

# Penerapan Data Mining untuk Klasifikasi Penyakit Demam Berdarah Dengue (DBD) Dengan Metode Naïve Bayes (Studi Kasus Puskesmas Taman Krocok)

Moh. Jasri<sup>a</sup>, Andi Wijaya<sup>b</sup>, Mochammad Rizqi Sunggara<sup>c</sup>

<sup>a</sup>Sistem Informasi Universitas Nurul Jadid, Probolinggo Jawa Timur Indonesia

<sup>b</sup>Rekayasa Perangkat Lunak Universitas Nurul Jadid, Probolinggo Jawa Timur Indonesia

<sup>c</sup>Teknik Informatika Universitas Nurul Jadid, Probolinggo Jawa Timur Indonesia

\*correspondence email : [jasri@unuja.ac.id](mailto:jasri@unuja.ac.id)

**Abstract**—Dengue Hemorrhagic Fever (DHF) is an infectious disease caused by the bite of *Aedes aegypti* and *Aedes albopictus* mosquitoes. This *Aedes* mosquito is one of the most widespread mosquitoes in Indonesia, except for areas at an altitude exceeding one hundred meters above sea level. The Puskesmas Taman Krocok locate in the east of Bondwoso city. Many patients are affected by dengue fever every year due to a lack of counselling from the health department. The location under review is almost at the foot of the mountain, so the potential for dengue fever is high-speed. The classification process uses Naïve Bayes, which acts as a feature selection on the Taman Krocok Health Center data. After a series of the processes, the accuracy rate for the Naïve Bayes classification model is 90%. With an *f1*-score of 89% for negative numbers and a positive *f1*-score of 92%.

**Index Terms**— Data Mining, Dengue Hemorrhagic Fever (DHF), Naïve Bayes, Puskesmas Taman Krocok.

**Abstrak**—Demam Berdarah Dengue (DBD) adalah sebuah Penyakit yang menular akibat gigitan nyamuk *Aedes aegypti* serta *Aedes albopictu*. Nyamuk *Aedes* ini adalah salah satu nyamuk yang tersebar luas di wilayah Indonesia, terkecuali wilayah yang berada pada ketinggian yang melebihi seratus meter dari permukaan dasar laut. Puskemas Taman Krocok terletak di sebelah timur kota Bondwoso, dimana disetiap tahunnya pada saat musim hujan berlangsung terdapat banyak pasien yang terkena penakit demam berdarah yang dikarenakan kurangnya penyuluhan dari dinas kesehatan dimana lokasi yang tinjau hampir berada di kaki gunung, sehingga potensi terkenanya penyakit DBD sangat cepat. Proses klasifikasi menggunakan Naïve Bayes yang berperan sebagai seleksi fitur pada data yang dipeoleh dari Puskesmas Taman Krocok. Setelah serangkaian proses diatas dilakukan maka didapat tingkat akurasi untuk model klasifikasi Naïve Bayes didapat 90%. Dengan *f1*-score 89% dari angka negatif, *f1*-Score positif di dapat 92%.

**Kata Kunci**—Data Ming, Demam Berdarah Dengue (DBD), Naïve Bayes, Puskesmas Taman Krocok.

## I. PENDAHULUAN

Demam Berdarah Dengue (DBD) adala sebuah Penyakit yang menular akibat gigitan nyamuk *Aedes aegypti* serta *Aedes albopictus* [1]. Nyamuk *Aedes* ini adalah salah satu nyamuk yang tersebar luas di wilayah Indonesia, terkecuali wilayah yang berada pada ketinggian yang melebihi seratus meter dari permukaan dasar laut. Beberapa peneliti sempat berpendapat bahwa nyamuk *Aedes* di suhu 28-32 derajat celcius maka disitulah Nyamuk *Aedes* cepat berkembang biak sehingga kehidupannya lebih panjang. Factor yang bisa mempengaruhi perkembangan biakan Nyamuk *Aedes* yaitu suhu yang selalu berubah-ubah, curah hujan yang tinggi, kondisi lingkungan tidak sehat, mobilitas penduduk serta perilaku masyarakat yang kurang sehat.

Indonesia termasuk negara dengan tingkat tertinggi pada kasus demam berdarah di setiap tahunnya pada saat musim hujan berlangsung [2]. WHO (*Word Health Organization*) menyatakan kasus penyakit demam berdarah di Indonesia merupakan yang tertinggi di Asia Tenggara. Pada bulan Februari 2022 Kementerian Kesehatan RI menyatakan terdapat 45.387 kasus DBD dari 34 provinsi yang ada di Indonesia untuk angka kematian yang terdiagnosa penyakit DBD sebanyak 432 orang [3].

