

PENERAPAN ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI PENYAKIT BERDASARKAN GEJALA DAN PROFIL PASIEN

Indi Ramadhani Sitorus¹, Dicky Apdilah²

Universitas Asahan, Asahan

email: ¹rahmadaniindi44@gmail.com, ²dickyapdi1404@gmail.com

Abstract: *This study aims to implement the Naive Bayes algorithm to develop a web-based disease classification system utilizing patient symptoms and profile data. The dataset, obtained from Kaggle, consists of 349 entries with five primary attributes: fever, cough, fatigue, difficulty breathing, and age group. Classification is performed by manually calculating the posterior probability of each disease class using the Naive Bayes approach. The model is integrated into a PHP-MySQL-based system and evaluated using 20 randomly selected test data. Performance evaluation resulted in an accuracy of 45%, with a precision of 0.45, recall of 0.45, and F1-score of 0.45. These results indicate that the Naive Bayes algorithm is applicable as an initial method in developing a symptom- and profile-based disease classification system, although further improvement is needed through feature enrichment and model optimization.*

Keyword: *naive bayes; disease classification; symptoms; patient profile; system evaluation*

Abstrak: Penelitian ini bertujuan menerapkan algoritma *Naive Bayes* untuk membangun sistem klasifikasi penyakit berbasis web dengan memanfaatkan data gejala dan profil pasien. Dataset yang digunakan diperoleh dari Kaggle, terdiri atas 349 entri dengan lima atribut utama: demam, batuk, kelelahan, kesulitan bernapas, dan kelompok umur. Proses klasifikasi dilakukan dengan menghitung probabilitas posterior setiap kelas penyakit secara manual menggunakan pendekatan *Naive Bayes*. Model diterapkan ke dalam sistem berbasis PHP dan MySQL, kemudian diuji menggunakan 20 data acak sebagai data uji. Evaluasi performa menghasilkan akurasi sebesar 45%, precision 0,45, recall 0,45, dan F1-score 0,45. Hasil ini menunjukkan bahwa algoritma *Naive Bayes* dapat diterapkan sebagai metode awal dalam pengembangan sistem klasifikasi penyakit berbasis data gejala dan karakteristik pasien, meskipun performanya masih perlu ditingkatkan melalui pengayaan fitur dan optimasi algoritma.

Kata Kunci: *naive bayes; klasifikasi penyakit; gejala; profil pasien; evaluasi sistem*

PENDAHULUAN

Kesehatan merupakan aspek fundamental dalam kehidupan manusia yang sangat memengaruhi kualitas hidup dan produktivitas. Di tengah meningkatnya kesadaran akan pentingnya menjaga kesehatan, tantangan dalam mengenali dan mendeteksi penyakit sejak dini masih kerap terjadi. Banyak kasus keterlambatan penanganan disebabkan oleh kurangnya pemahaman masyarakat terhadap gejala awal

penyakit. Seiring dengan kemajuan teknologi informasi, khususnya dalam bidang kecerdasan buatan (AI) dan pembelajaran mesin (machine learning), upaya deteksi penyakit secara otomatis menjadi semakin memungkinkan.

Salah satu metode pembelajaran mesin yang banyak digunakan dalam klasifikasi penyakit adalah *Naive Bayes*. Metode ini dikenal karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam

menghasilkan prediksi yang cukup akurat meskipun dengan asumsi independensi antar fitur (Purba & Syahputra, 2021). Penggunaannya telah terbukti efektif dalam berbagai kasus medis, seperti dalam klasifikasi penyakit kulit berbasis sistem pakar (Pohan & Chairunisah, 2024), penyakit jantung (Riani et al., 2019), serta penyakit ginjal kronik (A'yuniyah et al., 2022).

Meskipun begitu, sebagian besar penelitian sebelumnya hanya berfokus pada satu jenis penyakit atau atribut gejala saja tanpa mempertimbangkan profil pasien. Padahal, kombinasi antara data gejala dan data profil pasien seperti usia, jenis kelamin, tekanan darah, dan kolesterol dapat meningkatkan akurasi klasifikasi (Guines et al., 2022; Anggraini et al., 2021).

