

# Penerapan Algoritma (*Naïve Bayes*) Untuk Memprediksi Penyakit Diare

Repan<sup>1</sup>, Barry Ceasar Octariadi<sup>2</sup>, Sucipto<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer Universitas Muhammadiyah Pontianak  
[201220041repan@gmail.com](mailto:201220041repan@gmail.com)\*, [barry.ceasaro@unmuhpnk.ac.id](mailto:barry.ceasaro@unmuhpnk.ac.id), [sucipto@unmuhpnk.ac.id](mailto:sucipto@unmuhpnk.ac.id)

## Abstract

*Diarrhea is a serious health problem that can cause severe complications and even death, especially in children under the age of five in countries with low to medium mortality rates. At the Bunut Hilir Health Center, early detection of diarrhea is very important to prevent the risk of further complications. This study aims to develop a diarrhea disease prediction model using the Naïve Bayes method to improve the effectiveness of early detection. The data used came from medical records of Bunut Hilir Health Center patients, with a total of 490 case data covering clinical symptoms and environmental risk factors. The model was built using the Python programming language and implemented in the form of an interactive interface using Streamlit. Model evaluation was carried out using a confusion matrix, with an accuracy of 89%, precision 0.87, recall 0.90, and an F1-score value of 0.88. These results indicate that the Naïve Bayes method is effective in predicting diarrhea. The implementation of this model has the potential to improve the quality of health services and accelerate medical responses, especially in detecting diarrhea cases more quickly and accurately in the Bunut Hilir Health Center work area.*

*Keywords: Diarrhea, Prediction, Naïve Bayes, Data Mining*

## Abstrak

Diare merupakan salah satu masalah kesehatan serius yang dapat menyebabkan komplikasi berat hingga kematian, terutama pada anak-anak di bawah usia lima tahun di negara-negara dengan tingkat mortalitas rendah hingga menengah. Di Puskesmas Bunut Hilir, deteksi dini terhadap penyakit diare menjadi sangat penting untuk mencegah risiko komplikasi lanjutan. Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit diare dengan menggunakan metode *Naïve Bayes* guna meningkatkan efektivitas deteksi dini. Data yang digunakan berasal dari rekam medis pasien Puskesmas Bunut Hilir, dengan total sebanyak 490 data kasus yang mencakup gejala klinis dan faktor risiko lingkungan. Model dibangun menggunakan bahasa pemrograman *Python* dan diimplementasikan dalam bentuk antarmuka interaktif menggunakan *Streamlit*. Evaluasi model dilakukan menggunakan *confusion matrix*, dengan hasil akurasi sebesar 89%, *precision* 0.87, *recall* 0.90, dan nilai *F1-score* sebesar 0.88. Hasil tersebut menunjukkan bahwa metode *Naïve Bayes* efektif dalam memprediksi penyakit diare. Implementasi model ini berpotensi meningkatkan kualitas layanan kesehatan dan mempercepat respons medis, khususnya dalam mendeteksi kasus diare secara lebih cepat dan akurat di wilayah kerja Puskesmas Bunut Hilir.

Kata kunci: Diare, Prediksi, *Naïve Bayes*, Data Mining.

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

## 1. Pendahuluan

Dalam sektor kesehatan, rekam medis berfungsi sebagai instrumen penting untuk mengukur dan membandingkan guna memantau kemajuan kesehatan di suatu daerah. Selain itu, data rekam medis dapat diolah lebih lanjut untuk menghasilkan informasi berharga, seperti identifikasi pola penyakit.

Penelitian oleh N.M Putry. Komparasi Algoritma KNN dan *Naive bayes* untuk Klasifikasi Diagnosis Penyakit diabetes Melitus. Hasilnya menunjukkan akurasi sebesar 80% dan menunjukkan bahwa metode ini dapat membantu dalam sistem pendukung keputusan medis (*decision support system*) [1].

Sementara itu, studi oleh Cahyana dan A. Nurlayli menggunakan *Random Forest* untuk pendeteksi kanker Payudara dan menunjukkan akurasi mencapai 75%. Namun, kompleksitas dan waktu pelatihan model menjadi tantangan tersendiri dibanding *Naïve Bayes* [2].

Pengendalian penyakit menular, termasuk diare, melalui pengelolaan lingkungan terbukti lebih hemat biaya dibandingkan pengobatan, karena dapat mengurangi biaya perawatan medis [3]. Diare merupakan kondisi yang sering terjadi, di mana seseorang mengalami buang air besar lebih sering dari biasanya dengan tinja yang bertekstur cair atau encer. [4]. Beragam faktor lingkungan, seperti ketersediaan air bersih, fasilitas sanitasi seperti toilet, pengelolaan sampah rumah tangga, pengolahan limbah cair, serta kebersihan pribadi, menjadi faktor risiko utama yang berkontribusi terhadap terjadinya diare [5].