Pentingnya menjaga kesehatan lingkungan yaitu untuk menurunkan peningkatan penularan penyakit demam berdarah. penularan penyakin demam berdarah yaitu dengan gigitan nyamuk *Aedes* kemudian di

tularkan pada manusia lain. Penyebaran penyakit demam berdarah disebabkan oleh nyamuk *Aedes* saat setelah manusia mengidap virus viremia di dalam darah. Viremia sendiri adalah suatu masa virus dengue berada dalam aliran darah manusia sehingga virus dengue dapat berkembang biak di dalam tubuh manusia. Pengobatan penyakit demam berdarah yaitu dengan dibutuhkan diagnosa awal yang tepat disertai dengan pengobatan intensif. Demam Berdarah adalah penyakit yang perlu pengawasan dengan baik karena merupakan penyakit yang dapat menyebabkan kematian jika penanganannya kurang baik [4]. Oleh sebab itu perlunya dilakukan klasifikasi data mining secara baik agar dapat mempermudah diagnosa penyakit DBD.

Puskemas Taman Krocok terletak di sebelah timur kota Bondwoso, dimana disetiap tahunnya pada saat musim hujan berlangsung terdapat banyak pasien demam berdarah. Karena kurangnya penyuluhan dari dinas kesehatan dan lokasi yang berada di kaki gunung, sehingga berpotensi terkenanya penyakit DBD sangat cepat. Disamping itu hidup sehat masih jauh dari kata sempurna sehingga dari hasil observasi terdapat banyak pasien yang menderita DBD. Diharapkan Puskesmas Taman Krocok bisa meminimalisir penyakit DBD dengan mengklasifikasi data pasien yang terkena penyakit DBD sehingga dapat mengurangi pasien di setiap tahunnya [5]. Mempertimbangkan penelitian terdahulu maka pada kali dilakukan proses klasifikasi menggunakan Naïve Bayes yang berperan sebagai seleksi fitur pada data yang diperoleh dari Puskesmas Taman Krocok [6].

## II. METODE PENELITIAN

Pada penelitian ini, tahapan penelitian dilakukan melalui beberapa kegiatan yaitu pertama adalah pengumpulan data yang meliputi studi literatur yang dilakukan untuk mempelajari metode yang akan digunakan lalu dilanjutkan tahapan selanjutnya adalah pengumpulan data yang akan diolah dari proses wawancara dan observasi di lapangan lalu dilanjutkan tahap selanjutnya adalah preprocessing data untuk memilah data yang akan diolah sesuai variabel yang digunakan. Kemudian menghitung jumlah nilai rata-rata dari keseluruhan atribut, lalu menerapkan aturan asosiasi baik pada atribut status maupun atribut yang lainnya untuk mempermudah dalam proses perhitungannya yang dilanjutkan ketahap selanjutnya yaitu tahap analisis dengan menggunakan metode yang akan digunakan yaitu metode Naïve Bayes.

### Pengumpulan data

Pengumpulan data disini menggunakan observasi dan wawancara. Observasi dilakukan untuk menganalisa objek di lapangan secara langsung. Sedangkan, wawancara terhadap kepala Puskesmas Taman Krocok dimana pada hal ini akan tertuju langsung kepada dokter yang bersangkutan dalam penanganan pasien yang terindeksi penyakit DBD.

### Preprocessing Data

Preprocessing data adalah langkah pertama dalam data mining sebagai proses mengubah *raw data* (bentuk mentah) ke dalam data yang lebih mudah dipahami. Karena kebanyakan data mentah tidak memiliki format yang teratur dan tidak mudah dipahami. Tahap ini penting karena sebagian besar proses data mining terletak pada tahap preprocessing ini dan cukup menghabiskan waktu. Preprocessing ini disebut juga preparasi data atau proses sterilisasi data agar pada saat proses data dilakukan memiliki keakuratan pola data.

### Klasifikasi

Klasifikasi adalah salah satu teknik machine learning dan termasuk dalam tipe supervised learning. Yang bertujuan untuk memprediksi suatu kategori atau data. Langkah yang harus dilakukan melakukan proses pembelajaran. Proses pembelajaran disini memerlukan sebuah data yang mana data tersebut digunakan untuk data latih/training. Untuk data yang digunakan pada proses prediksi disebut dengan data uji/testing [7].

