

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Dalam penelitian ini, klasifikasi penyakit jantung menggunakan metode *naïve bayes*. Tahap awal yang dilakukan adalah mengumpulkan data tentang penyakit jantung serta gejala-gejala yang muncul pada orang yang terkena penyakit jantung. Dalam penelitian ini ada beberapa yang memuat masalah sejenis dengan metode analisis yang sama dengan penelitian yang sedang diteliti :

Penelitian klasifikasi penyakit paru-paru dengan menggunakan metode *naïve bayes classifier* (Haffandi *et al.*, 2022) Data yang diambil berupa rekam medis penyakit paru-paru pasien bulan Juli sampai Agustus sebanyak 134 data pasien yang berisi 19 gejala penyakit dan 6 penyakit diagnosa. Dari hasil pengujian menggunakan aplikasi Rapidminer dan pemisahan data berupa 34 data uji dan 100 data latih dengan perbandingan data 7:3 diperoleh nilai akurasi sebesar 97,06 diperoleh.

Penelitian perancangan system tentang penyakit jantung (Bianto, Kusrini and Sudarmawan, 2020) Hasil yang diperoleh dalam penelitian ini adalah penerapan system klasifikasi dengan menggunakan metode *naïve bayes* yang akan menghasilkan nilai rata-rata akurasi sebesar 90.61%, presisi sebesar 87.44% dan recall sebesar 87.95%.

Penelitian implementasi Data Mining tentang penyakit jantung (Riani, Susianto and Rahman, 2019) Penelitian ini akan melakukan perhitungan menggunakan metode Data Mining dengan Algoritma Naïve Bayes. Hasil dari penelitian ini mendapatkan akurasi sebesar 86% untuk 303 dataset yang diuji.

Penelitian klasifikasi tentang penyakit Stroke (Paramitha et al, Aang Nuryaman , Ahmad Faisol, Eri Setiawan, Dina Eka 2023). Tujuan penelitian ini adalah untuk mengklasifikasi pasien penyakit stroke menggunakan metode Naïve Bayes untuk beberapa proporsi data training dan data testing yang berbeda. Data yang digunakan adalah data pasien penyakit stroke dengan kelas terkena stroke atau tidak terkena stroke dan menggunakan 10 variabel bebas. Hasil klasifikasi dengan nilai akurasi terbesar 80% diperoleh, ketika proporsi data training dan data testing 80:20.

Penelitian system prediksi tentang penyakit jantung (Al Hafiz et al. Dimsyiar M Al Hafiz<sup>1</sup>, Khoirul Amaly<sup>1</sup>, Javen Jonathan<sup>1</sup>, M. Teranggono Rachmatullah<sup>1</sup>, Rosidi<sup>1</sup>, 2021). Eksperimen simulasi dilakukan untuk melihat apakah system yang dibangun menggunakan metode Naïve Bayes mampu untuk memprediksi apakah seseorang mengidap penyakit jantung Berdasarkan eksperimen data uji berjumlah 61 data, 52 diantaranya terprediksi benar dengan nilai keakuratan sebesar 85,25%.

Penelitian tentang Komparasi 3 Metode Algoritma Klasifikasi Hasil Prediksi Herregistrasi. Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Teknologi Digital Indonesia) oleh (Tyas Yonatama,2020). Hasil dari

penelitian ini adalah sistem yang dapat memprediksi data dengan kelas yaitu registrasi dan tidak registrasi dari 1704 data testing yang diproses menggunakan sistem didapatkan hasil prediksi registrasi sebanyak 1226 data dan tidak registrasi 478 data. Untuk pengujian dari 731 data didapatkan hasil prediksi 679 data terprediksi benar dan 52 data salah prediksi tingkat akurasi probabilitas yang didapatkan sebesar 92,88%.

Tabel 2. 1 Perbandingan dengan Peneliti Sebelumnya

| No | Penulis                         | Objek              | Masalah   | Metode                 | Hasil  |
|----|---------------------------------|--------------------|---|------------------------|--|
| 1. | (Haffandi <i>et al.</i> , 2022) | Penyakit Paru-Paru | Klasifikasi penyakit paru-paru dengan Menggunakan metode Naïve Bayes Classifier | Naïve Bayes Classifier | Dari proses perhitungan algoritma Naïve Bayes Classifier pada Rapid Miner dengan menggunakan 34 data testing dan 100 data training dengan perbandingan data 7:3 didapatkan nilai akurasi sebesar 97,06%. |