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan sistem klasifikasi penyakit berbasis web menggunakan algoritma *Naive Bayes* dengan mempertimbangkan data gejala dan profil pasien. Dataset yang digunakan merupakan data sekunder dari platform Kaggle, berjudul Disease Symptoms and Patient Profile Dataset, yang terdiri dari 349 entri dengan atribut demam, batuk, kelelahan, kesulitan bernapas, dan usia (Kaggle, 2024). Penelitian ini bertujuan untuk menerapkan metode *Naive Bayes* dalam membangun sistem klasifikasi penyakit serta mengevaluasi performanya dalam mendeteksi penyakit berdasarkan kombinasi atribut gejala dan karakteristik pasien.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis algoritma *Naive Bayes*. Tujuan utamanya adalah membangun sistem prediksi penyakit yang mampu mengklasifikasikan jenis penyakit berdasarkan data gejala dan profil pasien. Proses dilakukan secara bertahap, mulai dari persiapan data,

perhitungan probabilitas dengan *Naive Bayes*, hingga implementasi sistem dalam bentuk aplikasi web berbasis PHP dan MySQL.

Metode *Naive Bayes* dipilih karena kemampuannya dalam menangani data berskala besar, serta efisiensinya dalam melakukan klasifikasi dengan asumsi independensi antar fitur (Purba & Syahputra, 2021). Metode ini juga telah digunakan dalam berbagai penelitian bidang kesehatan, seperti klasifikasi penyakit ginjal (A'yuniyah et al., 2022), penyakit kulit (Pohan & Chairunisah, 2024), dan penyakit jantung (Riani et al., 2019).

Proses penelitian ini dirancang dan dilaksanakan secara sistematis melalui tahapan-tahapan berurutan, dimulai dari pengolahan data hingga evaluasi performa sistem klasifikasi. Bagan alir proses pelaksanaan metode penelitian ditampilkan pada Gambar 1 berikut:



Gambar 1 Alur Penelitian

1. Pembacaan dan Penyiapan Dataset
Data yang digunakan berasal dari platform Kaggle, berjudul Disease Symptoms and Patient Profile Dataset (Kaggle, 2024). Dataset ini berisi 349 data pasien dengan atribut gejala dan demografi, yaitu: demam, batuk, kelelahan, kesulitan bernapas, dan kelompok usia. Pra-pemrosesan dilakukan melalui seleksi atribut, penghapusan duplikasi, dan pengelompokan umur pasien.
2. Pra-pemrosesan (Pembersihan dan Transformasi Data)

Atribut gejala yang semula berbentuk nilai kategorikal seperti 'Yes' dan 'No' digunakan secara langsung dalam proses klasifikasi. Nilai-nilai ini kemudian diinterpretasikan sebagai logika biner dalam perhitungan probabilitas *Naive Bayes*, sedangkan atribut usia dikategorikan ke dalam kelompok umur tertentu agar lebih sesuai dalam proses klasifikasi.

3. Ekstraksi dan Seleksi Fitur
Setelah data dibersihkan dan dikodekan, dilakukan proses ekstraksi dan seleksi fitur untuk menentukan atribut mana yang paling relevan terhadap hasil klasifikasi. Dalam penelitian ini, seluruh atribut dalam dataset digunakan karena telah disesuaikan dengan kebutuhan klasifikasi berbasis gejala. Pemilihan fitur dilakukan berdasarkan distribusi frekuensi dan relevansi medisnya terhadap jenis penyakit.
4. Perhitungan *Naive Bayes*
Setelah data dipersiapkan, dilakukan klasifikasi menggunakan algoritma *Naive Bayes*. Rumus dasar yang digunakan adalah:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan:

 - $P(H|X)$: Probabilitas suatu penyakit H jika diketahui atribut X
 - $P(X|H)$: Probabilitas atribut X dalam kelas H
 - $P(H)$: Probabilitas awal dari kelas penyakit
 - $P(X)$: Probabilitas keseluruhan data
5. Klasifikasi
Setelah probabilitas masing-masing kelas penyakit dihitung, sistem akan memilih kelas dengan nilai probabilitas tertinggi sebagai hasil prediksi. Hasil klasifikasi ditampilkan kepada pengguna setelah mereka mengisi form data gejala dan profil pasien.
6. Evaluasi Performa