## III. HASIL DAN PEMBAHASAN

Dari hasil pengumpulan dataset yang dilakukan di puskesmas Taman Krocok didapatkan data sebanyak 215 pasien dari tahun 2014 sampai 2022 yang didapatkan di puskesmas Taman Krocok. Data yang didapat meliputi data pasien tiap harinya serta pengecekan apakah diagnosa penyakit DBD atau DBD status rendah.

Tabel 1. Data set tabel pasien Resiko DBD dari Tahun 2014 sampai 2022

No	Nama	Usia	Jenis Kelamin	Suhu Tubuh	Bintik Merah	Trombosit	Diagnosa
1	Suparman	35	Laki-laki	39	0	Resiko rendah	Negatif
2	Anggita	5	Perempuan	39	1	Resiko sedang	Positif
3	Silvia	25	Perempuan	40	4	Resiko tinggi	Positif
4	Ulfatun nazihah	21	Perempuan	38	3	Resiko tinggi	Positif
5	Cindy aulia putri	10	Perempuan	40	2	Resiko tinggi	Positif
6	Belqis aprilianti	8	Perempuan	38	3	Resiko tinggi	Positif
7	Zainal arifin	15	Laki-laki	39	1	Resiko sedang	Positif
8	Muhammad ali	35	Laki-laki	39	0	Resiko rendah	Negatif
9	Ahmad effendi	30	Laki-laki	39	0	Resiko rendah	Negatif
10	Indah nur umami	23	Perempuan	39	0	Resiko rendah	Negatif
11	Rafikatul hasanah	21	Perempuan	37	0	Resiko rendah	Negatif
12	Diana putri	22	Perempuan	41	3	Resiko tinggi	Positif
13	Muhammad anshari	18	Laki-laki	39	1	Resiko sedang	Positif
14	Supriyadi	22	Laki-laki	39	2	Resiko tinggi	Positif
15	Muhammad ghafur	23	Laki-laki	39	2	Resiko tinggi	Positif
16	Samsul arifin	25	Laki-laki	38	1	Resiko sedang	Positif
17	Novita sari	11	Perempuan	39	0	Resiko rendah	Negatif
18	Maghfirah	19	Perempuan	39	0	Resiko rendah	Negatif
19	Nikmatul husna	5	Perempuan	40	4	Resiko tinggi	Positif
20	Rahmatullah	26	Laki-laki	39	1	Resiko sedang	Positif
21	Rafi'atul maghfiroh	10	Perempuan	39	0	Resiko rendah	Negatif
22	Yazid	11	Laki-laki	37	2	Resiko tinggi	Positif
23	Muhammad ilham	22	Laki-laki	39	1	Resiko sedang	Positif
24	Ainun shofiyah	13	Perempuan	39	0	Resiko rendah	Negatif
215	Ahmad naufal	16	Laki-laki	39	0	Resiko rendah	Negatif

A. Preprocessing Data

Preprocessing data adalah langkah pertama dalam data mining untuk mengubah raw data (bentuk mentah) ke dalam data yang lebih mudah dipahami. Beberapa langkah yang dilakukan pada saat preprocessing data yaitu tokenisasi yaitu proses pemisahan suatu rangkaian karakter berdasarkan pada karakter spasi yang mungkin saja terjadi penghapusan karakter tertentu, lalu Case Folding yaitu proses mengubah semua huruf dalam dokumen menjadi huruf kecil untuk mempermudah pencarian, dan dilanjut Stemming yang dilakukan untuk menyeragamkan kata hingga mengurangi daftar kata yang ada pada data latih, lalu Filtering/eliminasi stopword yang bergungsi untuk mengurangi space pada tabel term index hingga 40% atau lebih, dan Labelling yang digunakan untuk memberi tanda khusus agar lebih kuat.

Tabel 2. Aturan Asosiasi Usia

Usia	Kategori
1 – 15 Tahun	Anak – Anak
15 – 25 Tahun	Remaja
25 – 40 tahun	Dewasa
40 – 60 Tahun	Tua
60 – 80 Tahun	Lansia

Tabel 3. Aturan Asosiasi Suhu Tubuh

Suhu Tubuh	Kategori
37 – 39	Rendah
40 -42	Tinggi

B. Implementasi Naïve Bayes

Setelah tahap preprocessing data di atas dilakukan selanjutnya data yang sudah siap di implementasikan untuk di analisis menggunakan metode naïve bayes. Mulai tahap mengintegrasikan sejumlah data yang didapat dari suatu lingkungan, menetapkan sasaran yang didapat secara spesifik, dan memilih langkah alternatif untuk mengatasi masalah dan menetapkan langkah yang tepat diantara yang terbaik untuk persiapan yang tepat guna sesuai kebutuhan. Sedangkan tujuan analisis sendiri adalah mengenali sejumlah data yang didapat dari suatu populasi, agar dapat menarik kesimpulan.