Puskesmas Bunut Hilir, sebagai fasilitas kesehatan utama di wilayah tersebut, memainkan peran vital dalam pencegahan, deteksi, dan penanganan penyakit diare. Meskipun jumlah pasien diare meningkat setiap tahunnya, dengan rata-rata satu kasus per hari, terutama akibat banjir dan kontaminasi air, Puskesmas ini menghadapi keterbatasan alat laboratorium untuk diagnosis dan penanganan efektif. Oleh karena itu, pengembangan sistem prediksi penyakit diare menjadi penting untuk memungkinkan *intervensi* yang lebih

cepat dan tepat. Penulis berencana mengembangkan sistem prediksi tersebut guna mendukung tenaga medis dalam mendeteksi kasus lebih awal dan memberikan intervensi yang sesuai.

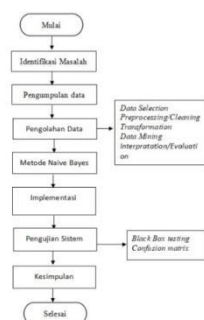
Algoritma *Naïve Bayes* merupakan metode klasifikasi yang digunakan untuk menangani ketidakpastian dalam data. Metode ini termasuk dalam kategori klasifikasi statistik yang memperkirakan probabilitas keanggotaan suatu kelas menganggap bahwa setiap fitur bersifat independen satu sama lain dalam menentukan kelas suatu data. Meskipun asumsi independensi ini jarang terpenuhi dalam praktik, algoritma ini sering memberikan hasil yang efektif dan efisien dalam berbagai aplikasi klasifikasi [6]. Metode ini berlandaskan asumsi bahwa setiap fitur dalam dataset tidak saling bergantung, yang dikenal sebagai asumsi *Naïve* [7]. Algoritma *Naïve Bayes* terkenal karena efektivitasnya dalam melakukan klasifikasi berbasis teks dan data kesehatan karena kesederhanaannya dan kemampuannya menangani data dengan banyak fitur. Algoritma ini juga mudah diimplementasikan dan diinterpretasikan.

Penelitian ini diharapkan dapat mempercepat proses diagnosis, meningkatkan akurasi identifikasi penyebab diare, serta memberikan rekomendasi pengobatan yang tepat dan cepat bagi pasien, sehingga dapat meningkatkan taraf hidup masyarakat di daerah tersebut.

Berdasarkan permasalahan yang telah dijelaskan, Dapat disimpulkan bahwa algoritma *Naïve Bayes* adalah metode yang efektif dalam memprediksi penyakit diare berdasarkan gejala yang dialami pasien. Dengan mengumpulkan serta menganalisis data gejala dan menerapkan prinsip probabilitas Bayes, algoritma ini mampu memberikan prediksi yang akurat. Hal ini pada akhirnya dapat mendukung upaya pencegahan serta pengendalian penyakit diare.

## 2. Metode Penelitian

Penelitian ini mengajukan penggunaan algoritma *Naïve Bayes* dalam memprediksi penyakit diare berdasarkan gejala yang dialami pasien. Metode ini diharapkan dapat meningkatkan ketepatan diagnosis serta membantu dalam upaya pencegahan dan pengendalian penyakit diare.



Gambar 1. Diagram Alur Metode Penelitian

### 2.1. Identifikasi Masalah

Penelitian ini difokuskan pada pengidentifikasian masalah di Puskesmas Bunut Hilir, di mana tingginya frekuensi kejadian penyakit diare menjadi isu utama yang memerlukan penanganan cepat dan efektif. Penerapan algoritma *Naïve Bayes* diusulkan sebagai solusi untuk memprediksi kemungkinan terjadinya penyakit diare, guna mendukung deteksi dini dan penanganan yang lebih baik.

### 2.2. Pengumpulan Data

Proses pengumpulan data dilakukan dengan menggunakan metode pengamatan langsung di lokasi penelitian, yaitu Puskesmas Bunut Hilir. Data dikumpulkan melalui wawancara bersama Bapak NS. Gunawan, S.Kep., dan Kepala Puskesmas Bunut Hilir. Penelitian ini memanfaatkan dataset yang berisi 490 data pasien yang tercatat selama periode tahun 2020 hingga 2023. Data tersebut disimpan dalam format file CSV untuk memudahkan pengolahan dan analisis. Analisis deskriptif terhadap dataset. Berikut adalah karakteristik data:

