

# **HEART DISEASE DETECTOR WITH MACHINE LEARNING USING NAIVE BAYES ALGORITHM**



## **Anggota Kelompok**

Irfan Rahmadan	(Ketua)	20.11.3792
Muhamad Gatot Supiadin		20.11.3771
Aditia Parmadi		20.11.3786
Galih Purnomo		20.11.3791
Muhammad Dzakwan Ar R		20.11.3797

**PRODI S1 INFORMATIKA  
UNIVERSITAS AMIKOM YOGYAKARTA  
TAHUN AJARAN 2021/2022**

## **Daftar Isi**

<b>Abstraksi</b>	<b>3</b>
<b>BAB I PENDAHULUAN</b>	<b>4</b>
• <b>Latar Belakang</b>	<b>4</b>
<b>Bab II.</b>	<b>5</b>
<b>Tinjauan Pustaka</b>	<b>5</b>
<b>1.1 Penelitian Terdahulu</b>	<b>5</b>
<b>Bab. III</b>	<b>7</b>
<b>Metodologi Research</b>	<b>7</b>
• <b>Flowchart/alur kerja</b>	<b>7</b>
• <b>Spesifikasi software/hardware/library</b>	<b>7</b>
<b>Bab IV</b>	<b>8</b>
<b>Hasil Pembahasan</b>	<b>8</b>
<b>Kesimpulan</b>	<b>10</b>

## **Abstraksi**

Karena institusi medis (rumah sakit, puskesmas) menghasilkan data yang sangat banyak, tetapi karena data ini tidak dimanfaatkan dengan baik. Kumpulan data tersebut mengandung banyak informasi tersembunyi. Dimungkinkan untuk mengubah data yang tidak berguna ini menjadi data yang bermanfaat. Kita dapat menggunakan berbagai pendekatan data mining untuk tujuan ini. Penelitian ini menjelaskan bagaimana Naive Bayes dapat digunakan untuk tujuan klasifikasi dan mengusulkan strategi pengklasifikasian untuk mengidentifikasi gangguan jantung. Data medis akan dibagi menjadi lima kategori dalam sistem kami: tidak, rendah, rata-rata, tinggi, dan sangat tinggi. Selain itu, jika sampel yang tidak diketahui diterima, sistem akan memperkirakan label kelasnya. Dengan demikian, kategorisasi (pelatihan) dan prediksi (pengujian) adalah dua tugas mendasar yang akan dilakukan. Algoritma dan database yang digunakan menentukan seberapa akurat sistem tersebut.

Hasilnya, pasien dapat dengan mudah mencegah atau bahkan menjaga kesehatan jantung mereka dengan melakukan hal-hal kecil seperti memperbaiki kebiasaan tidur mereka, menghindari begadang secara teratur, sering berolahraga, memperbaiki pola makan, jogging setiap pagi, dan lain-lain untuk memastikan bahwa jantung mereka berfungsi dengan baik.

## **Abstraction**

Because medical institutions (hospitals, health centers) produce a vast amount of data, but because this data is not well utilized. The data set contains a great deal of hidden information. It is possible to transform this useless data into helpful data. We can employ a variety of data mining approaches for this aim. This study explains how Naive Bayes can be used for classification purposes and proposes a classifier strategy to identify cardiac disorders. The medical data will be divided into five categories in our system: no, low, average, high, and very high. Additionally, if an unknown sample is received, the system will forecast its class label. Thus, categorization (training) and prediction (testing) are the two fundamental tasks that will be carried out. The algorithm and database that are employed determine how accurate the system is.

As a result, patients may easily prevent or even maintain their heart health by doing minor things like improving their sleep habits, avoiding staying up late regularly, exercising frequently, improving their diet, jogging every morning, and others to ensure that their heart is functioning at its best.

# **BAB I**

## **PENDAHULUAN**

- **Latar Belakang**

Manusia memiliki organ yang disebut jantung yang memompa darah. Kesehatan jantung, bersama dengan kesehatan fisik, adalah salah satu faktor paling penting yang harus diperhitungkan bagi setiap orang saat ini. Mayoritas orang secara konsisten mengabaikan kesehatan jantung karena tidak terlihat dan rumit untuk dideteksi, namun penyakit jantung yang tidak diobati meningkatkan kemungkinan kematian.

Mengingat masalah mereka yang secara konsisten mengabaikan kesehatan secara keseluruhan, kami sebagai mahasiswa berusaha menciptakan solusi. Solusi ini terdiri dari Detektor Penyakit Jantung / Detektor Penyakit Jantung yang dibuat dengan Kecerdasan Buatan.

