

Klasifikasi Penentuan Tingkat Penyakit Demam Berdarah dengan menggunakan Algoritma *Naïve Bayes* (Studi Kasus Puskesmas Nagreg)

Saeful Anwar¹, Revita Lestari Faujiah², Tuti Hartati³, Edi Tohidi⁴

Program Studi Teknik Informatika, STMIK IKMI Cirebon

saefulanwar@gmail.com, revitalestarifaujiah@gmail.com, tutihartati@gmail.com,
editohidi@gmail.com

Abstrak

Pesatnya perkembangan ilmu pengetahuan dan teknologi khususnya di bidang teknologi informasi dapat memunculkan suatu inovasi baru untuk menyajikan dan mengelola suatu informasi demi memenuhi kebutuhan informasi. Peranan teknologi pada bidang kesehatan maupun kedokteran telah banyak membantu menolong jiwa manusia dan telah menunjukkan peranan pentingnya. Demam Berdarah Dengue (DBD) adalah penyakit yang terdapat pada anak-anak dan dewasa dengan gejala utama demam, nyeri otot dan sendi, yang biasanya memburuk setelah dua hari pertama. Penyakit DBD merupakan salah satu masalah kesehatan masyarakat di Indonesia yang jumlah penderitanya cenderung meningkat dan penyebarannya mengakibatkan pendarahan. Penyakit DBD ditandai dengan ciri demam tinggi mendadak selama 2-7 hari tanpa sebab yang jelas disertai manifestasi seperti petekie, epistaxis kadang disertai muntah darah, berak darah, kesadaran menurun, bertandensi menimbulkan senjata (shock) dan kematian. Algoritma Naïve Bayes merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Dilanjut ke tahap yang terakhir adalah tahap atau langkah terakhir yaitu untuk melihat tingkat keakurasaan atau seberapa baik model klasifikasi yang kita pakai.

Kata kunci: Klasifikasi, Naïve Bayes, Penyakit DBD

Abstract

The rapid development of science and technology, especially in the field of information technology, can give rise to new innovations for presenting and managing information to meet information needs. The role of technology in the health and medical fields has helped a lot in helping the human spirit and has shown its importance. Dengue Hemorrhagic Fever (DHF) is a disease that occurs in children and adults with the main symptoms of fever, muscle and joint pain, which usually gets worse after the first two days. DHF is a public health problem in Indonesia where the number of sufferers tends to increase and its spread causes bleeding. Dengue fever is characterized by sudden high fever lasting 2-7 days without a clear cause accompanied by manifestations such as petechiae, epistaxis sometimes accompanied by vomiting of blood, diarrhea, decreased consciousness, tendency to cause shock and death. The Naïve Bayes algorithm is a form of data classification using probability and statistical methods. The algorithm uses Bayes' theorem and assumes that all attributes are independent or not interdependent given the values of the class variables. Another definition says that Naïve Bayes is a classification using probability and statistical methods discovered by the British scientist Thomas Bayes, namely predicting future opportunities based on previous experience. Proceeding to the final stage, the final stage or step is to see the level of accuracy or how well the classification of the model we are using is.

Keywords : Classification, Naïve Bayes, Dengue Disease

PENDAHULUAN

Pesatnya perkembangan ilmu pengetahuan dan teknologi khususnya di bidang teknologi informasi dapat memunculkan suatu inovasi baru untuk menyajikan dan mengelola suatu

informasi demi memenuhi kebutuhan informasi. Peranan teknologi pada bidang kesehatan maupun kedokteran telah banyak membantu menolong jiwa manusia dan telah menunjukkan peranan pentingnya. Demam Berdarah Dengue

(DBD) adalah penyakit yang terdapat pada anak-anak dan dewasa dengan gejala utama demam, nyeri otot dan sendi, yang biasanya memburuk setelah dua hari pertama. Penyakit DBD merupakan salah satu masalah kesehatan masyarakat di Indonesia yang jumlah penderitanya cenderung meningkat dan penyebarannya mengakibatkan pendarahan. Penyakit DBD ditandai dengan ciri demam tinggi mendadak selama 2-7 hari tanpa sebab yang jelas disertai manifestasi seperti petekie, epistaxis kadang disertai muntah darah, berak darah, kesadaran menurun, bertandensi menimbulkan senjata (shock) dan kematian.(Sembiring, 2021).