| No | Penulis  | Objek               | Masalah   | Metode      | Hasil  |
|----|--|---------------------|---|-------------|--|
| 2. | Mufti Ari<br>Bianto,<br>Kusrini,<br>Sudarmawan<br>(2019)   | Penyakit<br>Jantung | Perancangan Sistem<br>Klasifikasi Penyakit jantung<br>menggunakan<br>Naïve Bayes                    | Naïve Bayes | Hasil yang diperoleh dalam penelitian ini adalah penerapan system klasifikasi dengan menggunakan metode naïve bayes yang akan menghasilkan nilai rata-rata akurasi sebesar 90.61%, presisi sebesar 87,44% dan recall sebesar 87.95%. |
| 3. | Ade Riani ,<br>Yessy<br>Susianto , Nur<br>Rahman<br>(2019) | Penyakit<br>Jantung | Implementasi Data Mining<br>Untuk Memprediksi Penyakit<br>Jantung Menggunakan Metode<br>Naive Bayes | Naïve Bayes | Hasil yang diperoleh dalam penelitian ini adalah penerapan system klasifikasi dengan menggunakan metode naïve bayes yang akan menghasilkan nilai rata-rata akurasi sebesar 90.61%, presisi sebesar 87,44% dan recall sebesar 87.95%. |

| No | Penulis   | Objek                                      | Masalah  | Metode      | Hasil   |
|----|---|--|--|-------------|---|
| 4. | Nabilla<br>Yolanda<br>Paramita,<br>Aang<br>Nuryaman ,<br>Ahmad Faisol<br>, Eri Setiawan<br>, Dina Eka<br>Nurvazly<br>(2023) | Penyakit<br>Stroke                         | Klasifikasi Penyakit Stroke<br>Menggunakan Metode Naïve<br>Bayes   | Naïve Bayes | Hasil klasifikasi dengan nilai<br>akurasi terbesar 80%<br>diperoleh Ketika proporsi<br>data training dan data testing<br>80:20.   |
| 5. | Dimsyiar M<br>Al Hafiz1,<br>Khoirul<br>Amaly1 ,2021   | Penyakit<br>Jantung                        | Sistem Prediksi Penyakit<br>Jantung Menggunakan<br>Metode Naive Bayes  | Naïve Bayes | Berdasarkan eksperimen<br>data uji berjumlah 61 data,<br>52 diantaranya terprediksi<br>benar dengan nilai<br>keakuratan<br>sebesar 85,25%.  |
| 6. | Tyas<br>Yanotama,<br>2020   | Memprediks<br>i calon<br>mahasiswa<br>baru | Komparasi 3 Metode<br>Algoritma Klasifikasi Hasil<br>Prediksi Herregistrasi Pada<br>Penerimaan Mahasiswa Baru<br>(Studi Kasus: Universitas<br>Teknologi Digital Indonesia) | Naïve Bayes | Hasil prediksi registrasi<br>sebanyak 1226 data dan<br>tidak registrasi 478 data<br>Untuk pengujian dari 731<br>data didapatkan hasil<br>prediksi 679 data terprediksi<br>benar dan 52 data salah<br>prediksi Tingkat akurasi<br>probabilitas yang didapatkan<br>sebesar 92,88% |

| No | Penulis         | Objek               | Masalah   | Metode      | Hasil  |
|----|-----------------|---------------------|---|-------------|--|
| 7. | Debora,<br>2023 | Penyakit<br>Jantung | Klasifikasi Penyakit Jantung<br>Menggunakan Metode Naïve<br>Bayes | Naïve Bayes | Hasil akurasi dari dataset yang berjumlah 270 yaitu 87,03%, presisi 0,84%, recall 0,87,5% dengan jumlah data training 216 dan data testing 54. |

## 2.2 Dasar Teori

### 2.2.1 Penyakit Jantung

Penyakit jantung merupakan suatu kondisi dimana terdapat gangguan pada organ jantung. Penyakit jantung dapat ditandai dengan beberapa gejala yang perlu diwaspadai. Gejala-gejala penyakit jantung diantaranya adalah adanya rasa mual dan muntah, berkeringat dingin dan perasaan mudah lelah, sakit kepala, nyeri dada sebelah kiri, sesak napas, lemas jantung berdebar.(kamila,dkk 2023)

