksi Atribut

Seleksi atribut ini dilakukan untuk mengetahui atribut apa saja yang mempengaruhi hasil klasifikasi. Seleksi atribut ini menggunakan algoritma yang umum digunakan yaitu algoritma *Classification Attribute Evaluation* (*ClassificationAttributeEval*) pada tools WEKA.

#### 7. Penulisan Laporan

Setelah semua tahapan penulisan dilakukan, maka akan dibuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan pada waktu yang akan datang, baik oleh peneliti sendiri maupun peneliti lainnya.

## 7. ALAT DAN BAHAN PENELITIAN

Alat yang dibutuhkan dalam penelitian dibagi menjadi dua yaitu hardware dan software sebagai berikut:

### 7.1 Alat penelitian

#### a. Perangkat keras (*Hardware*)

Perangkat keras yang penulis gunakan dalam penelitian ini adalah sebagai berikut:

- 1.) Satu unit laptop ACER Aspire E 14 dengan spesifikasi prosessor Intel Core i5-7200U
- 2.) Smartphone
- 3.) Printer
- 4.) Flashdisk
- 5.) Printer
- 6.) Kertas dan Alat tulis

*b.* Perangkat Lunak (*software*)

Perangkat lunak yang penulis gunakan dalam penelitian ini adalah sebagai berikut:

- 1) Sistem Operasi Windows 10 Pro 64-bit
- 2) Microsoft Office 2016
- 3) Microsoft Excel 2016
- 4) Google Chrome
- 5) Mandeley
- 6) Tools WEKA

## **7.2 Bahan penelitian**

1. Objek penelitin

Objek penelitian ini diambil secara online , yaitu menggunakan data online dari website [www.kaggle.com](http://www.kaggle.com).

2. Data penelitian

## 8. JADWAL PENELITIAN

Berikut jadwal waktu penelitian yang direncanakan berdasarkan kerangka kerja (Frame Work) yang telah disusun yaitu dilaksanakan pada bulan september 2022 sampai dengan Desember 2022. Penelitian dilakukan selama 4 bulan dengan perincian seperti pada tabel berikut ini :

No	Kegiatan	september				oktober				november				desember			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Identifikasi Masalah	■	■														
2	Pengumpulan Data			■													
3	Studi Literatur				■	■	■										
4	Klasifikasi Penyakit Serangan Jantung						■	■	■	■	■	■					
5	Penulisan Laporan										■	■	■	■	■	■	■

**Tabel 8.1 Jadwal penelitian**

## DAFTAR PUSTAKA

- [1] M. A. Bianto, K. Kusriani, and S. Sudarmawan, "Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 75, 2020, doi: 10.24076/citec.2019v6i1.231.
- [2] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [3] S. Rahayu, A. Subekhi, D. Astuti, I. Widaningsih, and I. Sartika, "Upaya mewaspadaı serangan jantung melalui pendidikan kesehatan," *JMM (Jurnal Masy. Mandiri)*, vol. 4, no. 2, pp. 163–171, 2020.
- [4] H. M. Nawawi, J. J. Purnama, and A. B. Hikmah, "Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 189–194, 2019, doi: 10.33480/pilar.v15i2.669.
- [5] D. Evanko, "Optical imaging of the native brain," *Nat. Methods*, vol. 7, no. 1, p. 34, 2010, doi: 10.1038/nmeth.f.284.
- [6] I. D. Ayu Susilawati, "Mekanisme Hipotetikal Efek Konsumsi Ikan Menurunkan Resiko Serangan Jantung (Hypothetic Mechanisms of Fish Consumption Effects Reduce Risk of Heart Attack)," pp. 1–7, 2018.
- [7] A. Maulana and A. A. Fajrin, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, p. 27, 2018, doi: 10.20527/klik.v5i1.100.
- [8] D. Nofriansyah, K. Erwansyah, and M. Ramadhan, "Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat

Beli Pelanggan terhadap Kartu Internet XL ( Studi Kasus di CV. Sumber Utama Telekomunikasi),” *J. Saindikom*, vol. 15, no. 2, pp. 81–92, 2016.