Penyakit Demam Berdarah merupakan penyakit endemik di kecamatan nagreg dan sampai saat ini masih merupakan masalah utama kesehatan masyarakat. Sebagian menganggap hal ini terjadi karena kurangnya kesadaran akan kebersihan lingkungan dan sebagian lagi menganggap karena pemerintah lambat dalam mengantisifasi dan merespon kasus ini. Musim hujan tiba maka perlu diwaspada adanya genangan – genangan air yang terjadi pada selokan yang buntuh, gorong – gorong yang tidak lancar serta adanya banjir yang berkepanjangan. Maka perlu dilakukannya pemberantasan vektor penyakit karena vaksin yang efektif masih belum tersedia. Pemberantasan vektor ini meliputi pemberantasan sarang nyamuk dan jentik. Pemberantasan ini meliputi pembersihan tempat penampungan air yang bersih yang merupakan sarana utama perkembangbiakan nyamuk. Tempat air bersih perlu dilindungi dengan ditutup yang baik. Aedes Aegypti adalah salah satu nyamuk spesies tropis yang sangat efisien sebagai penampung virus, karena sangat ganas dan berkembang biak secara subur pada lingkungan manusia. Satu nyamuk dapat menjangkiti beberapa orang dalam waktu singkat dan lebih dari 1 kali.

Tujuan utama dari penelitian ini adalah untuk melakukan klasifikasi agar dapat meningkatkan sebuah akurasi terhadap penentuan penyakit Demam Berdarah dengan menggunakan Algoritma Naïve Bayes. Menentukan pola pelatihan sehingga mampu melakukan klasifikasi apakah pasien terdiagnosa parah atau tidak parah. Hasil dari penelitian ini berupa hasil pengolahan kualitatif dan kuantitatif yang telah dikumpulkan dengan perhitungan berdasarkan model yang eksperimen dan

pengujian dalam penelitian melakukan prediksi terhadap dataset dengan metode naïve bayes. Eksperimen akan dilakukan terhadap dataset yang telah divalidasi.

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Algoritma Naïve Bayes merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Implikasi hasilnya dapat memberikan perkembangan ilmu pengetahuan dan teknologi, terutama dalam teknologi informasi, telah membawa perubahan signifikan dalam penyajian dan pengelolaan informasi untuk memenuhi kebutuhan. Peran teknologi dalam bidang kesehatan dan kedokteran juga telah membuktikan keberhasilannya dalam menyelamatkan nyawa dan menjadi elemen penting dalam penanganan penyakit. Salah satu penyakit yang menjadi fokus di daerah kecamatan nagreg adalah Demam Berdarah Dengue (DBD). Penelitian ini memiliki relevansi penting dalam konteks pencegahan dan penanganan penyakit DBD di kecamatan nagreg, di mana data mining, terutama melalui algoritma Naïve Bayes, digunakan sebagai pendekatan untuk menemukan pola atau informasi menarik dalam data kesehatan. Kesimpulan yang dapat ditarik adalah bahwa pemanfaatan teknologi informasi dan data mining dapat menjadi alat yang efektif dalam mendukung upaya pencegahan dan penanganan penyakit DBD di wilayah tertentu.

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Naïve Bayes telah menjadi focus penelitian yang signifikan, dan banyak publikasi ilmiah telah mengalami lebih luas dalam pemodelan

algoritmanya, terutama dalam pengolahan data berskala besar. Metode ini terbukti berhasil dalam berbagai konteks, menjadikannya populer dan umum digunakan dalam mengklasifikasi data besar terkait kasus tertentu.

Sejumlah penelitian sebelumnya telah mengulas algoritma ini dengan mendalam, terdapat beberapa penelitian sebelumnya yang mendukung dan referensi dalam penelitian ini. Seperti penelitian yang berjudul "Implementasi Metode Naïve Bayes untuk Prediksi Daerah Terjangkit Demam Berdarah di Provinsi Jambi" Membahas mengenai Demam Berdarah Dengue (DBD) ialah suatu kondisi penyakit yang diakibatkan oleh salah satu dari empat virus dengue, yakni virus dengue dengan serotipe-1, serotipe-2, serotipe-3, dan serotipe-4. Patogen ini dapat dengan mudah menular. Untuk mengantisipasi serta mengatasi penyebaran penyakit ini, diperlukan suatu metode yang mampu memproyeksikan daerah-daerah yang berisiko terjangkit Demam Berdarah Dengue (DBD). Penelitian ini bertujuan untuk memprediksi tingkat penyebaran DBD di wilayah Provinsi Jambi. Analisis dilakukan dengan menggunakan Metode Naïve Bayes untuk memproyeksikan daerah-daerah yang memiliki potensi terjangkit DBD di Provinsi Jambi. Hasil penelitian menunjukkan bahwa aplikasi Rapidminer mencapai tingkat akurasi tertinggi sebesar 100%.(Waru & Wahyuning Astuti, 2021).