1. Proporsi Kelas Positif dan Negatif  
Dataset terbagi menjadi dua kelas, yaitu pasien yang terdiagnosis dengan gejala (kelas positif) dan pasien tanpa gejala (kelas negatif). Dari total 490 data pasien, sebanyak 60% termasuk dalam kelas positif, sementara 40% sisanya termasuk dalam kelas negatif.
2. Distribusi Gejala  
Dataset mencatat enam gejala utama yang diamati, yaitu demam, kehilangan nafsu makan, haus terus-menerus, mual dan muntah, feses lembek atau cair, serta perut kembung. Distribusi gejala menunjukkan bahwa gejala yang paling sering terjadi adalah feses lembek atau cair (70%) dan demam (65%), sedangkan gejala yang lebih jarang muncul adalah perut kembung (30%).
3. Faktor Demografis
  1. Umur: Sebagian besar pasien berusia di bawah 5 tahun (50%), dengan sisanya tersebar di berbagai kelompok usia lainnya.
  2. Jenis Kelamin: Pasien laki-laki (55%) sedikit lebih banyak dibandingkan pasien perempuan (45%).
  3. Wilayah: Data mencakup pasien dari beberapa desa di sekitar Puskesmas Bunut Hilir, dengan distribusi yang merata.

### 2.3. Pengolahan Data

Data mining, atau penambangan data, merupakan proses menemukan pola tersembunyi dari sekumpulan data besar menggunakan algoritma spesifik. Proses ini sering dikenal sebagai *Knowledge Discovery in*

Database (KDD), yang melibatkan identifikasi pola yang valid, inovatif, bermanfaat, dan mudah dipahami dari sekumpulan data [8].

a) Seleksi Data (*Data Selection*)

Data pasien mencakup Informasi seperti nama pasien, nomor registrasi atau NIK, jenis kelamin, usia, alamat, desa atau kelurahan, pekerjaan, serta tanggal kunjungan. serta gejala yang dialami seperti demam, hilang nafsu makan, haus terus-menerus, mual dan muntah, feses lembek atau cair, dan perut kembung. Untuk keperluan analisis, atribut terkait status kehamilan dan konfirmasi laboratorium telah dihapus.

b) Pemilihan Data (*Preprocessing/Cleaning*)

Tahap *preprocessing* dalam penelitian ini mencakup penyusunan dataset dengan melakukan pembersihan dan pemrosesan data. Proses ini penting untuk memastikan kualitas data sebelum analisis lebih lanjut. Pembersihan data mencakup identifikasi dan penghapusan duplikasi serta koreksi kesalahan, seperti kesalahan pengetikan atau *tipografi*. Setelah proses pembersihan, dataset yang terdiri dari 490 entri berhasil disiapkan untuk tahap analisis berikutnya [9]. Pada penelitian ini, tahap *preprocessing* dimulai dengan pengumpulan data melalui observasi langsung di Puskesmas Bunut Hilir. Data yang dikumpulkan meliputi nama pasien, nomor registrasi atau NIK, jenis kelamin, usia, alamat, desa atau kelurahan, pekerjaan, serta tanggal kunjungan, status kehamilan, serta gejala-gejala seperti demam, hilang nafsu makan, haus terus-menerus, mual dan muntah, feses lembek dan cair, perut kembung, konfirmasi laboratorium, dan jenis keracunan makanan. Setelah data terkumpul, dilakukan pembersihan dengan menghapus duplikasi dan memperbaiki kesalahan ketik (*tipografi*). Atribut yang tidak relevan, seperti status kehamilan dan konfirmasi laboratorium, juga dihapus untuk fokus pada gejala-gejala utama. Setelah proses pembersihan, dataset yang terdiri dari 490 entri siap untuk analisis lebih lanjut.

c) Transformasi (*transformation*)

Menyesuaikan data ke dalam format yang tepat untuk proses data mining merupakan langkah penting dalam memprediksi penyakit diare. Tahap ini mencakup pemrosesan data agar lebih mudah diinterpretasikan oleh algoritma data mining. Proses transformasi data yang diterapkan dapat dilihat pada Tabel dibawah ini.

Tabel 1. *Transformation* Suhu Badan (Demam)

Demam	Transformasi
>=37°- 38°	Rendah Sedang

39°- 40°

Tinggi

> 40°

Sangat Tinggi

d) *Data mining*

Data mining adalah proses menganalisis kumpulan data besar untuk menemukan pola dan hubungan yang berguna, menggunakan teknik dari pembelajaran mesin, statistik, dan manajemen basis data. Tujuannya adalah mengekstraksi informasi yang dapat dipahami dan digunakan untuk pengambilan keputusan yang lebih baik [10]. Tahap transformasi data adalah proses penting dalam mempersiapkan data untuk analisis menggunakan metode *Naïve Bayes*. Proses ini melibatkan pengolahan data yang telah diseleksi, dibersihkan, dan ditransformasikan menjadi format yang lebih sesuai untuk algoritma data mining. Transformasi data ini mempermudah algoritma dalam memahami dan memproses informasi yang ada. Detail mengenai transformasi data dapat dilihat pada Tabel 1.

e) Interpretasi/Evaluasi

Metode *Naïve Bayes* diterapkan untuk menghitung *probabilitas* berdasarkan data uji dengan kelas hasil yang belum diketahui. Sebagai ilustrasi, biasanya data X yang digunakan memiliki karakteristik tertentu. Setelah dilakukan perhitungan, nilai probabilitas untuk kelas positif lebih tinggi dibandingkan kelas negatif, yaitu sebesar 0,0066, sehingga data X diklasifikasikan ke dalam kelas positif.