Naive Bayes mempunyai algoritma posteriori sebagai berikut :

$$p(H|X) = \frac{P(X|H). P(H)}{P(X)}$$

Keterangan :

- $P(X|H)$  adalah probabilitas untuk mengamati sampel.
- $P(H)$  adalah probabilitas awal.
- $P(X)$  adalah probabilitas terhadap data sampel yang diamati.

Langkah pertama adalah membagi menjadi 2 kelas yaitu "POSITIF" dan "NEGATIF".

$P(C_i)$  :

$$P(\text{Status} = \text{"Positif"}) = 12/20 = 0,6$$

$$P(\text{Status} = \text{"Negatif"}) = 8/20 = 0,4$$

Hitung  $P(X|C_i)$  untuk tiap kelas :

$$P(\text{Nilai Usia} = \text{"Anak - Anak"} | \text{Status} = \text{"positif"}) = 3/12 = 0,25$$

$$P(\text{Nilai Usia} = \text{"Anak - Anak"} | \text{Status} = \text{"Negatif"}) = 4/8 = 0,5$$

$$P(\text{Nilai Usia} = \text{"Remaja"} | \text{Status} = \text{"Positif"}) = 2/12 = 0,16$$

$$P(\text{Nilai Usia} = \text{"Remaja"} | \text{Status} = \text{"Negatif"}) = 2/8 = 0,25$$

$$P(\text{Nilai Usia} = \text{"Dewasa"} | \text{Status} = \text{"positif"}) = 3/12 = 0,25$$

$$P(\text{Nilai Usia} = \text{"Dewasa"} | \text{Status} = \text{"Negatif"}) = 0/8 = 0$$

$$P(\text{Nilai Usia} = \text{"Tua"} | \text{Status} = \text{"positif"}) = 4/12 = 0,3$$

$$P(\text{Nilai Usia} = \text{"Tua"} | \text{Status} = \text{"Negatif"}) = 2/8 = 0,25$$

$$P(\text{Jenis Kelamin} = \text{"Perempuan"} | \text{Status} = \text{"Positif"}) = 7/12 = 0,58$$

$$P(\text{Jenis Kelamin} = \text{"Perempuan"} | \text{Status} = \text{"Negatif"}) = 5/8 = 0,625$$

$$P(\text{Jenis Kelamin} = \text{"Laki - Laki"} | \text{Status} = \text{"Positif"}) = 5/12 = 0,416$$

$$P(\text{Jenis Kelamin} = \text{"Laki - Laki"} | \text{Status} = \text{"Negatif"}) = 3/8 = 0,375$$

$$P(\text{Suhu Tubuh} = \text{"Rendah"} | \text{Status} = \text{"positif"}) = 9/12 = 0,75$$

$$P(\text{Suhu Tubuh} = \text{"Rendah"} | \text{Status} = \text{"Negatif"}) = 8/8 = 1$$

$$P(\text{Suhu Tubuh} = \text{"Tinggi"} | \text{Status} = \text{"positif"}) = 3/12 = 0,25$$

$$P(\text{Suhu Tubuh} = \text{"Tinggi"} | \text{Status} = \text{"Negatif"}) = 0/8 = 0$$

$$P(\text{Trombosit} = \text{"Resiko Rendah"} | \text{Status} = \text{"positif"}) = 0/12 = 0$$

$$P(\text{Trombosit} = \text{"Resiko Rendah"} | \text{Status} = \text{"Negatif"}) = 7/8 = 0,875$$

$$P(\text{Trombosit} = \text{"Resiko Sedang"} | \text{Status} = \text{"positif"}) = 9/12 = 0,75$$

$$P(\text{Trombosit} = \text{"Resiko Sedang"} | \text{Status} = \text{"Negatif"}) = 0/8 = 0$$

$$P(\text{Trombosit} = \text{"Resiko Tinggi"} | \text{Status} = \text{"positif"}) = 3/12 = 0,25$$

$$P(\text{Trombosit} = \text{"Resiko Tinggi"} | \text{Status} = \text{"Negatif"}) = 1/8 = 0,125$$

$P(X|C_i)$

$$P(X|\text{Status} = \text{"Positif"}) = 0,25 \times 0,16 \times 0,25 \times 0,3 \times 0,58 \times 0,416 \times 0,75 \times 0,25 \times 0,75 \times 0,25 = 0,0000254475$$

$$P(X|\text{Status} = \text{"Negatif"}) = 0,5 \times 0,25 \times 0,25 \times 0,625 \times 0,375 \times 1 \times 0,875 \times 0,125 = 0,00080108642578125$$