Paper yang berjudul "Penerapan Metode Algoritma K-Means Clustering untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (DBD)" Membahas mengenai DBD umumnya dipicu oleh gigitan nyamuk Aedes Aegypti, yang cenderung berkembang biak di lingkungan yang kurang bersih. Penelitian ini memiliki tujuan untuk menerapkan prinsip data mining dengan menggunakan algoritma K-Means (klastering) untuk memetakan desa-desa yang terdampak oleh Demam Berdarah Dengue (DBD). Pemetaan ini dilakukan untuk mengidentifikasi wilayah yang rentan terhadap penyakit DBD, serta wilayah yang memiliki risiko lebih rendah terjangkit. Hasil penelitian menunjukkan adanya tiga klaster untuk tingkat penyebaran DBD, yaitu klaster 1 dengan tingkat keparahan sedang, klaster 2 dengan tingkat keparahan tinggi, dan klaster 3 dengan tingkat keparahan rendah. (Sembiring, 2021).

Paper yang berjudul "Perbandingan Kinerja Algoritma Dempster Shafer dan Fuzzy-Naïve Bayes dalam Klasifikasi Penyakit Demam Berdarah dan Tifus" Membahas mengenai Demam Berdarah dan Tifus secara konsisten menjadi salah satu dari sepuluh penyakit terbanyak yang memerlukan perawatan rawat inap di Rumah Sakit Umum Daerah Kabupaten Buleleng. Pentingnya memberikan penanganan yang tepat sejak gejala awal muncul untuk mengurangi jumlah pasien rawat inap menjadi dasar untuk mendekripsi penyakit ini melalui pengamatan gejala awal. Penelitian ini fokus pada deteksi Demam Berdarah dan Tifus, dengan menggunakan metode Dempster Shafer atau sistem pakar berdasarkan pengetahuan dari seorang Dokter Spesialis di Rumah Sakit Umum Daerah Kabupaten Buleleng.(Sastrawan et al., 2019).

Paper yang berjudul "Prediksi Penyakit Demam Berdarah di Puskesmas Ngemplak Simongan Menggunakan Algoritma C4.5" Membahas mengenai Penyebaran penyakit ini cenderung lebih cepat di daerah perkotaan dibandingkan dengan daerah pedesaan, disebabkan oleh tingginya kepadatan penduduk. Nyamuk Aedes aegypti memiliki kemampuan yang memudahkan penyebaran virus dengue dari satu orang ke orang lain, terutama karena sifat domestiknya. Dalam upaya penanggulangan penyakit DBD, Departemen Kesehatan bekerja sama dengan puskesmas di sekitar, termasuk Puskesmas Ngemplak Simongan yang terletak di Kecamatan Semarang Barat. Dalam konteks ini, algoritma C4.5 digunakan untuk melakukan prediksi terhadap penyakit demam berdarah dengan tujuan menghasilkan pohon keputusan. Pemilihan algoritma ini disebabkan oleh kemampuannya dalam menggambarkan pola, pengetahuan, atau informasi secara eksplisit dalam bentuk pohon keputusan. Hasil pengujian menunjukkan tingkat akurasi sebesar 94.44%, menandakan bahwa aplikasi ini dapat digunakan dengan tepat.(Cholil et al., 2020).

Paper yang berjudul "Prediksi Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kabupaten Bandung menggunakan Hybrid Random Forest (RF) dan Genetic Algorithm (GA)" Membahas mengenai Demam Berdarah Dengue (DBD) merupakan suatu penyakit yang disebabkan oleh virus yang ditularkan melalui nyamuk Aedes Aegypti. Nyamuk tersebut umumnya ditemukan di wilayah tropis dan sub-

tropis, seperti di kawasan Asia Tenggara. Di Indonesia, khususnya di Kabupaten Bandung, jumlah kasus penyakit ini mengalami kecenderungan peningkatan, dan wilayah penyebarannya semakin meluas. Oleh karena itu, perolehan informasi mengenai penyebaran DBD sangat penting untuk membantu analisis dan menentukan tindakan yang perlu diambil untuk penanganan di masa depan. Metode yang digunakan adalah Hybrid Random Forest dan Genetic Algorithm untuk meramalkan penyebaran penyakit. Hasil prediksi disajikan menggunakan Geographic Information System (GIS). Dalam perbandingan antara penggunaan algoritma Random Forest tanpa Genetic Algorithm dan Random Forest dengan Genetic Algorithm, terdapat peningkatan akurasi rata-rata dari 83,42% menjadi 89,12%. (Utami et al., 2019).