Penyakit jantung merujuk pada sekelompok kondisi yang mempengaruhi fungsi normal jantung. Penyakit jantung termasuk di antaranya penyakit jantung koroner, gagal jantung, aritmia, dan penyakit jantung bawaan. Kondisi ini dapat mempengaruhi kemampuan jantung untuk memompa darah dengan efektif, menyebabkan gangguan sirkulasi darah, dan berpotensi menjadi ancaman serius bagi kesehatan seseorang. Penyebab penyakit jantung dapat bervariasi tergantung pada jenisnya. Beberapa faktor risiko umum yang dapat menyebabkan penyakit jantung antara lain:

1. Kolesterol Tinggi: Tingginya kadar kolesterol dalam darah dapat menyebabkan penyakit jantung koroner.
2. Tekanan Darah Tinggi: Tekanan darah tinggi yang tidak terkontrol dapat memengaruhi kesehatan jantung dan menyebabkan komplikasi seperti gagal jantung.
3. Merokok: Merokok dan paparan asap rokok pasif dapat meningkatkan risiko penyakit jantung koroner.
4. Diabetes: Diabetes yang tidak terkontrol dapat merusak pembuluh darah dan menyebabkan kerusakan pada jantung.
5. Obesitas: Kelebihan berat badan atau obesitas meningkatkan risiko penyakit jantung koroner dan gagal jantung.
6. Kurang Aktivitas Fisik: Gaya hidup yang kurang aktif secara fisik dapat meningkatkan risiko penyakit jantung.
7. Riwayat Keluarga: Riwayat keluarga dengan penyakit jantung dapat meningkatkan risiko seseorang terkena kondisi serupa.

Gejala penyakit jantung dapat bervariasi tergantung pada jenis penyakit jantung yang dialami seseorang. Beberapa gejala umum yang mungkin timbul meliputi:

1. Nyeri atau ketidaknyamanan di dada (angina).
2. Sesak napas atau sulit bernapas.
3. Kelelahan yang berlebihan.
4. Pusing atau pingsan.

5. Pembengkakan pada kaki, pergelangan kaki, atau area lainnya.
6. Detak jantung yang tidak teratur atau berdebar-debar.
7. Nyeri atau ketidaknyamanan di bagian atas perut, punggung, atau lengan.
8. Stres: Stres yang berkepanjangan dapat memengaruhi kesehatan jantung.

Proses diagnosis penyakit jantung melibatkan langkah-langkah berikut:

1. Riwayat Medis: Dokter akan mewawancarai pasien mengenai gejala yang dialami, riwayat kesehatan, dan riwayat keluarga terkait penyakit jantung.
2. Pemeriksaan Fisik: Dokter akan melakukan pemeriksaan fisik untuk mencari tanda-tanda penyakit jantung, seperti mendengarkan suara jantung dan memeriksa tekanan darah.
3. Tes Penunjang: Tes penunjang yang mungkin dilakukan meliputi tes darah, elektrokardiogram (EKG), echocardiogram, tes stres, angiografi koroner, atau tes pencitraan jantung.

Pilihan pengobatan penyakit jantung akan disesuaikan dengan jenis dan tingkat keparahan penyakit yang dialami. Beberapa opsi pengobatan yang umum meliputi:

1. Perubahan Gaya Hidup: Mengadopsi gaya hidup sehat seperti berhenti merokok, menjaga berat badan yang sehat, mengatur pola makan seimbang, berolahraga secara teratur, dan mengelola stres.
2. Obat-obatan: Dokter dapat meresepkan obat-obatan seperti aspirin, beta-blocker, statin, atau anti-koagulan untuk mengendalikan tekanan darah, kolesterol, dan mencegah pembekuan darah.



3. **Prosedur Medis:** Prosedur medis seperti angioplasti koroner dengan stent, pemasangan ring jantung, atau operasi bypass jantung dapat direkomendasikan dalam kasus-kasus tertentu.
4. **Terapi Rehabilitasi Jantung:** Terapi rehabilitasi jantung melibatkan program latihan terapeutik, pendidikan kesehatan, dan dukungan psikososial untuk membantu pemulihan dan pengelolaan penyakit jantung.

Pencegahan penyakit jantung sangat penting untuk menjaga kesehatan jantung.

Beberapa langkah yang dapat diambil untuk mencegah penyakit jantung antara lain:

1. **Mengadopsi Gaya Hidup Sehat:** Mengonsumsi makanan sehat, berolahraga secara teratur, berhenti merokok, dan mengelola stres.
2. **Memantau Tekanan Darah dan Kolesterol:** Memantau tekanan darah dan kolesterol secara teratur dan mengambil langkah-langkah untuk menjaga agar dalam batas normal.
3. **Menjaga Berat Badan yang Sehat:** Memelihara berat badan yang sehat dengan menjaga keseimbangan antara asupan kalori dan aktivitas fisik.
4. **Menghindari Konsumsi Alkohol Berlebihan:** Konsumsi alkohol berlebihan dapat meningkatkan risiko penyakit jantung.
5. **Menjaga Kesehatan Mental:** Mengelola stres dan menjaga kesehatan mental secara keseluruhan dapat membantu mengurangi risiko penyakit jantung.