Selain evaluasi model berdasarkan hasil klasifikasi, analisis kinerja Model dievaluasi dengan menggunakan beberapa metrik berikut:

- 1) Akurasi: Mengukur proporsi prediksi yang benar dari seluruh prediksi yang dilakukan. Hasil akurasi model adalah 89%, menunjukkan bahwa model cukup andal dalam klasifikasi data.
- 2) prediksi: Mengukur seberapa banyak positif yang benar-benar relevan. prediksi model tercatat sebesar 0,89, yang berarti sekitar 89% dari pasien yang diprediksi positif benar-benar tergolong positif.
- 3) *Recall*: Menilai sejauh mana model berhasil mengidentifikasi data positif dengan benar. Nilai *recall* sebesar 0,89 menunjukkan bahwa model dapat mengidentifikasi sebagian besar data positif dengan akurat data positif.
- 4) *F1-Score*: Menggabungkan *precision* dan *recall* menjadi metrik tunggal untuk mengevaluasi keseimbangan antara keduanya. *F1-score* model

adalah 0,89, menandakan performa model yang konsisten.

Metrik-metrik ini mengindikasikan bahwa model memiliki performa yang cukup baik, dengan keseimbangan antara nilai *precision* dan *recall*. *F1-score* yang mendekati nilai keduanya menunjukkan keandalan model dalam klasifikasi. Nilai *recall* menunjukkan bahwa model dapat mengidentifikasi kasus positif dengan cukup baik tanpa banyak menghasilkan kesalahan *false positive*.

#### 2.4. Metode Naïve Bayes

Metode *Naïve Bayes* merupakan metode klasifikasi yang efisien untuk menangani dua atau lebih kelas. Metode ini menggunakan teknik klasifikasi terawasi untuk memprediksi kategori masa depan dengan memberikan label kelas pada setiap *instance* atau catatan berdasarkan probabilitas bersyarat [11]. Algoritma ini dipilih karena *Naïve Bayes* merupakan algoritma yang sederhana namun sangat efektif dalam proses klasifikasi. Algoritma ini mudah diimplementasikan dan tidak memerlukan banyak parameter untuk diatur. Selain itu, *Naïve Bayes* dapat digunakan dengan data kualitatif (kategori) dan kuantitatif (*numerik*).

Metode *Naïve Bayes* merupakan metode klasifikasi probabilistik yang sederhana, yang menghitung probabilitas dengan menganalisis frekuensi serta kombinasi nilai dalam dataset yang tersedia [12].

$$P(H | X) = (P(X | H)P(H)) / (P(X)) \quad (1)$$

Dimana :

H = H merupakan hipotesis bahwa X termasuk dalam suatu kelas tertentu.

X = X merupakan kelasnya belum kita ketahui.

P(H|X) merupakan probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior).

P(H) merupakan probabilitas hipotesis H (probabilitas prior).

P(X|H) probabilitas X dengan kondisi hipotesis H.

P(X) merupakan probabilitas X secara keseluruhan.

Tahap ini, perhitungan dilakukan berdasarkan data uji di mana hasil kelasnya belum diketahui. Kelas yang mungkin terdiri dari:

C1 adalah Diagnosis Positif

C2 adalah Diagnosis Negatif

Diketahui terdapat X yang kelasnya belum kita tentukan.

Langkah Perhitungan :

1. Menghitung probabilitas kelas dalam data pelatihan:

$$P(\text{hasil diagnosis} = \text{positif}) = 222/490 = 0.45$$

$$P(\text{hasil diagnosis} = \text{negatif}) = 268/490 = 0.54$$

2. Menghitung probabilitas kondisi setiap atribut dalam data X untuk masing-masing kelas:

$$P(\text{demam} = \text{"39"} - \text{40} \mid \text{hasil diagnosis} = \text{positif}) = 180/222 = 0,810$$

$$P(\text{demam} = \text{"39"} - \text{40} \mid \text{hasil diagnosis} = \text{negatif}) = 110/268 = 0,410$$