Paper yang berjudul “Penerapan Metode K-Nearest Neighbor untuk Mengklasifikasi Penyebaran Kasus Demam Berdarah Dengue (DBD) di Kabupaten Maluku Tenggara” Membahas mengenai Penyakit ini dianggap sebagai ancaman serius karena perkembangan penyakitnya yang cepat dan dapat menyebabkan kematian dalam waktu singkat, sehingga menimbulkan kekhawatiran di kalangan masyarakat. Oleh karena itu, diperlukan penelitian mengenai penyebaran DBD untuk memungkinkan pengambilan tindakan pencegahan yang cepat terhadap kasus-kasus tersebut. Salah satu metode yang digunakan dalam penelitian ini adalah metode K-Nearest Neighbor (KNN), sebuah model pendukung keputusan yang mampu mengklasifikasikan data berdasarkan jarak terdekat. Hasil penelitian menunjukkan adanya dua kelompok penyebaran DBD di Kabupaten Maluku Tenggara. Pertama, terdapat Kecamatan yang memiliki potensi tinggi untuk penyebaran DBD, yakni Kecamatan Kei Besar, Kecamatan Kei Kecil, dan Kecamatan Kei Besar Selatan Barat. Kedua, terdapat kecamatan yang memiliki potensi penyebaran DBD yang rendah, yaitu Kecamatan Kei Besar Utara Barat, Kecamatan Kei Besar Selatan, Kecamatan Kei Besar Timur Selatan, Kecamatan Kei Besar Utara Timur, Kecamatan Hoat Sorbay, Kecamatan Kei Kecil Barat, Kecamatan Kei Kecil Timur, dan Kecamatan Manyeuw.(Matdoan, 2022).

Paper yang berjudul “Implementasi Algoritma Naïve Bayes Pada Sistem Pakar untuk Diagnosis Penyakit Demam Berdarah Dengue

berbasis Website” Membahas mengenai Sistem pakar untuk mendiagnosa penyakit DBD dibuat dalam bentuk aplikasi berbasis web, sehingga dapat diakses oleh masyarakat secara luas. Selain itu, aplikasi ini juga dapat membantu tenaga medis dalam pengambilan keputusan saat mendiagnosa penyakit DBD. Metode Naive Bayes merupakan suatu pendekatan klasifikasi probabilistik sederhana yang menghitung sekelompok probabilitas dengan merangkum frekuensi dan kombinasi nilai dari dataset yang diberikan (Saleh, 2015:209) dalam penelitian ini. Hasil dari penelitian ini ada 198 kasus yang terdiri dari 103 pasien yang terjangkit penyakit DBD dan 95 pasien yang tidak terjangkit penyakit DBD yang diperoleh dari UPTD Puskesmas Caringin, ditemukan 12 aturan yang dihasilkan dari wawancara dengan para pakar. (Nurlelah et al., 2019).

Dari penelitian terdahulu dalam hal tujuan, metode, ruang lingkup data, implikasi dan kesimpulan, serta inovasi yang ditawarkan. Penelitian penulis lebih fokus pada penggunaan algoritma Naïve Bayes dalam klasifikasi penyakit DBD dan memperluas pengetahuan dalam pencegahan dan penanganan penyakit tersebut melalui teknologi informasi dan data mining.

2.2 Data Mining

Data mining didefinisikan sebagai suatu proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Artinya, data mining merupakan upaya untuk mengekstrak pengetahuan yang bernilai dari volume data yang besar. Lebih lanjut, data mining juga dapat diartikan sebagai kegiatan pengekstrakan informasi baru yang diperoleh dari data yang sangat besar, dengan tujuan membantu dalam pengambilan keputusan.

Istilah data mining seringkali dikaitkan dengan knowledge discovery, karena melibatkan proses penemuan pengetahuan baru dari data yang ada. Salah satu teknik yang digunakan dalam data mining adalah penelusuran data untuk membangun model, yang selanjutnya dapat digunakan untuk mengenali pola data yang tidak terdapat dalam basis data yang disimpan. Proses ini sangat berguna dalam konteks pengambilan keputusan dan memenuhi kebutuhan prediksi.

Dalam konteks data mining, pengelompokan data juga dapat dilakukan. Teknik ini memungkinkan pengelompokan data

berdasarkan pola atau karakteristik tertentu, memberikan pemahaman lebih lanjut tentang struktur dan hubungan antar data. Sehingga, data mining menjadi alat yang sangat berharga dalam menggali wawasan dari data besar untuk mendukung proses pengambilan keputusan. (Sulistio et al., 2023)

2.3 Algoritma Naïve Bayes

Data Mining adalah sebuah istilah yang digunakan untuk menggali pengetahuan dalam big data. Data Mining berguna untuk mengidentifikasi informasi yang bermanfaat dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning dalam sebuah dataset besar.(Walid & Halimiyah, 2022)

Dalam penelitian ini akan menerapkan algoritma naïve bayes untuk menerapkan prediksi untuk data penyakit DBD untuk kategori parah dan tidak parah.

