

## **BAB II**

### **KAJIAN PUSTAKA, KERANGKA PEMIKIRAN DAN HIPOTESIS**

#### **2.1 Kajian Pustaka**

##### **1.1.1 *Data Mining***

*Data Mining* atau penambangan data adalah teknik yang relatif cepat dan mudah untuk menemukan pengetahuan, pola dan atau relasi antar data, secara otomatis. Dengan menggabungkan empat disiplin ilmu komputer seperti pada definisi di atas, pengetahuan bisa diketemukan dalam lima proses berurutan: seleksi, pra-pemrosesan, transformasi, *Data Mining*, dan interpretasi/evaluasi (Suyanto, 2017).

*Data Mining* adalah proses menemukan pola-pola menarik dari jumlah data yang sangat kecil. Sebagai proses penemuan pengetahuan, biasanya melibatkan pembersihan data, integrasi data, pemilihan data, transformasi data, penemuan pola, evaluasi pola, dan presentasi pengetahuan (Han dkk, 2012).

*Data Mining* merupakan suatu proses untuk mengidentifikasi pola yang memiliki potensi dan berguna untuk mengelola dataset yang besar (Witten dkk, 2011). *Data Mining* merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data (Aprilla C dkk, 2013).

Teknik *Data Mining* seperti algoritma *Decision Tree* dapat digunakan untuk mengevaluasi aktivitas bisnis, misalnya dengan memetakan produk terpopuler dan memeriksa hubungan antara produk tersebut dengan produk lainnya (Rafdhi, A. Dkk, 2023).

Dapat disimpulkan bahwa *Data Mining* adalah proses analisa dengan menggali informasi dalam basis data dengan cara mengekstraksi untuk

menemukan suatu pola pengetahuan baru yang tidak diketahui sebelumnya dan akan bermanfaat pada masa yang akan datang.

### 1.1.2 Metode Klasifikasi

Salah satu tugas utama dari *Data Mining* adalah klasifikasi. Klasifikasi digunakan untuk menempatkan bagian yang tidak diketahui pada data ke dalam kelompok yang sudah diketahui. Klasifikasi menggunakan variabel target dengan nilai nominal. Dalam satu set pelatihan, variabel target sudah diketahui. Dengan pembelajaran dapat ditemukan hubungan antara fitur dengan variabel target. (Han dan Kamber, 2006)

Klasifikasi adalah suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan sebelumnya. Teknik ini dapat memberikan hasil berupa klasifikasi pada data baru dengan memanipulasi data yang telah diklasifikasi dan dengan menggunakan hasilnya untuk membuahkan sejumlah aturan (Aprilla C dkk, 2013).

Klasifikasi adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas atau konsep data. Model diturunkan berdasarkan analisis satu set data pelatihan (yaitu, objek data yang dikenal oleh label kelas). Model ini digunakan untuk memprediksi label kelas objek yang label kelasnya tidak diketahui (Han dkk, 2012).

Unsur-unsur untuk membuat suatu data dalam klasifikasi di antaranya:

- Kelas, merupakan variabel tidak bebas atau dapat dikatakan hasil dari klasifikasi.
- Prediktor, merupakan variabel bebas yang digunakan untuk menentukan pola atau karakteristik data yang akan diklasifikasikan.
- Set data *training*, merupakan sekumpulan data yang berisi kelas dan prediktor yang akan diuji untuk mendapat pengelompokan sesuai dengan kelas yang tepat.
- Set Data Uji, merupakan data-data baru yang akan di kelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

### 1.1.3 *Decision Tree*

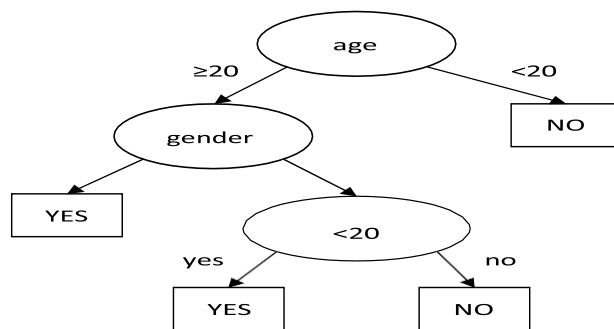
Di antara beberapa metode yang dapat digunakan untuk klasifikasi adalah metode *Decision Tree*. Metode *Decision Tree* merupakan sebuah metode yang dapat mengubah fakta yang sangat besar menjadi sebuah *Decision Tree* yang merepresentasikan aturan.