$$P(\text{hilang\_nafsu\_makan} = \text{Ya} \mid \text{hasil diagnosis} = \text{positif}) = 172/222 = 0,774$$

$$P(\text{hilang\_nafsu\_makan} = \text{Ya} \mid \text{hasil diagnosis} = \text{negatif}) = 99/268 = 0,369$$

$$P(\text{haus\_terus\_menerus} = \text{Ya} \mid \text{hasil diagnosis} = \text{positif}) = 147/222 = 0,662$$

$$P(\text{haus\_terus\_menerus} = \text{Ya} \mid \text{hasil diagnosis} = \text{negatif}) = 147/268 = 0,548$$

$$P(\text{mual\_muntah} = \text{Tidak Sering} \mid \text{hasil diagnosis} = \text{positif}) = 75/222 = 0,337$$

$$P(\text{mual\_muntah} = \text{Tidak Sering} \mid \text{hasil diagnosis} = \text{negatif}) = 72/268 = 0,268$$

$$P(\text{fases\_lembek\_cair} = \text{Tidak} \mid \text{hasil diagnosis} = \text{positif}) = 99/222 = 0,445$$

$$P(\text{fases\_lembek\_cair} = \text{Tidak} \mid \text{hasil diagnosis} = \text{negative}) = 97/268 = 0,361$$

$$P(\text{perut\_kembung} = \text{Ya} \mid \text{hasil diagnosis} = \text{positif}) = 147/222 = 0,662$$

$$P(\text{perut\_kembung} = \text{Ya} \mid \text{hasil diagnosis} = \text{negatif}) = 97/268 = 0,361$$

3. Menghitung probabilitas gabungan untuk masing-masing kelas:

$$P(X \mid \text{hasil diagnosis} = \text{positif}) = 0,810 \times 0,774 \times 0,662 \times 0,337 \times 0,445 \times 0,662 = 0,0147$$

$$P(X \mid \text{hasil diagnosis} = \text{negatif}) = 0,410 \times 0,369 \times 0,548 \times 0,268 \times 0,361 \times 0,361 = 0,0106$$

4. Menghitung probabilitas akhir dengan memperhitungkan probabilitas kelas dalam data pelatihan:

$$P(X \mid \text{hasil diagnosis} = \text{positif})$$

$$P(\text{hasil diagnosis} = \text{positif})$$

$$= 0,0147 \times 0,45$$

$$= 0,0066$$

$$P(X \mid \text{hasil diagnosis} = \text{negatif})$$

$$P(\text{hasil diagnosis} = \text{negatif})$$

$$= 0,0106 \times 0,54$$

$$= 0,0057$$

#### 2.5. Implementasi

Tahap ini, model yang telah dikembangkan menggunakan Algoritma *Naïve Bayes* akan diimplementasikan melalui pemrograman dalam bahasa *Python* serta memanfaatkan *framework* *Streamlit* untuk membangun aplikasi web interaktif. Pemilihan Algoritma *Naïve Bayes* didasarkan data yang telah dikumpulkan dan diproses sebelumnya.

Proses ini mencakup pengkodean algoritma, pengelolaan dependensi, serta integrasi dengan sistem secara keseluruhan. Tujuan dari implementasi ini adalah untuk menciptakan sistem prediksi yang mampu memberikan hasil berdasarkan gejala yang dialami pasien. Dengan adanya sistem ini, petugas dapat menggunakan model prediksi penyakit diare secara lebih mudah dan efisien.

### 2.6. Pengujian Sistem

Dalam Pengujian ini, metode yang diterapkan yaitu *Blackbox Testing* dengan menggunakan teknik *Boundary Value Analysis*. Metode ini berfokus pada pengujian nilai yang berada di sekitar (*valid*) dan tidak (*valid*), yang bertujuan menentukan jumlah digit maksimum dan minimum yang akan jadi pengujian [13].

*Confusion matrix* adalah metode yang biasanya digunakan untuk membandingkan tingkat akurasi serta mengukur performa model dalam klasifikasi. Melalui evaluasi yaitu yang digunakan terhadap penelitian ini adalah *confusion matrix*, dapat diperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score* [14]. Akurasi dapat dihitung dengan membandingkan jumlah data yang berhasil diklasifikasikan dengan benar terhadap total jumlah sampel data yang diuji [15].