*(Kemenkes, 2023) <https://ayosehat.kemkes.go.id/topik-penyakit/penyakit-kardiovaskular/penyakit-jantung>*

### **2.2.2 Klasifikasi**

Klasifikasi adalah jenis analisis data yang dapat membantu orang memprediksi label kelas sampel harus diklasifikasikan, berbagai teknik klasifikasi telah diusulkan dalam bidang-bidang seperti pembelajaran mesin, system pakar dan statistik. Perkembangan teknologi informasi dan ilmu komputasi telah membuka peluang yang sangat besar untuk menjadi bagian dari solusi alternatif penyelesaian masalah terkait sulitnya identifikasi penyakit jantung. Hal ini membuat banyak para peneliti melakukan penelitian terhadap penyakit jantung, salah satunya dengan cara metode yang berbasis komputer. Cara ini sudah banyak digunakan dengan komputasi yang begitu cerdas dan memproses data dengan jumlah yang begitu banyak, pengolahan data tersebut tentunya dilakukan dengan cara klasifikasi menggunakan beberapa metode dengan hasil yang akurat dan cepat. (B.Hirwono, dkk, 2023)

### **2.2.3 Python**

Python merupakan Bahasa pemrograman yang dapat dijalankan pada berbagai sistem operasi baik Linux, MacOS ataupun Windows dengan melakukan konfigurasi terlebih dahulu. Python merupakan bahasa pemrograman tingkat tinggi dikarenakan kode-kode yang ditulis akan decompile menjadi byte code serta di eksekusi, sehingga membuat python cocok digunakan untuk bahasa pemrograman scripting aplikasi web dan lain sebagainya. (S.Suryono, dkk. 2018)

#### 2.2.4 Pengklasifikasi Naïve Bayes Gaussian

Pengklasifikasi naïve bayes Gaussian adalah pengklasifikasi probabilistik berdasarkan teorema bayes, yang mengasumsikan bahwa setiap fitur memberikan kontribusi yang independen dan setara ke atas target. Pengklasifikasi naïve bayes gaussian mudah di implementasikan secara komputasi cepat dan berkinerja baik pada kumpulan data besar yang memiliki dimensi tinggi. kondusif untuk aplikasi waktu nyata dan tidak sensitive terhadap noise. Pengklasifikasi naïve bayes gaussian memproses data set pelatihan untuk menghitung probabilitas kelas dan probabilitas bersyarat, yang menentukan frekuensi setiap nilai fitur untuk nilai kelas tertentu dibagi dengan frekuensi instance dengan nilai kelas tersebut. Pengklasifikasi naïve bayes gaussian berkinerja terbaik, ketika fitur yang berkorelasi dihapus karena fitur yang berkorelasi akan di pilih dua kali dalam model yang mengarah kepenekanan berlebihan pada pentingnya fitur yang berkorelasi. (Hafiz, 2021).

Distribusi normal memiliki persamaan matematika yang menentukan probabilitas satu observasi berada dalam salah satu kelompok. Rumusnya adalah:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Dimana  $\mu$  adalah mean dan  $\sigma$  adalah simpangan baku.

Penjelasan:

- $f(x)$ : Menyatakan fungsi distribusi probabilitas yang bergantung pada variabel acak  $x$ .

- $\mu$  (mu): Menyatakan mean (rata-rata) dari distribusi.
- $\sigma$  (sigma): Menyatakan standar deviasi dari distribusi.
- $\pi$  (pi): Merupakan konstanta matematika, sekitar 3.14159.
- $e$ : Merupakan basis dari logaritma natural, sekitar 2.71828.

Rumus ini memberikan nilai probabilitas suatu variabel acak  $x$  berada pada nilai tertentu dalam distribusi normal. Distribusi normal adalah distribusi probabilitas yang berbentuk lonceng dan simetris, yang banyak digunakan dalam statistik dan ilmu sosial untuk menggambarkan data yang secara alami menyebar di sekitar mean.

Fungsi distribusi normal digunakan untuk menentukan seberapa besar kemungkinan suatu nilai  $x$  terjadi berdasarkan rata-rata  $\mu$  dan standar deviasi  $\sigma$  dari populasi.