$P(X|C_i) \times P(C_i)$  :

$$P(X|\text{Status} = \text{"Positif"}) \times P(\text{Status} = \text{"Positif"}) = 0,0000254475 \times 0,6 = 0,0000152685$$

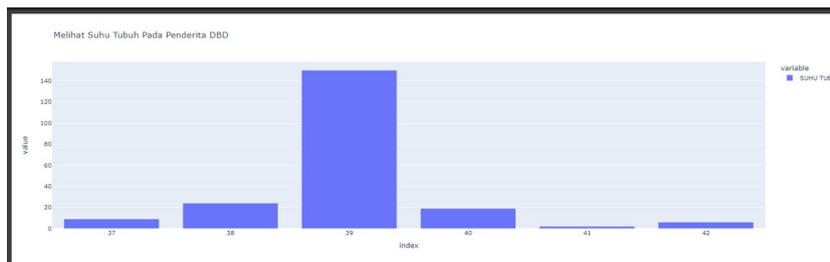
$$P(X|\text{Status} = \text{"Negatif"}) \times P(\text{Status} = \text{"Negatif"}) = 0,00080108642578125 \times 0,4 = 0,0003204345703125$$

Oleh karena itu, X milik kelas ("Status = Negatif")

Dengan dataset sebanyak 210 penderita Demam Berdarah dengan 4 atribut disetiap data. Setelah menampilkan 5 data dari atas maka selanjutnya menampilkan 5 data terakhir supaya data dari atas sampai akhir tidak ada yang terkurangi. Visualisasi data usia, suhu tubuh, dan trombosit dapat dilihat pada gambar 1, 2, dan 3.



Gambar 1 Hasil Visualisasi Data Usia



Gambar 2 Hasil Visualisasi Data Suhu Tubuh



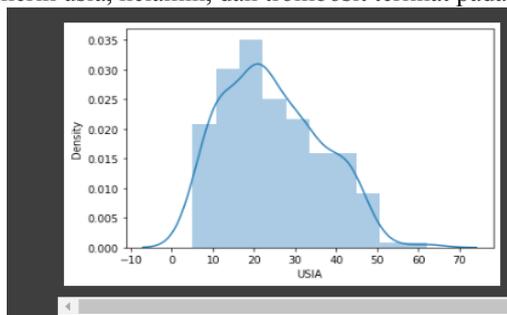
Gambar 3 Hasil Visualisasi Data Trombosit

Setelah mengetahui atribut dari setiap kolom, untuk setiap atribut tidak semua akan dieksekusi dalam perhitungan naïve bayes sehingga perlunya pemangkasan kolom yang tidak diperlukan dalam proses naïve bayes nantinya. Data yang sebelumnya berbentuk objek atau lebih di kenal dengan huruf harus diubah dalam bentuk numeric sehingga pada proses selanjutnya yaitu cleaning data dimana bertujuan untuk mengubah data objek menjadi numeric.

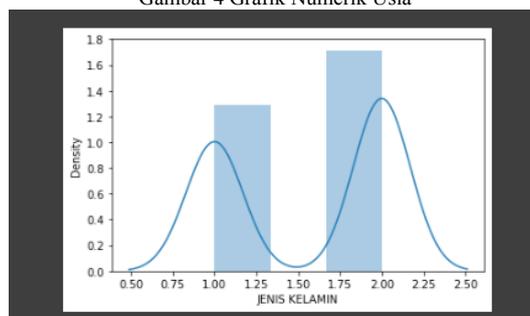
Tabel 4. Proses Cleaning data

Nama Atribut	Jenis Atribut	Numerik
Jenis Kelamin	Laki -Laki	1
	Perempuan	2
	Resiko Rendah	0
Trombosit	Resiko Sedang	1
	Resiko Tinggi	2
	Negatif	0
Diagnosa	Positif	1

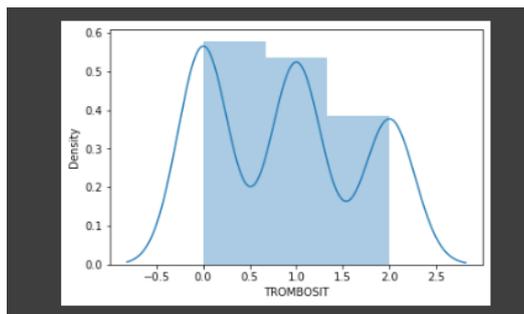
Setelah tahapan *cleaning* data selesai maka data yang sebelumnya berbentuk objek diubah menjadi numeric. Setelah proses cleaning data atau perubahan string ke integer berikutnya dapat ditampilkan data grafik dengan format numeric dengan density. Density sendiri adalah ukuran kepadatan pada setiap atribut tingkat tertinggi dalam penyakit DBD yaitu terletak pada usia 18 sampai 20 tahun dengan density 0.035. Visualisasi atribut numeric usia, kelamin, dan trombosit terlihat pada gambar 4, 5, dan 6.