Sebuah *Decision Tree* adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan-kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagi, anggota himpunan hasil menjadi mirip satu dengan yang lainnya. Data dalam *Decision Tree* biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Proses pada *Decision Tree* adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*. Beberapa Algoritma yang bisa digunakan dalam pembentukan *Decision Tree*, di antaranya IDE3, CART, dan C4.5. (Mardi, 2017).

Dibandingkan dengan metode yang juga digunakan *Decision Tree* adalah salah satu metode yang paling populer untuk klasifikasi dalam berbagai aplikasi *Data Mining* dan membantu proses pengambilan keputusan. *Decision Tree* adalah pohon terarah dengan simpul akar yang tidak memiliki tepi masuk dan semua simpul lainnya dengan tepat satu ujung yang masuk, yang dikenal sebagai simpul keputusan. Pada tahap pelatihan, setiap simpul internal membagi ruang contoh menjadi dua atau lebih bagian dengan tujuan mengoptimalkan kinerja pengklasifikasi. Setelah itu, setiap jalur dari simpul akar ke simpul daun membentuk aturan keputusan untuk menentukan kelas mana yang merupakan turunan baru (Dai dkk, 2014).

*Decision Tree* khas terdiri dari *node* internal, tepi dan simpul daun. Setiap simpul *internal* disebut simpul keputusan yang mewakili pengujian pada atribut atau subkumpulan atribut, dan setiap sisi diberi label dengan nilai atau rentang nilai tertentu dari atribut masukan. Dengan cara ini, *node internal* yang terkait dengan tepiannya membagi ruang instansi menjadi dua atau lebih partisi.

Setiap simpul daun adalah simpul terminal pohon dengan label kelas. Misalnya, Gambar 2.2 memberikan ilustrasi dari *Decision Tree* dasar, di mana lingkaran berarti simpul keputusan dan persegi berarti simpul daun. Dalam contoh ini, memiliki tiga atribut pemisahan, yaitu usia, jenis kelamin, dan kriteria 3, bersama dengan dua label kelas, yaitu, YA dan TIDAK. Setiap jalur dari simpul akar ke daun *node* membentuk aturan klasifikasi (Dai dkk, 2014).



Gambar 2. 1 Contoh model *Decision Tree*

#### 1.1.4 Bahasa Pemrograman *Python*

Bahasa pemrograman *Python* merupakan bahasa pemrograman yang dapat digunakan untuk melakukan pemrograman atau pengembangan pada bidang yang sangat luas seperti untuk analisis data statistik, pengembangan Internet dan situs web, akses *database*, GUI desktop, komputasi ilmiah, serta *game* dan pengembangan perangkat lunak lainnya. Bahasa pemrograman *Python* mendapatkan popularitas luar biasa di kalangan ilmuwan data dan pengembang perangkat lunak (Robinson, 2017). Terdapat dua seri versi utama *Python*, versi 2.x dan 3.x, dan keduanya tidak sepenuhnya kompatibel meskipun sebagian besar serupa. Versi 2.x adalah versi lawas, yang dukungan dan pemeliharannya dijadwalkan berakhir sekitar tahun 2020. Versi 3.x merupakan desain ulang berdasarkan versi 2.x dan dianggap sebagai masa depan *Python*. Dengan adanya dua versi utama tersebut referensi penggunaan tetap diserahkan sepenuhnya kepada pengguna baik menggunakan *Python* versi 2.x atau pun

versi 3.x, tergantung kebutuhan dari digunakannya bahasa *Python* itu sendiri (Hao, J., & Ho, T. K, 2019).

*Python* bukanlah bahasa yang dikompilasi, artinya ia tidak melakukan prakompilasi kode ke dalam biner. Sebaliknya, lingkungan perangkat lunak, juru bahasa *Python*, menerjemahkan skrip ke dalam biner selama eksekusi kode secara *real time*. Dengan distribusinya, *Python* hadir dengan beberapa fungsi dasar tetapi bergantung pada paket eksternal untuk melakukan hampir semua perhitungan numerik. Setelah proses seleksi alam selama 10 tahun terakhir, sekumpulan kecil paket yang menyediakan beberapa kemampuan komputasi dasar telah diterima secara luas di komunitas pengguna bahasa pemrograman *Python* (Hao, J., & Ho, T. K, 2019).