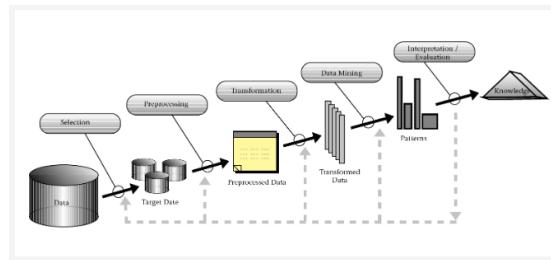
2.4 Klasifikasi

Klasifikasi adalah suatu metode yang digunakan untuk mendefinisikan sekelompok pola atau properti yang dapat menggambarkan serta membedakan satu kelas data dari kelas lainnya. Tujuannya adalah untuk menentukan apakah suatu objek memiliki perilaku dan atribut yang sesuai dengan kelompok yang telah ditentukan sebelumnya. Dalam konteks ini, klasifikasi tidak hanya mengidentifikasi kelas objek, tetapi juga menunjukkan hubungan antar atribut. Dengan menganalisis pola dan properti, klasifikasi memberikan pemahaman mendalam tentang karakteristik yang membedakan setiap kelas, memfasilitasi proses pengenalan dan pengelompokan data secara akurat(Washilaturrizqi, 2023).

METODE PENELITIAN

Pada tahap ini untuk mengolah data menggunakan Knowledge Discovery In Database (KDD) adalah suatu rangkaian proses nontrivial yang bertujuan untuk mencari dan mengidentifikasi pola dalam data. Proses ini melibatkan teknik integrasi dan penemuan ilmiah untuk memastikan bahwa pola yang ditemukan bersifat sah, baru, dapat bermanfaat, dan dapat dimengerti. KDD tidak hanya terkait dengan ekstraksi informasi dari data, tetapi juga melibatkan interpretasi dan visualisasi dari pola-

pola yang ditemukan dalam sejumlah kumpulan data.(Purnomo et al., 2023)



Gambar 1. Tahapan KDD

1. Data Selection

Penelitian ini akan membahas tentang pengelolaan data yang akan digunakan dalam penelitian, di mana data tersebut akan diproses menggunakan algoritma Naïve Bayes dengan metode Knowledge Discovery in Database (KDD). Dalam penelitian ini, sumber data yang digunakan adalah Data Kasus Jumlah Demam Berdarah Dengue (DBD) dari Puskesmas Nagreg pada rentang waktu 2021-2023. Variabel yang dianalisis dalam penelitian ini mencakup sebanyak 11 atribut, termasuk Jenis Kelamin, Demam, Sakit Ulu Hati, Perdarahan, Shock, Muntah, Pusing, Torniket, dan Trombosit, dengan jumlah data mencapai 204 record yang akan diuji.

2. Data Preprocessing

Pada tahap ini peneliti memilih atribut dari data demam berdarah yang ingin digunakan dan menghapus atribut lain yang tidak diperlukan dalam penelitian dengan bantuan operator di Google Colab.

3. Pemodelan algoritma Naïve bayes

Teknik pengklasifikasikan digunakan untuk mengembangkan model yang dapat mengklasifikasi data ke dalam kategori atau kelas yang sudah ditentukan. Dalam konteks penelitian ini teknik pengklasifikasian dapat digunakan untuk mengklasifikasikan pasien berdasarkan riwayat penyakit mereka, misalnya pasien dengan penyakit tertentu atau pasien yang berisiko tinggi terhadap penyakit tertentu.

4. Evaluasi

Hasil dari teknik data mining berupa pola dan model evaluasi untuk menentukan apakah benar-benar dapat dicapai.

HASIL DAN PEMBAHASAN

Hasil penelitian akan menyajikan hasil penemuan dan analisis selama penelitian ini berjalan. Setelah melakukan penelitian ini penulis menemukan pembahasan yang akan menjelaskan secara rinci langkah-langkah yang dilewati mulai dari pengambilan data, pengolahan data mentah menjadi data jadi pemrosesan algoritma dan evaluasi. Hasil dari penelitian ini berupa grafik dan table agar mudah dibaca dan memberi pemahaman yang mendalam oleh pembaca. Temuan yang didapatkan juga akan menjadi informasi yang mendasar bagi penelitian selanjutnya.

4.1 Data Selection

Dalam penelitian ini, sumber data yang digunakan adalah Data Kasus Jumlah Demam Berdarah Dengue (DBD) dari Puskesmas Nagreg pada rentang waktu 2021-2023. Variabel yang dianalisis dalam penelitian ini mencakup sebanyak 11 atribut, termasuk Jenis Kelamin, Demam, Sakit Ulu Hati, Perdarahan, Shock, Muntah, Pusing, Torniket, dan Trombosit, dengan jumlah data mencapai 204 record yang akan diuji. Format file yang dihasilkan adalah .csv.