- [9] B. Zahedi, B. Nahid-Mobarakeh, S. Pierfederici, and L. E. Norum, “A robust active stabilization technique for dc microgrids with tightly controlled loads,” *Proc. - 2016 IEEE Int. Power Electron. Motion Control Conf. PEMC 2016*, vol. VI, no. 1, pp. 254–260, 2016, doi: 10.1109/EPEPEMC.2016.7752007.
- [10] D. Sartika and D. I. Sensuse, “Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 3, no. 2, pp. 151–161, 2017, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/78>
- [11] S. Dewi, “Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan,” *Techno Nusa Mandiri*, vol. 13, no. 1, pp. 60–66, 2016.

### LAMPIRAN

No	umur	nis_kelami	yeri_dad	akan_dar	kolesterol	gula_darah	lektrokardiyut_per_m	angina	depresi	ringan_sej	erdiaknosa	gan_sej	erdiaknosa	
1	63	laki-laki	simtomati	145	233	benar	normal	150	tidak	2.3	ring_ke_ai	penyakit	g_ke_ai	penyakit
2	37	laki-laki	ri_non-an	130	250	salah	depresi	187	tidak	3.5	ring_ke_ai	penyakit	g_ke_ai	penyakit
3	41	perempu	angina_ati	130	204	salah	normal	172	tidak	1.4	ing_ke_ba	penyakit	_ke_ba	penyakit
4	56	laki-laki	gina_ati	120	236	salah	depresi	178	tidak	0.8	ing_ke_ba	penyakit	_ke_ba	penyakit
5	57	perempu	angina_tip	120	354	salah	depresi	163	ya	0.6	ing_ke_ba	penyakit	_ke_ba	penyakit
6	57	laki-laki	gina_tip	140	192	salah	depresi	148	tidak	0.4	datar	penyakit	datar	penyakit
7	56	perempu	angina_ati	140	294	salah	normal	153	tidak	1.3	datar	penyakit	datar	penyakit
8	44	laki-laki	gina_ati	120	263	salah	depresi	173	tidak	0	ing_ke_ba	penyakit	_ke_ba	penyakit
9	52	laki-laki	ri_non-an	172	199	benar	depresi	162	tidak	0.5	ing_ke_ba	penyakit	_ke_ba	penyakit
10	57	laki-laki	ri_non-an	150	168	salah	depresi	174	tidak	1.6	ing_ke_ba	penyakit	_ke_ba	penyakit
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	
294	58	perempu	angina_tip	170	225	benar	normal	146	ya	2.8	datar	lak_pen	datar	lak_pen
295	67	laki-laki	ri_non-an	152	212	salah	normal	150	tidak	0.8	datar	lak_pen	datar	lak_pen
296	44	laki-laki	gina_tip	120	169	salah	depresi	144	ya	2.8	ring_ke_at	lak_pen	g_ke_at	lak_pen
297	63	laki-laki	gina_tip	140	187	salah	normal	144	ya	4	ing_ke_ba	lak_pen	_ke_ba	lak_pen
298	63	perempu	angina_tip	124	197	salah	depresi	136	ya	0	datar	lak_pen	datar	lak_pen
299	59	laki-laki	gina_tip	164	176	benar	normal	90	tidak	1	datar	lak_pen	datar	lak_pen
300	57	perempu	angina_tip	140	241	salah	depresi	123	ya	0.2	datar	lak_pen	datar	lak_pen
301	45	laki-laki	simtomati	110	264	salah	depresi	132	tidak	1.2	datar	lak_pen	datar	lak_pen
302	68	laki-laki	gina_tip	144	193	benar	depresi	141	tidak	3.4	datar	lak_pen	datar	lak_pen
303	57	laki-laki	gina_tip	130	131	salah	depresi	115	ya	1.2	datar	lak_pen	datar	lak_pen
304	57	perempu	angina_ati	130	236	salah	normal	174	tidak	0	datar	lak_pen	datar	lak_pen