#### **1.1.4.1 Scikit-Learn**

Scikit-learn adalah pustaka atau modul untuk *data mining* dan pembelajaran mesin paling komprehensif dan *open source* dari bahasa pemrograman *Python* (Gholizadeh, S, 2022). Karena pembelajaran mesin sering kali merupakan komponen aplikasi yang lebih umum (seperti layanan Web), sebaiknya komponen tersebut dilengkapi dengan bahasa pemrograman yang sama dengan bagian lain aplikasi untuk integrasi yang lancar. Memanfaatkan beragam penerapan *Python*, Scikit-learn menjadi paket yang sangat populer untuk aplikasi terkait pembelajaran mesin (Hao, J., & Ho, T. K, 2019).

Selain backup yang kuat dari ekosistem *Python*, Scikit-learn sendiri memiliki banyak fitur yang membuatnya menonjol di antara software machine learning. Yang pertama adalah cakupan metode pembelajaran mesin yang komprehensif. Prosedur tinjauan komunitas tersedia untuk mengidentifikasi dan memutuskan metode pembelajaran mesin mana yang harus disertakan dalam paket Scikit-learn. Mekanisme seperti itu memastikan keseimbangan antara cakupan luas dan selektivitas metode pembelajaran mesin yang terdapat dalam paket. Kedua, implementasi algoritma metode pembelajaran mesin di Scikit-learn dioptimalkan untuk efisiensi komputasi. Meskipun *Python* adalah bahasa

pemrograman interpretatif, sebagian besar metode pembelajaran mesin di Scikit-learn didasarkan pada pustaka biner terkompilasi yang awalnya diprogram dalam Fortran, C, atau C++. Implementasi berbasis biner ini secara signifikan meningkatkan efisiensi komputasi. Yang ketiga adalah Scikit-learn memiliki dukungan komunitas yang kuat untuk dokumentasi, pelacakan bug, dan jaminan kualitas. Komunitas Scikit-learn mengelola dokumentasi umum, proses pelacakan/perbaikan bug terpadu di GitHub, dan prosedur jaminan kualitas yang ketat berdasarkan kesepakatan komunitas. Terakhir, Scikit-learn menerapkan konvensi data input/output terpadu dan memiliki prosedur penyesuaian model tetap yang membuat peralihan dari satu metode ke metode lainnya menjadi mudah (Hao, J., & Ho, T. K, 2019).

Singkatnya, Scikit-learn adalah kumpulan metode yang dapat digunakan dalam *data mining* dan pembelajaran mesin yang diterapkan secara efisien dan didokumentasikan dengan baik serta dipelihara oleh komunitas penggunanya.

#### **1.1.4.2 Pandas**

*Pandas* adalah pustaka atau modul yang berisi alat analisis dan manipulasi data sumber terbuka yang cepat, kuat, fleksibel, dan mudah digunakan yang dibangun di atas bahasa pemrograman *Python*. *Pandas* digunakan untuk melakukan proses yang berurusan dengan jumlah data yang besar dan melakukan proses seperti analisis data serta menyediakan cara sederhana untuk membersihkan, memanipulasi, dan mengubah data (Gholizadeh, S. 2022). Fitur unggulan di *Pandas* dapat dikategorikan sebagai berikut.

1. Jelajahi & Analisis data dengan cepat
2. Membaca berbagai format file
3. Membersihkan data
4. Memanipulasi data

*Pandas* bekerja dengan objek *Data Frame*, *Pandas Dataframe* adalah Struktur Data 2 Dimensi di mana data disimpan secara tabel berupa baris &