4.2 Data Processing

Proses *pre-processing* melibatkan langkah-langkah seperti membersihkan data dari nilai yang tidak valid, normalisasi atau penyesuaian data, serta pengelompokan data untuk mempersiapkannya agar dapat digunakan dalam analisis menggunakan algoritma *Naïve Bayes* dengan metode *Knowledge Discovery in Database* (KDD).

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score, classification_report
```

Gambar 2. Proses instalasi

Pada Gambar 2 menunjukkan Library **pandas** untuk mengimpor pustaka dan memberi alias umumnya digunakan untuk membaca data dan melakukan data frame.

```
df = pd.read_csv('DBD_fix.csv')
df.head(1000)
```

Gambar 3. Membaca File Csv

Untuk membaca file CSV dengan nama 'DBD_fix.csv' menggunakan pustaka pandas (pd) dan menampilkan 1000 baris pertama dari DataFrame yang dibaca.

```
# Buatlah nilai 'ya' dengan 1 dan 'tidak' null dengan 0 pada kolom kolom tertentu
df.columns_to_process = df.columns_to_process.replace(['YA': 1]).fillna(0)
df.columns_to_process = df.columns_to_process.replace([''YA': 1]).fillna(0)
df = df.loc[(df.columns_to_process) != 0].any(axis=1)]
# Menghapus baris yang memiliki semua nilai 0 pada kolom tertentu adalah 0.
df = df.loc[(df.columns_to_process) != 0].any(axis=1)]
df['TROMBOSIT'] = df['TROMBOSIT'].replace(['-'], np.nan).astype(int)
# Tampilkan DataFrame yang telah dibersihkan
df.head(1000)
```

Gambar 4. DataFrame

Pada Gambar 4 menjelaskan untuk membersihkan DataFrame dengan mengganti nilai 'Ya' menjadi 1, mengisi nilai NaN dengan 0, menghapus baris yang memiliki semua nilai 0 pada kolom tertentu, dan mengonversi kolom 'TROMBOSIT' ke tipe data integer.

4.3 Data Transformation

Setelah melalui proses *preprocessing*, yakni tahap transformation, data telah mengalami modifikasi atau perubahan sehingga dapat digunakan dengan efektif dalam proses data mining.

	JENIS KELAMIN	DEMAM	SAKIT ULU HATI	PERDARAHAN	SHOCK	MUNTAH	PUSING	TORNIKET	TROMBOSIT	HEPATOBESIT	ISTIA	Tingkat penyakit
0	P	YA	YA	NAN	YA	NAN	NAN	NAN	NAN	NAN	48	paruh
1	P	YA	YA	YA	NAN	YA	NAN	NAN	NAN	NAN	30	tidak
2	P	YA	YA	YA	YA	YA	NAN	NAN	NAN	NAN	42	tidak
3	P	YA	YA	YA	YA	NAN	YA	NAN	NAN	NAN	19	tidak
4	P	YA	YA	YA	NAN	NAN	NAN	NAN	NAN	NAN	10	tidak
...
198	L	YA	YA	YA	NAN	YA	YA	NAN	NAN	27	tidak	
199	P	YA	NAN	NAN	NAN	YA	YA	NAN	19.000	NAN	13	tidak
200	P	YA	YA	YA	YA	YA	YA	YA	12.000	NAN	13	tidak
201	L	YA	YA	YA	NAN	NAN	YA	YA	57.000	48.4	23	tidak
202	L	YA	YA	YA	NAN	YA	YA	NAN	32.000	35.0	24	tidak

203 rows × 12 columns

Gambar 5. Proses Membersihkan Data

Pada Gambar 5 melakukan serangkaian pemrosesan pada DataFrame yang memuat informasi mengenai pasien demam berdarah.

	JENIS KELAMIN	DEMAM	SAKIT ULU HATI	PERDARAHAN	SHOCK	MUNTAH	PUSING	TORNIKET	TROMBOSIT	HEPATOBESIT	ISTIA	Tingkat penyakit
0	P	1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	paruh
1	P	1	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	tidak
2	P	1	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	42	tidak
3	P	1	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	19	tidak
4	P	1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	tidak
...
198	L	1	1.0	1.0	0.0	0.0	1.0	1.0	0	0.0	27	tidak
199	P	1	0.0	0.0	0.0	1.0	1.0	0.0	19000	0.0	13	tidak
200	P	1	1.0	1.0	1.0	1.0	1.0	1.0	12000	0.0	19	tidak
201	L	1	1.0	0.0	0.0	1.0	1.0	0.0	37000	48.4	23	tidak
202	L	1	1.0	1.0	0.0	1.0	1.0	0.0	34000	35.0	24	tidak

203 rows × 12 columns

Gambar 6. Hasil Pembersihan Data

Hasil dari Gambar 6 memiliki nilai 0 dan 1 yang data awalnya YA menjadi 1 dan NAN menjadi 0.

4.4 Data Mining

Data mining proses untuk menemukan pola atau informasi menarik dalam kumpulan data yang telah dipilih, menggunakan berbagai teknik atau metode.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

# Mengubah variabel kategorikal ke dalam bentuk numerik
le = LabelEncoder()
df['JENIS KELAMIN'] = le.fit_transform(df['JENIS KELAMIN'])
df['Tingkat penyakit'] = le.fit_transform(df['Tingkat penyakit'])

# Memisahkan kumpulan data menjadi fitur dan variabel target
X = df.drop('Tingkat penyakit', axis=1)
y = df['Tingkat penyakit']

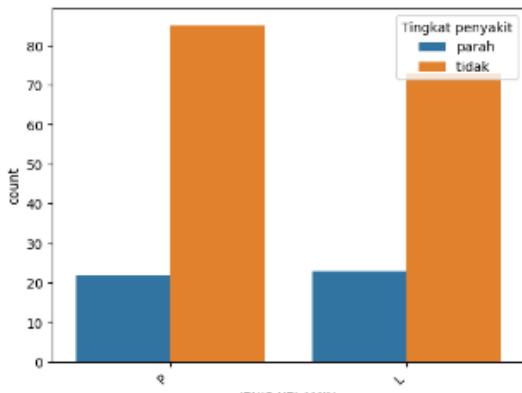
# Membagi dataset menjadi set pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Menampilkan bentuk (shape) dari dataset yang dihasilkan
(X_train.shape, X_test.shape, y_train.shape, y_test.shape)

((162, 11), (41, 11), (162,), (41,))
```