Gambar 4 Grafik Numerik Usia



Gambar 5 Grafik Jenis Kelamin



Gambar 6 grafik numeric Trombosit

Setelah proses pemanggilan dataset kembali yang asli akan dibandingkan atribut mana saja yang sudah di ganti numeric, tujuannya adalah mencocokkan kembali data yang sudah kita cleaning. Setelah semua proses diatas sudah selesai, dilanjut ke tahapan perhitungan naïve bayes dengan memanggil fungsi naïve bayes.

```

      USIA  JENIS  KELAMIN  SUHU  TUBUH  TROMBOSIT
0         35         1         39         0
1         5         2         39         1
2         25         2         40         2
3         21         2         38         2
4         10         2         40         2
..      ...      ...      ...      ...
205        45         2         39         1
206        41         2         39         2
207        18         2         39         2
208         20         1         38         1
209         43         1         39         0
[210 rows x 4 columns]
    
```

Gambar 7 Data Variable Endependen Variable X

Setelah proses pembuatan variable endependen dengan variable x diatas dan dilanjutkan dengan variable independen variable y dari tingkat diagnose apakah positif atau negative.

```

0      0
1      1
2      1
3      1
4      1
..
205    1
206    1
207    1
208    1
209    0
Name: DIAGNOSA, Length: 210, dtype: object
    
```

Gambar 7 Data Variable Diagnose Apakah Positif Atau Negative

Pembagian data *testing* dan data *training* dengan perbandingan 1 : 4 atau 20% dan 80%. Selanjutnya mentraining data X sebanyak 20 data secara acak. Dan mentraining data diagnose sebagai atribut Y. Didapatkan akurasi diantara 0 = negative dan 1= positif . sebanyak 20 data yang di training. Setelah proses training selesai antara x dan y selanjutnya teting data antara x dan y sebanyak 20% dari 210 data yang di siapkan. Untuk tingkat keakurasian x testing yaitu 42, 4. Setelah data latih dan data uji didapat berikutnya mengubah data ke dalam variansi angka 0 dan 1 agar untuk meminimalisir atau mengurangi jarak atau interval data yang terlalu jauh.

```
#hasil data training x yang sudah ditransformasi
print(x_train)

[[[-0.96763785 -1.11355287 -0.02122222 -1.10068362]
 [ 0.93875314  0.89802651 -2.3981114  0.14874103]
 [ 1.37202381 -1.11355287 -0.02122222  1.39816568]
 [-0.44771303  0.89802651 -0.02122222 -1.10068362]
 [-1.05429198  0.89802651 -0.02122222 -1.10068362]
 [ 1.54533208 -1.11355287 -1.20966681 -1.10068362]
 [-1.05429198 -1.11355287  1.16722236  1.39816568]
 [-0.3610589  0.89802651 -0.02122222 -1.10068362]
 [-0.10109649  0.89802651 -0.02122222 -1.10068362]
 [ 0.15886592 -1.11355287  3.54411154  1.39816568]
 [-1.05429198  0.89802651 -0.02122222 -1.10068362]
 [-1.22760025  0.89802651 -0.02122222  0.14874103]
 [ 0.50548246 -1.11355287 -1.20966681 -1.10068362]
 [ 1.02540727  0.89802651 -1.20966681  0.14874103]
 [ 1.89194863  0.89802651 -0.02122222 -1.10068362]
 [-0.70767544 -1.11355287 -0.02122222 -1.10068362]
 [-0.70767544 -1.11355287 -0.02122222  1.39816568]
 [-1.5742168 -1.11355287  1.16722236  1.39816568]
 [-1.05429198  0.89802651 -0.02122222 -1.10068362]
 [-0.96763785  0.89802651 -0.02122222 -1.10068362]]
```