kolom (Gholizadeh, S. 2022). Beberapa perusahaan menggunakan *Pandas* sebagai sistem rekomendasi, seperti *Netflix* yang menggunakan kumpulan besar data tentang preferensi pelanggannya untuk memberikan saran kepada penggunanya. *Amazon* melakukan analisis data ekstensif untuk menciptakan sistem rekomendasi yang kuat. *YouTube* menggunakan analisis data untuk merekomendasikan video kepada penggunanya. *Pandas* juga digunakan di berbagai domain seperti Layanan Kesehatan untuk menilai risiko penyakit kronis dan kanker (Meniaailov, I, 2019), Sektor Energi untuk meningkatkan kinerja dan mengurangi biaya pemeliharaan dengan memprediksi kegagalan perangkat (Wang, K, 2020), Organisasi e-niaga menggunakan *Pandas* untuk pelanggan segmentasi (Khandelwal, R, 2019). Kemudian perusahaan dapat menganalisis data pelanggan untuk menyediakan Iklan yang dipersonalisasi dan operator maskapai penerbangan menganalisis perilaku pelanggan mereka untuk memotong biaya (Safara, F, 2022).

### **1.1.5 Algoritma CART**

Algoritma CART (*Classification and Regression Trees*) adalah algoritma pembangunan pohon keputusan yang digunakan untuk pemodelan klasifikasi dan regresi. Pohon keputusan adalah model pembelajaran mesin yang populer karena kemudahan interpretasinya. Pohon ini memecah data menjadi subset yang lebih kecil berdasarkan serangkaian keputusan berhierarki. (Aziza, E. Z, 2019).

Berikut adalah langkah-langkah utama dalam algoritma CART untuk klasifikasi:

#### **1. Pemilihan Fitur Pemisahan:**

Algoritma CART memulai dengan memilih fitur yang akan digunakan untuk pemisahan. Ini dilakukan dengan mengukur *impurity* (ketidakmurnian) setiap fitur untuk setiap simpul (*node*) saat ini. *Impurity* diukur dengan metrik seperti Gini *impurity* atau Entropy. Pemilihan fitur yang menghasilkan pemisahan dengan *impurity* paling rendah dipilih sebagai fitur pemisahan terbaik.

## 2. Pemisahan Data:

Data dibagi menjadi dua (atau lebih) subset berdasarkan fitur pemisahan terbaik dan nilai ambang tertentu. Setiap subset sesuai dengan salah satu cabang pohon.

## 3. Rekursi:

Langkah di atas diulang secara rekursif untuk setiap anak simpul (*child nodes*). Pemilihan fitur pemisahan dan pemisahan data berlanjut hingga mencapai kriteria berhenti, seperti kedalaman maksimum pohon atau jumlah sampel minimum dalam simpul.

## 4. Penyusunan Prediksi:

Ketika pembangunan pohon selesai, pohon keputusan digunakan untuk membuat prediksi klasifikasi untuk data uji dengan mengikuti cabang-cabang pohon dari simpul akar hingga mencapai simpul daun (*leaf node*).

### 1.1.6 Penyakit Jantung

Penyakit Jantung merupakan sebutan umum yang digunakan untuk menggambarkan gangguan terhadap fungsi kerja jantung. Penyakit atau gangguan jantung sendiri memiliki banyak jenis dan macam nama penyakitnya seperti kardiovaskuler, jantung koroner dan serangan jantung. Penyakit Jantung merupakan salah satu penyakit yang paling sering terjadi kasusnya pada kalangan masyarakat, di mana penyakit jantung ini dapat menimpa dan menyerang siapa pun tanpa memandang usia, jenis kelamin dan gaya hidup. Menurut WHO (Organisasi Kesehatan Dunia) dan CDC, penyakit jantung adalah penyebab utama kematian di Inggris, Amerika Serikat, Kanada dan Australia. Jumlah orang dewasa yang didiagnosis dengan penyakit jantung terdiri dari 26,6 Juta Jiwa (11,3%) dari populasi orang dewasa. (R. Annisa, 2019)

Menurut D. B. Umadevi, (2017) Penyakit jantung salah satu penyakit yang memiliki angka kematian dengan tingkat yang tinggi lebih dari 12 juta jiwa kematian yang terjadi pada seluruh dunia dikarenakan penyakit jantung ini.



Dengan demikian diagnosa secara dini sangat penting untuk dilakukan, diagnosa pada penyakit jantung merupakan hal yang sangat menantang dikarenakan saling ketergantungan yang kompleks dari beberapa faktor atribut. Permasalahan yang sering kali dihadapi adalah kurangnya akurasi pada proses klasifikasi. (A. Jain, 2019).