Gambar 7. Proses Pengujian dan Pelatihan

Untuk mengonversi variabel kategorikal ('JENIS KELAMIN' dan 'Tingkat penyakit') ke dalam bentuk numerik. Selanjutnya, data dibagi menjadi set pelatihan dan pengujian menggunakan `train_test_split`, dengan proporsi pengujian sebesar 20% dan pelatihan 80%. Akhirnya, bentuk (shape) dari dataset yang dihasilkan ditampilkan untuk memastikan bahwa pemisahan data telah dilakukan dengan benar.



Gambar 8. Grafik Count Plot

Grafik pada Gambar 8 menunjukkan jumlah frekuensi jenis kelamin dengan memperhitung tingkat penyakit untuk setiap kategori.

4.5 Evaluasi

Evaluasi model machine learning adalah proses mengukur seberapa baik model yang dilatih dapat melakukan prediksi atau klasifikasi pada data baru. Tujuannya adalah untuk menilai

performa model sebelum diimplementasikan, sehingga diketahui apakah model sudah cukup baik atau perlu perbaikan.

```
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import classification_report, accuracy_score

# Inisialisasi klasifikasi Gaussian Naive Bayes
model = GaussianNB()

# Melatih model
model.fit(X_train, y_train)

# Melakukan prediksi pada set pengujian
y_pred = model.predict(X_test)

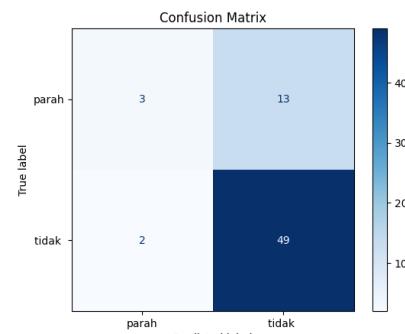
# Menghitung akurasi
accuracy = accuracy_score(y_test, y_pred)

# Menampilkan akurasi dan laporan klasifikasi
report = classification_report(y_test, y_pred, target_names=le.classes_)

print(accuracy, report)
```

Gambar 9. Menghitung Akurasi

Kode ini menggunakan model Gaussian Naive Bayes untuk melatih dan melakukan prediksi pada set pengujian.



Gambar 10. Confusion Matrix

Dari hasil confusion matrix pada Gambar 10 maka dapat dilihat hasil data testing atau data uji 20% yang kita olah dari 204 data sejumlah 67 data yang mendapat hasil prediksi.

	precision	recall	f1-score	support
parah	0.60	0.19	0.29	16
tidak	0.79	0.96	0.87	51
accuracy			0.78	67
macro avg	0.70	0.57	0.58	67
weighted avg	0.74	0.78	0.73	67

Gambar 11. 1 Output Dari Hasil Menghitung Akurasi Data

Setelah serangkaian proses diatas dilakukan maka didapat tingkat akurasi untuk model klasifikasi *Naïve Bayes* di dapat 78%. Dengan *f1-score* 29% dari angka parah, *f1-Score* tidak parah di dapat 87%.