Untuk pengujian menggunakan *confusion matrix*, terdapat beberapa metrik evaluasi, yaitu;

1. Akurasi (*accuracy*): Akurasi mengukur Perbandingan antara jumlah prediksi yang benar dengan total keseluruhan hasil prediksi. Dalam konteks diagnosis diabetes, presisi menunjukkan rasio prediksi benar terhadap seluruh hasil yang diprediksi positif. Nilai presisi diperoleh dengan membagi *True Positive* (TP) dengan jumlah *True Positive* dan *False Positive* (TP + FP). Presisi sangat penting dalam penelitian di bidang Kesehatan [16].
2. Presisi (*precision*): Presisi mengukur persentase Proporsi prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan oleh model.
3. *Recall*: *Recall* mengukur seberapa banyak data positif yang sebenarnya berhasil terdeteksi dengan benar dibandingkan dengan total data positif yang ada.
4. *F1-score*: *F1-score* adalah rata-rata harmonik antara presisi dan *recall*. Metrik ini berguna dalam situasi di mana perlu mempertimbangkan keseimbangan antara presisi dan *recall* secara bersamaan.

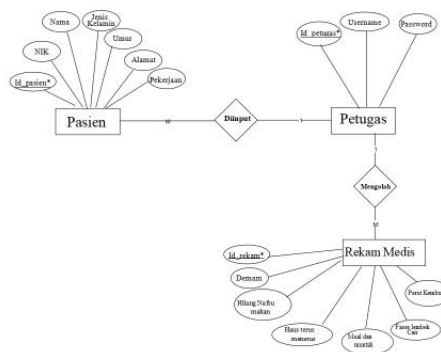
### 2.7 Kesimpulan

Kesimpulan diperoleh berdasarkan temuan dari hasil pengujian yang telah dilakukan serta analisis selama proses pengembangan aplikasi. Kesimpulan ini mencerminkan hasil penelitian yang sesuai dengan alur metode yang digambarkan dalam diagram penelitian.

## 3. Hasil dan Pembahasan

### 3.1. Spesifikasi

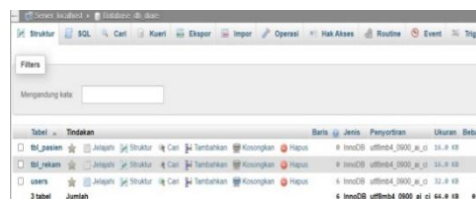
Perancangan sistem penelitian ini menggunakan *Entity-Relationship Diagram* (ERD) pada sistem Data Mining untuk prediksi penyakit diare. *Entity Relationship* adalah metode pemodelan basis data yang menggunakan skema konseptual sebagai salah satu jenis model data semantik sistem. Dalam pendekatan ini, basis data relasional dirancang menggunakan metode *top-down*. Model *Entity-Relationship* divisualisasikan melalui *Entity-Relationship Diagram* (ERD), yang untuk sebagai representasi dari entitas serta hubungan antar entitas dalam sistem [17].



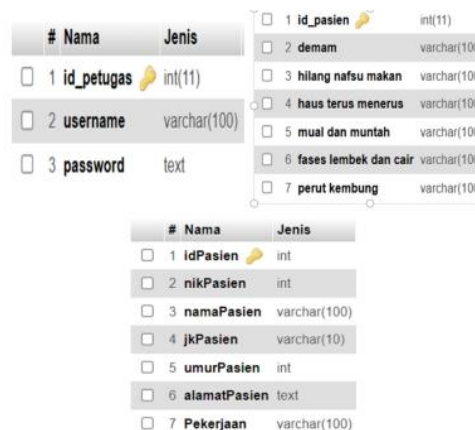
Gambar 2 Entity Relational Diagram

### 3.2. Rancangan Database

Rancangan Desain basis data serta hubungan antar tabel untuk prediksi penyakit diare dapat dilihat pada gambar di bawah ini.



Gambar 3 Rancangan Database



Gambar 4 Relasi Antar tabel

### 3.3. Implementasi

#### a. Halaman Login

Ketika ingin masuk petugas puskesmas harus *login* terdahulu, agar dapat melihat *menu* yang ada di website.



Gambar 5 Halaman *Login Dan Register*

- b. Halaman Beranda  
Tampilan Beranda pada Puskesmas Bunut Hilir berisi lokasi Puskesmas.



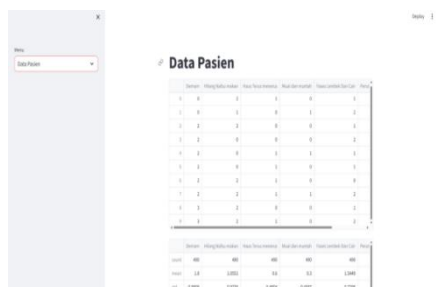
Gambar 6 Halaman Beranda

- c. Tampilan Grafik  
Menyajikan grafik berdasarkan atribut gejala yang dialami oleh pasien diare.



Gambar 7 Grafik Rekam Medis

- d. Halaman Data  
Tampilan dari data pasien dan penginputan.