### **1.1.7 COVID-19**

COVID-19 merupakan jenis *coronavirus* yang pertama kali ditemukan di Wuhan, Tiongkok, bulan Desember 2019 dan dapat menular melalui saluran pernapasan pada hewan atau manusia, COVID-19 juga dinyatakan sebagai sebuah pandemi di banyak negara di seluruh dunia termasuk Indonesia (Setiati, 2020). COVID-19 pun menyebabkan banyak kerugian di berbagai sektor. Hasil pendataan menunjukkan adanya penurunan dari berbagai sektor pada masa pandemi Covid-19 akibat kebijakan pembatasan sosial lainnya yang menghambat aktivitas. (Soeryanto Soegoto, E., 2022).

Kasus penyebaran Covid-19 sudah dimulai di benua Asia sejak tahun 2020, di Indonesia sendiri Covid-19 masuk pada Maret 2020 (Putra, Y. H. dkk, 2022). Kemudian COVID-19 menyebar sehingga menyebabkan ribuan orang meninggal dan terus bertambah tiap harinya. Oleh karena itu, sesuai anjuran World Health Organization (WHO) pemerintah Indonesia menerapkan *social distancing*, di mana tiap warga Indonesia harus menjaga jarak sejauh 2 meter dan mematuhi protokol kesehatan seperti memakai masker, rajin mencuci tangan, melakukan deteksi suhu tubuh saat berada di tempat umum agar terhindar dari COVID-19. Akan tetapi masih banyak masyarakat Indonesia yang masih tidak melakukan *social distancing* atau mematuhi protokol kesehatan saat berada di tempat umum seperti di pasar atau saat menggunakan transportasi umum karena rendahnya kesadaran dan kewaspadaan masyarakat Indonesia. Hal ini terbukti dari jumlah angka kematian yang dipaparkan dari web resmi Kementerian Kesehatan Indonesia per tanggal 20 Mei 2020 mencapai 1.242 jiwa sejak kasus COVID-19 pertama di Indonesia.

### 1.1.8 Penelitian Terdahulu

No	Peneliti	Judul	Persamaan	Perbedaan
1	Riani, A., Susianto, Y., & Rahman, N. <i>Journal of Innovation Information Technology and Application (JINITA). 1(01), 25-34. 2019</i>	Implementasi <i>Data Mining</i> Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes	Menggunakan <i>Data Mining</i> untuk analisis prediksi penyakit jantung	Penelitian ini menggunakan metode <i>naive bayes</i>
2	Utomo, D. P., Sirait, P., & Yunis, R. <i>Jurnal Media Informatika Budidarma, 4(4), 994-1006. 2020</i>	Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5. 0	Menerapkan <i>Data Mining</i> Klasifikasi untuk analisis pada data set penyakit jantung	Penelitian ini bertujuan untuk reduksi atribut dan menggunakan metode <i>algoritma C5.0</i>
3	Annisa, R. <i>JTIK (Jurnal Teknik Informatika Kaputama), 3(1), 22-28. 2019</i>	Analisis Komparasi Algoritma Klasifikasi <i>Data Mining</i> Untuk Prediksi Penderita Penyakit Jantung	Menerapkan <i>Data Mining</i> Klasifikasi untuk analisis pada data set penyakit jantung	Penelitian ini bertujuan untuk komparasi algoritma dan menggunakan metode <i>algoritma C4.5, naive bayes, dan k-NN</i>