SIMPULAN

Metode *Naïve Bayes* memberikan pendekatan yang efektif dalam mengklasifikasi penyakit demam berdarah. Dalam proses ini, langkah-langkah penting melibatkan pengumpulan dan *preprocessing* data, pemilihan fitur yang relevan, perhitungan probabilitas kelas dan kondisional, pengujian model, evaluasi kinerja, dan kemungkinan peningkatan model. Meskipun *Naïve Bayes* memiliki asumsi independensi antarfitur yang cukup sederhana, metode ini tetap mampu memberikan hasil yang baik, terutama pada dataset yang cukup besar dan dengan asumsi independensi yang dapat diterima. Tingkat akurasi yang memadai dalam klasifikasi penyakit demam berdarah hasilnya dapat bervariasi tergantung pada kualitas data, kompleksitas gejala penyakit. Meskipun *Naïve Bayes* memiliki asumsi independensi antarfitur yang sederhana, metode ini sering kali memberikan hasil yang baik. Setelah serangkaian proses diatas dilakukan maka didapat tingkat akurasi untuk model klasifikasi *Naïve Bayes* di dapat 78%. Dengan *f1-score* 29% dari angka parah, *f1-Score* tidak parah di dapat 87%.

Untuk mengembangkan lebih jauh dari penelitian ini, disarankan untuk mengembangkan lebih lanjut adanya pemetaan kesesuaian pengujian dengan menggunakan algoritma yang berbeda sebagai pembandingan untuk mengetahui metode tersebut dengan tingkat keakurasaan optimal sesuai kebutuhan.

DAFTAR PUSTAKA

- Cholil, S. R., Dwijayanto, A. F., & Ardianita, T. (2020). Prediksi Penyakit Demam Berdarah Di Puskesmas Ngemplak Simongan Menggunakan Algoritma C4.5. *Sistemasi*, 9(3), 529. <https://doi.org/10.32520/stmsi.v9i3.898>
- Matdoan, M. Y. (2022). Penerapan Metode K-Nearest Neighbor untuk Mengklasifikasi Penyebaran Kasus Demam Berdarah Dengue (DBD) di Kabupaten Maluku Tenggara. *Square : Journal of Mathematics and Mathematics Education*, 4(2), 75–82. <https://doi.org/10.21580/square.2022.4.2.13056>
- Nurlelah, E., Abdilah, A., & Ghani, M. A. (2019). Implementasi Algoritma Naive Bayes Pada Sistem Pakar Untuk Diagnosis Penyakit Demam Berdarah Dengue Berbasis Website. *Journal Speed-Sentra Penelitian Engineering Dan Edukasi*, 11(3), 1–8.
- Purnomo, R. K., Saufi, H. A., & Hs, R. (2023). Experimental Student Experiences. *Experimental Student Experience*, 1(1), 1–7.
- Sastrawan, A. S., Gunadi, I. G. A., & Sukajaya, I. N. (2019). Perbandingan Kinerja Algoritma Dempster Shafer Dan Fuzzy-Naive Bayes Dalam Klasifikasi Penyakit Demam Berdarah Dan Tifus. *Jurnal Ilmu Komputer Indonesia*, 4(2), 24–32. <https://ejournal-pasca.undiksha.ac.id/index.php/jik/article/view/3125>
- Sembiring, M. A. (2021). Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (Dbd). *Journal of Science and Social Research*, 4(3), 336. <https://doi.org/10.54314/jssr.v4i3.712>
- Sulistio, M. R., Suarna, N., & Nurdyawan, O. (2023). Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang. *Jurnal Teknologi Ilmu Komputer*, 1(2), 37–42. <https://doi.org/10.56854/jtik.v1i2.49>
- Utami, M. P., Fhira Nhita, M. ., & Annisa Aditsania, M. S. (2019). Prediksi Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kabupaten Bandung menggunakan Hybrid Random Forest (RF) dan Genetic Algorithm (GA). 6(2), 9963–9977.
- Walid, M., & Halimiyah, F. (2022). Klasifikasi Kemandirian Siswa SMA/MA Double Track Menggunakan Metode Naive Bayes. *Jurnal ICT : Information Communication & Technology*, 22, 190–197.
- Waru, D., & Wahyuning Astuti, R. (2021). Implementasi Metode Naïve Bayes Untuk Prediksi Daerah Terjangkit Demam Berdarah Dengue Di Provinsi Jambi. *Jurnal Teknik Informatika Kaputama (JTIK)*, 5(2), 240–245.
- Washilaturrizqi, O. N. (2023). Implementasi Algoritma C4 . 5 Untuk Menentukan Penerima. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 373–377.