Gaussian Naive Bayes adalah teknik klasifikasi pembelajaran mesin yang didasarkan pada pendekatan probabilistik yang mengasumsikan setiap kelas mengikuti distribusi normal. Ia mengasumsikan setiap parameter memiliki kapasitas independen untuk memprediksi variabel output. Ia mampu memprediksi probabilitas variabel dependen untuk diklasifikasikan dalam setiap kelompok. *(Carla Martins, 2023)*

Gaussian Naive Bayes (GNB) merupakan salah satu teknik klasifikasi yang digunakan dalam machine learning berdasarkan pendekatan probabilistik dan distribusi Gaussian. Gaussian Naive Bayes mengasumsikan bahwa setiap parameter, yang juga disebut fitur atau prediktor, memiliki kapasitas independen untuk memprediksi variabel output. Kombinasi prediksi untuk semua parameter adalah prediksi akhir yang menghasilkan probabilitas variabel dependen untuk

diklasifikasikan dalam setiap kelompok. Klasifikasi akhir diberikan kepada kelompok dengan probabilitas lebih tinggi. (Carla Martins,2023)

Distribusi Gaussian juga disebut distribusi normal. Distribusi normal adalah model statistik yang menggambarkan distribusi variabel acak kontinu di alam dan didefinisikan oleh kurva berbentuk lonceng. Dua fitur terpenting dari distribusi normal adalah mean ( $\mu$ ) dan simpangan baku ( $\sigma$ ). Mean adalah nilai rata-rata dari suatu distribusi, dan simpangan baku adalah "lebar" distribusi di sekitar mean. Variabel ( $X$ ) yang terdistribusi normal didistribusikan secara kontinu (variabel kontinu) dari  $-\infty < X < +\infty$ , dan total luas di bawah kurva model adalah 1. (Carla Martins,2023)

### 2.2.5 Confusion Matrix

*Confusion matrix* merupakan tabel pencatat hasil kerja klasifikasi. *Confusion matrix* melakukan pengujian untuk memperkirakan objek yang benar dan salah. Tiap kolom pada matrix adalah contoh kelas prediksi, sedangkan tiap garis mewakili kejadian dikelas yang sebenarnya. *Confusion matrix* berisi informasi aktual(*actual*) dan prediksi (*predicted*) pada sistem klasifikasi

Tabel 2. 2 Confusion Matrix

|              |          | Predicted Class           |                            |
|--------------|----------|---------------------------|----------------------------|
|              |          | Positive                  | Negative                   |
| Actual Class | Positive | True positives count (TP) | False negatives count (FN) |
|              | Negative | False positive count (FP) | True negative count (TN)   |

*True Positif*(TP) adalah jumlah *record positif* yang diklasifikasikan sebagai *positif*, *False Positives*(FP) adalah jumlah *record positif* yang diklasifikasikan sebagai *negatif*, *False Negative*(FN) adalah jumlah *record negatif* yang diklasifikasikan sebagai *positif*, *True Negative*(TN) adalah jumlah *record negative* yang diklasifikasikan sebagai negatif. Data uji yang dimasukkan ke dalam *confusion matrix*, akan dihitung nilai-nilai *accuracy*, *precision* dan *recall*.

- a. Akurasi adalah metode yang didasari Tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. Akurasi adalah hasil dari penjumlahan nilai diagonal dibagi dengan jumlah total keseluruhan data dan selanjutnya dikalikan 100% (1)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \times 100\% \quad (1)$$

$$\text{akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}}$$

- b. Precision (p) = jumlah sampel berkategori positif diklasifikasi benar dibagi dengan total sampel yang diklasifikasi sebagai sample positif. (2)

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

- c. *Recall* (r) = jumlah sampel diklasifikasi positif dibagi total sampel dalam testing set berkategori positif. (3)

$$\mathbf{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

- d. *F1-score* adalah rata-rata harmonik dari presisi dan *recall*. *F1-score* juga bisa disebut *F1-measure* (4)

$$\mathbf{F1 - score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} + \times 100 \quad (4)$$

### 2.2.6 Jupyter Notebook

Jupyter notebook adalah software yang sangat populer beberapa tahun terakhir. Jupyter adalah organisasi non-profit untuk mengembangkan software interaktif dalam berbagai bahasa pemrograman. Notebook adalah suatu software buatan jupyter, adalah aplikasi web open-source yang memungkinkan kita untuk membuat berbagai dokumen interaktif yang berisi kode live, persamaan, visualisasi, dan teks naratif yang kaya.