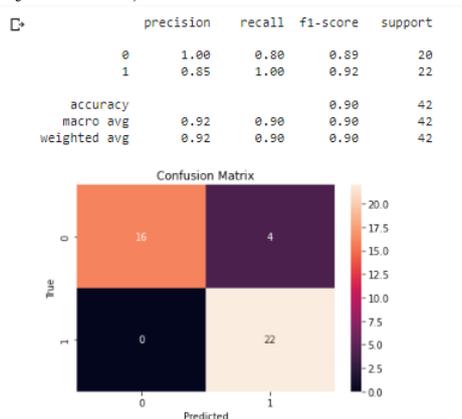
Gambar 8 Output Data X Training

```
#hasil data testing x yang sudah ditransformasi
print(x_test)

[[[ 0.67879073  0.89802651 -0.02122222  0.14874103]
 [ 1.37202381 -1.11355287 -0.02122222  1.39816568]
 [ 1.54533208 -1.11355287 -0.02122222 -1.10068362]
 [-0.44771303  0.89802651 -1.20966681  0.14874103]
 [ 0.93875314 -1.11355287 -0.02122222 -1.10068362]
 [ 0.93875314 -1.11355287 -1.20966681  1.39816568]
 [ 1.89194863  0.89802651 -0.02122222  0.14874103]
 [ 1.45867795  0.89802651  1.16722236  1.39816568]
 [-0.01444236 -1.11355287 -0.02122222  0.14874103]
 [-1.40090853  0.89802651 -0.02122222  0.14874103]
 [-1.05429198  0.89802651 -0.02122222 -1.10068362]
 [ 1.37202381  0.89802651 -0.02122222  0.14874103]
 [ 1.89194863 -1.11355287  1.16722236  1.39816568]
 [-1.14094612  0.89802651 -0.02122222 -1.10068362]
 [ 0.15886592 -1.11355287 -0.02122222  0.14874103]
 [-0.53436717 -1.11355287 -0.02122222  0.14874103]
 [-1.22760025  0.89802651 -0.02122222 -1.10068362]
 [-1.48756266  0.89802651 -0.02122222  0.14874103]]]
```

Gambar 9 Output Data X Testing

Selanjutnya memprediksi hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang sebenarnya untuk menggambarkan kinerja dari model klasifikasi yang digunakan menggunakan *confusion matrix*. Menampilkan dan membandingkan nilai aktual (sesungguhnya) dengan nilai hasil prediksi model yang dapat menghasilkan matrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *f-1 score (f-Measure)*.



Gambar 10 Hasil *Confusion Matrix*

Dari hasil *confusion matrix* diatas maka dapat dilihat hasil *data testing* atau data uji 20% yang kita olah dari 210 data sejumlah 42 data yang mendapat hasil prediksi 4 pasien yang diprediksi trombosit sedang dan dinyatakan positif, 22 pasien yang diprediksi trombosit sedang dan dinyatakan negatif, 16 pasien yang diprediksi trombosit rendah dan nyatanya positif, dan 0 pasien juga yang diprediksi trombosit rendah dan nyatanya negatif. Dan hasil dari grafik *confusion matrix* adalah pada pasien yang diprediksi berstatus positif atau yang dilambangkan angka 1 mendapat precision 1.00 atau 100% dan recall sebesar 0.80 atau 80% sedangkan untuk pasien yang diprediksi berstatus

negatif atau yang dilambangkan dengan angka 0 mendapat precision 0.85 atau 85% dan recall 1.00 atau 100%.

Dilanjut ke tahap yang terakhir adalah tahap atau langkah terakhir yaitu untuk melihat tingkat keakuratan atau seberapa baik model klasifikasi yang kita pakai.

	precision	recall	f1-score	support
0	1.00	0.80	0.89	20
1	0.85	1.00	0.92	22
accuracy			0.90	42
macro avg	0.92	0.90	0.90	42
weighted avg	0.92	0.90	0.90	42

Tingkat Akurasi : 90 persen

Gambar 11 Output Dari Hasil Menghitung Akurasi Data

Setelah serangkaian proses di atas dilakukan maka didapat tingkat akurasi untuk model klasifikasi Naïve Bayes didapat 90%. Dengan *f1-score* 89% dari angka negatif, *f1-Score* positif di dapat 92%.

#### IV. KESIMPULAN

Penelitian ini dilakukan dengan tujuan menganalisa penyakit Demam Berdarah di Puskesmas Taman Krocok Wonosari. Penerapan metode naïve bayes bertujuan memudahkan dalam menganalisis penyakit pasien di Puskesmas. Dari beberapa tahapan yang sudah dilalui dapat kita simpulkan bahwa penyakit DBD ini adalah penyakit serius dalam hal penanganannya, terlebih pada usia anak-anak dan remaja karena pada usia itu sangat rentan terkena penyakit DBD, diantaranya kurangnya menjaga kebersihan lingkungan dan pola hidup sehat.