Dengan menggunakan Kecerdasan Buatan dapat diketahui bahkan dideteksi bagaimana kondisi jantung seseorang dimana akan ditentukan oleh kebiasaan orang tersebut dalam menjaga pola makan, olahraga, pola tidur, dan lainnya. Naive Bayes kami pilih sebagai algoritma yang cocok menangani kasus ini dimana pada algoritma ini mengadaptasi proses clustering pada data.

Detektor Penyakit Kardiovaskular ini berfungsi dengan menggunakan kumpulan data sebagai dasar untuk mengumpulkan informasi tentang gejala penyakit jantung yang dialami banyak orang. Karena sebagian besar dari individu-individu ini tidak menyadari manifestasi klinis dan bahkan mungkin gejala penyakit kardiovaskular yang parah, tujuan alat ini adalah untuk menurunkan penyakit kadar kolesterol.

## **Bab II.**

### **Tinjauan Pustaka**

#### **1.1 Penelitian Terdahulu**

Penelitian sebelumnya berupaya mengumpulkan informasi tentang deteksi penyakit jantung. Masalah yang diangkat dalam penelitian ini berkaitan dengan bahaya penyakit kardiovaskuler, yang mempengaruhi semakin banyak orang. Dataset tentang penyakit jantung berfungsi sebagai sumber penelitian kami. Untuk mengenali risiko penyakit jantung koroner lebih awal dan mengurangi angka kematian terkait penyakit jantung, kami akan mengubah dataset tersebut ke dalam pembelajaran mesin.

Selain itu, penelitian ini berharap dapat mempermudah para profesional medis untuk mengenali penyakit jantung.

Teknik Naive Bayes akan digunakan dalam penelitian kami. Dengan bantuan bahasa pemrograman Python, algoritma ini akan membagi data ke dalam kategori yang berbeda.

#### **1.2 Dasar Teori**

Ferry Christian, Sp.JP mengklaim bahwa pemantauan jantung adalah proses untuk menemukan masalah jantung. Pemeriksaan jantung dapat menentukan risiko seseorang terkena penyakit jantung sebelum timbulnya gejala selain untuk mengidentifikasi penyakit jantung. Dengan kata lain, pemeriksaan ini juga membantu dalam mencegah penyakit jantung. Sejak menginjak usia 20 tahun, seseorang harus mulai memantau kesehatan jantung mereka sebagai organ utama. Anda dapat melakukan pemeriksaan kardio sekitar setahun sekali saat Anda berusia 30 tahun atau lebih.

Mengutip dari Jurnal “ N. Aaditya Sunder, P. PushpaLatha, “Performance analysis of classification data mining techniques over heart disease database” International Journal Of Engineering Science and Advance Technology”-vol-2 issue-3,470-478,May-June 2012..” Ini menawarkan cara baru untuk menyelidiki dan memahami data. Dengan menentukan kesamaan antara variabel target (juga dikenal sebagai variabel dependen) dan variabel lain (juga dikenal sebagai variabel independen), Naive Bayes belajar dari "bukti" yang dikumpulkannya. Weighted Associative Classifier (WAC) memang telah dianggap sebagai metode baru untuk menghasilkan aturan yang signifikan.

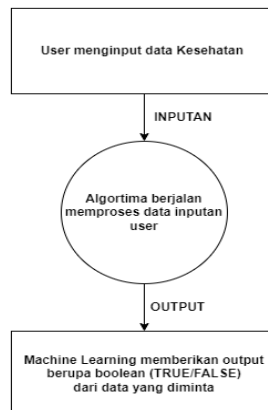
Berikut referensi dan sitasi yang dikutip :

- International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624. Vol 3, Issue 3, 2012, pp 290-294 <http://bipublication.com>
- Mai Shouman, Tim Turner, Rob Stocker, "Using data mining techniques in heart disease diagnosis and treatment", JapanEgypt Conference on Electronics, Communications and Computers 978-1-4673-0483-2 c\_2012 IEE
- N. Aaditya Sunder, P. PushpaLatha, "Performance analysis of classification data mining techniques over heart disease database" International Journal Of Engineering Science and Advance Technology"-vol-2 issue-3,470-478,May-June 2012.
- Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006
- IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August, 2008.
- R. Bhuvaneswari and K. Kalaiselvi, Naive Bayesian Classification Approach in Healthcare Applications International Journal of Computer Science and Telecommunications, [Volume 3, Issue 1, January 2012].
- Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni," Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011

## Bab. III

### Metodologi Research

- **Flowchart/alur kerja**



Alur kerja pada kasus pendeteksi penyakit jantung ini dimulai dari urutan berikut

- User/pengguna menginput data seperti BMI, Mental Health, etc
- Algoritma memproses data dari user dengan clustering dari Naive Bayes memproses data
- Sistem menentukan serta memberikan output hasil sehingga bisa dilihat oleh user dimana output tersebut berupa boolean (TRUE/FALSE)