Gambar 8 Data pasien



Gambar 9 Hasil Pengimputan

### 3.4. Pengujian sistem

Dalam penelitian ini, pengujian *Black Box Testing* dilakukan untuk memastikan bahwa sistem prediksi diare berfungsi dengan baik sesuai dengan spesifikasi yang telah ditetapkan.

Tabel 2. Hasil *Black Box Testing*

No	Skenario pengujian	Kasus pengujian	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
1	Login dan Register	Username dan Password	Berhasil	Sesuai	Valid
2	Beranda	Titik lokasi	Berhasil	Sesuai	Valid
3	Grafik	Grafik Data pasien	Berhasil	Sesuai	Valid
4	Data pasien dan pengimputan	Data Pasien dan Gejala	Berhasil	Sesuai	Valid
5	Pola <i>Naive Bayes</i>	Prediksi	Berhasil	Sesuai	Valid
6	Keluar	Tombol Keluar	Berhasil	sesuai	Valid

Pengujian dalam penelitian ini juga menggunakan *confusion matrix* untuk mengevaluasi kinerja sistem dalam memprediksi penyakit diare, seperti pada gambar di bawah ini.

```
print(confusion_matrix(y_test,naive_bayes_prediction))
[[43  5]
 [ 6 44]]

print(classification_report(y_test,naive_bayes_prediction))

          precision    recall  f1-score   support

     0       0.88      0.90      0.89         48
     1       0.90      0.88      0.89         50

 accuracy          0.89
 macro avg         0.89      0.89      0.89
weighted avg         0.89      0.89      0.89
```

Gambar 10. Hasil Pengujian

### 4. Kesimpulan

Penelitian ini berhasil mengimplementasikan model prediksi menggunakan metode *Naive Bayes* pada Puskesmas Bunut Hilir. Model ini dikembangkan dengan memanfaatkan data rekam medis pasien yang mencakup berbagai *variabel* yang berkaitan dengan

kondisi pasien serta gejala yang mereka alami. Proses implementasi meliputi tahap pengumpulan data, pra-pemrosesan, pelatihan model, dan pengujian model.

Hasil implementasi menunjukkan bahwa metode *Naïve Bayes* dapat digunakan secara efektif untuk memprediksi penyakit diare pada pasien dengan mempertimbangkan faktor-faktor risiko yang relevan. Model yang dikembangkan menunjukkan tingkat akurasi sebesar 89%, yang merupakan indikator bahwa metode ini cukup efektif dalam melakukan prediksi penyakit diare.

Meskipun demikian, terdapat beberapa keterbatasan yang perlu diperhatikan. Salah satu keterbatasan utama dari metode ini adalah asumsi independensi antar fitur, yang berarti *Naïve Bayes* menganggap bahwa setiap fitur yang digunakan dalam model bersifat saling independen. Pada kenyataannya, dalam banyak kasus, fitur-fitur tersebut mungkin saling berkorelasi. Ketika asumsi ini tidak terpenuhi, akurasi model dapat terpengaruh karena model mungkin gagal menangkap hubungan yang lebih kompleks antar fitur.

Selain itu, keterbatasan lain yang perlu dicatat adalah keterbatasan jumlah dan kualitas data yang digunakan. Data yang diperoleh berasal dari satu sumber, yaitu Puskesmas Bunut Hilir, sehingga hasil dan model yang dikembangkan mungkin belum dapat digeneralisasi ke wilayah atau populasi yang lebih luas. Faktor lain seperti ketidakseimbangan kelas pada data (*class imbalance*), kemungkinan adanya data yang tidak lengkap atau tidak akurat, serta keterbatasan variabel yang tersedia juga dapat memengaruhi kinerja model.

Saran untuk pengembangan di masa depan, antara lain:

1. Penggunaan metode lain yang lebih kompleks: Untuk meningkatkan akurasi dan menangani keterbatasan asumsi independensi, dapat dipertimbangkan penggunaan metode lain seperti *Random Forest*, *Support Vector Machine (SVM)*, atau model berbasis *deep learning*.
2. Pemanfaatan teknik seleksi dan rekayasa fitur: Menggunakan teknik feature selection dan feature engineering yang lebih canggih dapat membantu meningkatkan kualitas input model serta mengurangi dampak korelasi antar fitur.
3. Peningkatan dan perluasan data: Menambahkan data dari berbagai fasilitas kesehatan lain serta memperluas jumlah data historis dapat membantu membuat model yang lebih *robust* dan *generalizable*.
4. Penanganan data tidak seimbang: Menggunakan teknik seperti *SMOTE (Synthetic Minority Over-sampling Technique)* untuk mengatasi ketidakseimbangan data antar kelas penyakit.

5. Validasi lintas lokasi dan populasi: Untuk menguji keandalan model, perlu dilakukan pengujian lebih lanjut dengan data dari wilayah atau populasi yang berbeda.