Dalam penelitian ini menggunakan data training dan data testing sebanyak 20% dari 210 data pasien yang diagnose positif dan juga negative dengan menggunakan 4 atribut yang sama. Berdasarkan hasil uji coba yang dilakukan menggunakan naïve bayes adalah variable satu dengan yang lain memiliki keterkaitan dalam menganalisa diagnose penyakit DBD. Jumlah atribut yang digunakan juga mempengaruhi keakuratan terhadap model algoritma yang digunakan. Setelah analisis dilakukan didapatkan hasil diagnose negative dan positif. Untuk negative mendapatkan *precision* 10%, *recall* sebesar 80%, *f1-score* sebesar 89%, *support* 20. Sedangkan untuk diagnose penyakit DBD positif mendapatkan *precision* sebesar 85%, *recall* sebesar 100%, *f1-score* sebesar 92% , dan *support* sebesar 22. Dan mendapatkan tingkat akurasi sebesar 90% yang baik untuk hasil proses analisis.

#### DAFTAR PUSTAKA

- [1] D. N. Anggina, A. Ghiffari, and M. Abidinsyah, "Pengaruh penyuluhan terhadap tingkat pengetahuan pengunjung pasien DBD yang dirawat di RS Palembang," *Jurnal Kesehatan Medika Saintika*, 12(1), 21-25, 2021.
- [2] E. A. Zurroh, *Klasifikasi diagnosa penyakit Demam Berdarah Dengue menggunakan metode Hybrid Naive Bayes-K Nearest Neighbor*, 2021, (Doctoral dissertation, UIN Sunan Ampel Surabaya).
- [3] Rokom. (2022, juni 15). *sehatnegeriku*. Retrieved agustus 15, 2022, from Kasus DBD Meningkat, Kemenkes Galakkan Gerakan 1 Rumah 1 Jumentik (G1R1J): <https://sehatnegeriku.kemkes.go.id/baca/umum/20220615/0240172/kasus-dbd-meningkat-kemenkes-galakkan-gerakan-1-rumah-1-jumentik-g1r1j/>
- [4] A Riansyah, Pratama (2021). *Pernan Pemerintah Daerah dan Masyarakat dalam Upaya Kewaspadaan Dini Mengenai Covid-19 Perspektif Fiqh Siyash (Studi di Kecamatan Natar Kabupaten Lampung Selatan)* (Doctoral dissertation, UIN Raden Intan Lampung).
- [5] F. Qatada, *Hubungan Spiritualitas Dengan Tingkat Kecemasan Penderita Tuberkulosis Dalam Kesiapan Menerima Informasi Terapi Lanjutan Di Puskesmas Taman Krocok Kabupaten Bondowoso, 2020*, (Doctoral dissertation, Universitas Muhammadiyah Jember).
- [6] S. Umami, "Relationship Between Burnout And Self Efficacy In Nurses In Puskesmas Taman Kecamatan Taman Krocok Bondowoso District". *Jurnal Kesehatan dr. Soebandi*, 6(1), 2018.
- [7] A. Z. Mafakhir, A. Solichin, "Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta," *Fountain of Informatics Journal*, 5(1), 21, 2020.
- [8] R. Maya, F. Lubis, "Pengembangan Analisa Algoritma Autoregressive Integrated Moving Average ( Arima- Box Jenkins ) Pemodelan Menggunakan Google Colab ( Phytion )". July, 44-56, 2019.

**Moh. Jasri**, Meraih gelar sarjana komputer (S.Kom) dari Sekolah Tinggi Teknologi Nurul Jadid pada tahun 2009. Kemudian meraih gelar Master (M.Kom) dari Universitas Dian Nuswantoro pada tahun 2013 Saat ini Penulis menjadi dosen Program Studi Sistem Informasi di Universitas Nurul Jadid.

Andi Wijaya, Meraih gelar sarjana komputer (S.Kom) dari Sekolah Tinggi Teknologi Nurul Jadid pada tahun 2009. Kemudian meraih gelar Master (M.Kom) dari Universitas Dian Nuswantoro pada tahun 2013 Saat ini Penulis menjadi dosen Program Rekayasa Perangkat Lunak di Universitas Nurul Jadid.

*Penerapan Data Mining untuk Klasifikasi Penyakit Demam Berdarah Dengue (DBD) Dengan Metode Naïve Bayes (Studi Kasus Puskesmas Taman Krocok) (Moh. Jasri)*