- **Spesifikasi software/hardware/library**

Pada kasus berikut ini merupakan library-library yang dipakai yang digunakan oleh machine learning

1. Pandas
2. Matplotlib
3. Numpy
4. Sklearn

Pada kasus ini berikut merupakan spesifikasi hardware yang dibutuhkan meliputi minimum requirement

1. Ram 4gb
2. SSD 256gb
3. Processor Intel Core i3-5005U

Untuk Software pada kasus ini menggunakan 2 software yang bisa dipilih salah satu diantaranya

1. Anaconda/Jupyter Notebook
2. Google Collab

## Bab IV

### Hasil Pembahasan

Algoritma Naive Bayes adalah kumpulan metode pembelajaran yang diawasi berdasarkan asumsi "naif" bahwa setiap pasangan fitur secara linear independen jumlah pengklasifikasi dasar. Diberikan variabel kelas, teorema Bayes menggambarkan hubungan berikut antara  $y$  dan vektor fitur dependen  $x_1$  sampai  $x_n$ :

$$P(y | x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n | y)}{P(x_1, \dots, x_n)}$$

Menggunakan asumsi independensi bersyarat naif bahwa

$$P(x_i | y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | y),$$

untuk semua  $i$ , hubungan ini disederhanakan menjadi

$$P(y | x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x_1, \dots, x_n)}$$

Sejak  $P(x_1, \dots, x_n)$  konstan diberikan input, kita dapat menggunakan aturan klasifikasi berikut:

$$\begin{aligned} P(y | x_1, \dots, x_n) &\propto P(y) \prod_{i=1}^n P(x_i | y) \\ &\Downarrow \\ \hat{y} &= \arg \max_y P(y) \prod_{i=1}^n P(x_i | y), \end{aligned}$$

dan kita dapat menggunakan estimasi Maximum A Posteriori (MAP) untuk memperkirakan  $P(y)$  dan  $P(x_i | y)$  : yang pertama adalah frekuensi relatif kelas  $y$  dalam set pelatihan.

Pengklasifikasi naif Bayes yang berbeda berbeda terutama oleh asumsi yang mereka buat mengenai distribusi  $P(x_i | y)$



Metode Klasifikasi telah bekerja dengan cukup baik dalam konteks dunia nyata tertentu, terutama kategorisasi dokumen dan penyaringan spam, meskipun asumsi mereka seolah-olah sederhana. Bagi mereka untuk mengestimasi parameter yang dibutuhkan, hanya dibutuhkan data training dalam jumlah yang sedikit. (Lihat sumber-sumber di bawah ini untuk penjelasan teoritis mengapa naive Bayes efektif dan untuk jenis data apa).

Jika dibandingkan dengan teknik-teknik yang lebih kompleks, pelatih dan klasifikasi naive Bayes bisa sangat cepat. Setiap distribusi dapat diestimasi secara individual menjadi satu distribusi karena pemisahan distribusi fitur perkuliahan. Sebagai hasilnya, masalah yang ditimbulkan oleh kompleksitas komputasi berkurang.

```
## Example people
BMI = 24.21
PhysicalHealth = 3.0
MentalHealth = 1.0
SleepTime = 1.0

##testing
Data_Testing = [[BMI,PhysicalHealth,MentalHealth,SleepTime]]
```

```
if y_pred == 'No':
    hasil = "Tidak terkena"
elif y_pred == 'Yes':
    hasil = "Akan terkena"
else:
    hasil = "error"

print("Resiko Terkena Penyakit Jantung =",hasil)
```

Resiko Terkena Penyakit Jantung = Akan terkena

```
from sklearn.metrics import accuracy_score
print("nilai akurasi = %0.2f " % accuracy_score(y_test,y_predict))
```

nilai akurasi = 0.87

## **Bab V**

### **Kesimpulan**

Naive Bayes classifier adalah salah satu metodologi yang digunakan di mana-mana dan berpengetahuan luas untuk dimasukkan untuk melakukan prediksi yang berbeda, bersama dengan probabilitas. Lebih jauh lagi, sistem otomatis ini tampaknya hanya membutuhkan sumber daya komputasi yang cukup rendah karena desainnya yang sederhana, maka Naive Bayes sangat cocok untuk menangani pada kasus deteksi penyakit jantung ini, kasus penyakit jantung yang diangkat yang pastinya ini membutuhkan mengandung banyak sekali data data dari pasien maupun non pasien, data yang diambil dari dataset yang nantinya akan diolah melalui machine learning yang menggunakan algoritma Naive Bayes dimana prediksi dan probably sangat dibutuhkan dalam kasus ini, sementara itu algoritma ini cocok untuk menangani kasus detektor penyakit jantung.