#### Daftar Rujukan

- [1] N. M. Putry, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [2] C. W. Cahyana and A. Nurlayli, "Analisis Performa Logistic Regression, Naïve Bayes, dan Random Forest sebagai Algoritma Pendeteksi Kanker Payudara," *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 4, no. 1, pp. 51–64, 2023.
- [3] Kholis Ernawati, *Malaria: Penyebab Lingkungan Dan Pengendalian*, no. September. 2022.
- [4] M. B. A. Sinum, "Hubungan Program Open Defecation Free (ODF) oleh Pemerintah dengan Kejadian Diare," *J. Med. Hutama*, vol. 2, no. 3, pp. 928–933, 2021, [Online]. Available: <http://jurnalmedikahutama.com/index.php/JMH/article/view/191/125>
- [5] A. A. Iryanto, T. Joko, and M. Raharjo, "Literature Review : Faktor Risiko Kejadian Diare Pada Balita Di Indonesia," *J. Kesehat. Lingkung.*, vol. 11, no. 1, pp. 1–7, 2021, doi: 10.47718/jkl.v11i1.1337.
- [6] G. W. N. Wibowo and M. A. Manan, "Penerapan Algoritma Naive Bayes Untuk Prediksi Heregistrasi Calon Mahasiswa Baru," *JTINFO (Jurnal Tek. Inform.)*, vol. 1, no. 1, pp. 1–10, 2022, [Online]. Available: <https://journal.unisnu.ac.id/JTINFO/article/view/126>
- [7] D. Laila Sari, M. Saputra, and H. Gemasih, "Penerapan Data Mining Dalam Proses Prediksi Perceraian Menggunakan Algoritma Naive Bayes Di Kabupaten Aceh Tengah," *J. Tek. Inform. dan Elektro*, vol. 4, no. 1, pp. 23–35, 2022, doi: 10.55542/jurtie.v4i1.112.
- [8] D. Ramdhan, G. Dwilestari, R. D. Dana, A. Ajiz, and K. Kaslani, "Clustering Data Persediaan Barang Dengan Menggunakan Metode K-Means," *MEANS (Media Inf. Anal. dan Sist.)*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.54367/means.v7i1.1826.
- [9] F. P. Nursyamsy and F. N. Hasan, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Sentimen Terhadap Aplikasi Identitas Kependudukan Digital Menggunakan Algoritma Naïve Bayes dan SVM," *Media Online*, vol. 4, no. 3, pp. 1788–1798, 2023, doi: 10.30865/klik.v4i3.1517.
- [10] B. N. Sari *et al.*, "PENERAPAN DATA MINING DALAM KLASIFIKASI DATA TRANSAKSI PRODUK KOPERASI DI SMK PGRI 2 KARAWANG," vol. 9, no. 1, pp. 263–269, 2025.
- [11] M. Irfan and A. A. Syukron, "IMPLEMENTASI DATA MINING UNTUK MENGLASIFIKASI PENJUALAN PRODUK TERLARIS PADA KOPERASI AL-BAEDLOWI," vol. 9, no. 1, pp. 1423–1428, 2025.
- [12] Raja Rizki Alanta Nasution and Relita Buatun, "Penerapan Metode Naive Bayes dalam Menentukan Diagnosa Kerusakan pada Smartphone," *Indones. J. Educ. Comput. Sci.*, vol. 2, no. 1, pp. 24–33, 2024, doi: 10.60076/indotech.v2i1.383.
- [13] S. R. Yulistina, T. Nurmala, R. M. A. T. Supriawan, S. H. I. Juni, and A. Saifudin, "Penerapan Teknik Boundary Value Analysis untuk Pengujian Aplikasi Penjualan Menggunakan Metode Black Box Testing," *J. Inform. Univ. Pamulang*, vol. 5, no. 2, p. 129, 2020, doi: 10.32493/informatika.v5i2.5366.
- [14] I. N. Abrar, A. Abdullah, and S. Sucipto, "Liver Disease Classification Using the Elbow Method to Determine Optimal K in the K-Nearest Neighbor (K-NN) Algorithm," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 2, pp. 218–228, 2023, doi: 10.32736/sisfokom.v12i2.1643.
- [15] N. S. Fauziah and R. D. Dana, "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari," *Blend Sains J. Tek.*, vol. 1,

- no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.
- [16] S. A. Hicks *et al.*, “On evaluation metrics for medical applications of artificial intelligence,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–9, 2022, doi: 10.1038/s41598-022-09954-8.
- [17] D. T. B. Purba, H. Tawaqal, R. Fachrudin, and F. Sinlae, “Design Database Pada Sistem Informasi Reservasi Hotel,” *J. Siber Multi Disiplin*, vol. 2, no. 2, pp. 144–151, 2024.