No	Peneliti	Judul	Persamaan	Perbedaan
4	Utomo, D. P., & Mesran, M. <i>Jurnal Media Informatika Budidarma</i> , 4(2), 437-444. 2020	Analisis Komparasi Metode Klasifikasi <i>Data Mining</i> dan Reduksi Atribut Pada Data Set Penyakit Jantung	Menerapkan <i>Data Mining</i> Klasifikasi untuk analisis pada data set penyakit jantung	Penelitian ini bertujuan untuk komparasi algoritma dan reduksi atribut menggunakan metode <i>algoritma C5.0 dan naive bayes</i>
5	Azis, H., Purnawansyah, P., Fattah, F., & Putri, I. P. <i>ILKOM Jurnal Ilmiah</i> , 12(2), 81-86. 2020	Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung	Menerapkan <i>Data Mining</i> Klasifikasi untuk analisis pada data set penyakit jantung	Menggunakan metode <i>k-NN dan Cross Validation</i>
6	Firdlous, D. A. <i>Infoman's: Jurnal Ilmu-ilmu Manajemen dan Informatika</i> , 16(1), 79-84. 2022	Komparasi Algoritma Klasifikasi <i>Data Mining</i> untuk Memprediksi Penyakit Jantung	Menerapkan <i>Data Mining</i> Klasifikasi untuk analisis pada data set penyakit jantung	Menggunakan metode <i>naive bayes, random forest, Decision Tree</i> , dan <i>support vector machine</i>

No	Peneliti	Judul	Persamaan	Perbedaan
7	Maulana, D., & Yahya, R.	Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Penderita Penyakit Jantung Di Indonesia Menggunakan Rapid Miner	Menerapkan <i>Data Mining</i> Klasifikasi untuk analisis pada data set penyakit jantung	Menggunakan metode <i>naive bayes</i> dan <i>rapid miner</i>

Tabel 2. 1 Penelitian terdahulu

## 2.2 Kerangka Pemikiran

*Data Mining* merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data. Penerapan *Data Mining* sudah dilakukan pada beberapa bidang di antaranya bisnis, ilmu pengetahuan dan kesehatan. Pada bidang kesehatan, *Data Mining* dapat dimanfaatkan untuk memprediksi suatu penyakit dari data rekam medis pasien. Dengan metode klasifikasi pada *Data Mining*, data seperti umur, jenis kelamin, tekanan darah dan atribut lainnya, dapat digunakan untuk memprediksi kemungkinan pasien terkena suatu penyakit.

Klasifikasi adalah proses analisis data yang menghasilkan model untuk menggambarkan kelas-kelas yang ada dalam data. Model-model ini disebut classifier. Oleh karena itu, classifier ini akan digunakan untuk menyusun kelas-kelas yang terkandung dari data, misalnya untuk *Decision Tree* maka kelas-kelas tersebut digambarkan dalam bentuk pohon.

Algoritma *Decision Tree* digunakan untuk mempelajari klasifikasi dan prediksi model dari data dan menggambarkan hubungan antara variabel atribut x dan variabel target y sebagai pohon. Sebuah *Decision Tree* adalah struktur seperti *flowchart* di mana setiap *node* dalam (*node non-leaf* atau bukan *node*

terluar) adalah pengujian variabel atribut, setiap cabang adalah hasil pengujian, sedangkan *node* terluar yaitu daun, adalah labelnya.

Berdasarkan faktor-faktor yang belum diketahui secara pasti faktor yang mana yang menjadi pengaruh terbesar dari meningkatnya risiko penyakit jantung pada penderita COVID-19, penelitian ini bertujuan untuk melakukan identifikasi dan ekstraksi dari informasi yang sudah ada untuk mendapatkan pengetahuan untuk mengetahui faktor manakah yang paling mempengaruhi risiko penyakit jantung pada penderita COVID-19 untuk dilakukan prediksi risiko penyakit jantung pada penderita COVID-19.

### 2.3 Hipotesis

Sugiyono (2002:39) mengemukakan pengertian tentang hipotesis sebagaiberikut : “ hipotesis merupakan jawaban sementara terhadap rumusan masalah penelitian, dikatakan sementara karena jawaban yang diberikan baru didasarkan pada fakta-fakta yang empiris yang diperoleh melalui pengumpulan data”

Berdasarkan pendapat tersebut, maka disimpulkan bahwa hipotesis merupakan jawaban sementara yang jawabannya belum final karena harus dibuktikan kebenarannya melalui penelitian. Berdasarkan rumusan permasalahan di atas, maka hipotesis utama yang penulis ajukan bahwa **“*Data Mining* Klasifikasi Dengan Algoritma *Decision Tree* dapat Memprediksi Risiko Penyakit Jantung pada Penderita Covid-19 dengan Sangat Akurat dan dapat Menghasilkan Pengetahuan Berupa Atribut yang Paling Berpengaruh Dalam Risiko Penyakit Jantung Pada Penderita COVID-19”**.