Evaluasi dilakukan dengan cara menguji sistem menggunakan 20 data uji secara acak dari dataset utama. Performa sistem diukur menggunakan 4 metrik klasifikasi yaitu akurasi, presisi, *recall*, dan *F1-score*. Evaluasi dilakukan secara manual berdasarkan hasil klasifikasi dari sistem dan label penyakit sebenarnya pada data uji. Penghitungan metrik didasarkan pada *confusion matrix* sederhana, mengacu pada pendekatan dari Putro et al. (2020).

7. Implementasi dan Pengujian Sistem
Aplikasi dibangun berbasis web menggunakan PHP dan MySQL. Desain antarmuka dikembangkan dengan *Visual Studio Code*, dan seluruh fungsi berjalan di atas server lokal (*localhost*) menggunakan XAMPP. Pengguna dapat mengakses form input gejala dan data pasien untuk memperoleh hasil prediksi penyakit dari sistem secara otomatis.

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan secara bertahap untuk merancang dan membangun sistem klasifikasi penyakit berbasis web menggunakan algoritma *Naive Bayes*. Proses dimulai dengan perhitungan klasifikasi secara manual, dilanjutkan dengan evaluasi hasil prediksi, serta diiringi oleh perancangan dan implementasi sistem yang sesuai dengan kebutuhan pengguna.

Perhitungan *Naive Bayes*

Metode *Naive Bayes* merupakan metode klasifikasi yang menggunakan pendekatan probabilistik berdasarkan Teorema Bayes. Tujuan utamanya adalah menghitung kemungkinan suatu data termasuk ke dalam salah satu dari beberapa kelas (dalam hal ini, kategori penyakit) berdasarkan atribut atau fitur yang dimilikinya.

Sebagai contoh, berikut ditampilkan salah satu proses perhitungan klasifikasi *Naive Bayes* secara manual berdasarkan data uji ke-1.

Data Uji 1:

Demam= yes, Batuk= Yes, Kelelahan= No, Sesak Napas= Yes, Kelompok Umur= Young

Tujuan dari proses ini adalah menghitung probabilitas pasien tersebut ke dalam masing-masing kelas penyakit. Perhitungan:

1. Probabilitas Kelas (*Prior*)

$$P(H) = \frac{\text{Jumlah data pada kelas } H}{\text{Total seluruh data}} \quad (2)$$

Contoh perhitungan untuk kelas “Penyakit Pernafasan”

$$P(\text{Pernapasan}) = \frac{70}{349} \approx 0,20057$$

2. Probabilitas Fitur (*Likelihood*)

$$P(X_i|H) = \frac{\text{Jlh data dalam kelas } H \text{ yang memiliki nilai } X_i}{\text{Jumlah total data dalam kelas } H} \quad (3)$$

- $P(\text{Demam}=\text{yes}|H) = 50/70 \approx 0,71429$
- $P(\text{Batuk}=\text{yes}|H) = 50/70 \approx 0,71429$
- $P(\text{Kelelahan}=\text{no}|H) = 21/70 \approx 0,3$
- $P(\text{Sesak Nafas}=\text{yes}|H) = 52/70 \approx 0,74286$
- $P(\text{umur}=\text{young}|H) = 18/70 \approx 0,25714$

3. Probabilitas *Posterior*

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (4)$$

Maka untuk menghitung nilai probabilitas *posterior* penyakit pernafasan adalah:

$$\begin{aligned} & P(\text{Pernafasan}) \\ & P(\text{demam}|\text{pernafasan}) * \\ & P(\text{batuk}|\text{pernafasan}) * \\ & P(\text{kelelahan}|\text{pernafasan}) * \\ & P(\text{sesak nafas}|\text{pernafasan}) * \\ & P(\text{umur}|\text{pernafasan}) * \\ & = \frac{P(\text{pernafasan})}{P(X)} \end{aligned}$$

Karena (X) merupakan konstanta yang sama untuk seluruh kelas penyakit, maka

dalam implementasi praktis cukup dilakukan perbandingan nilai dari $P(X|H)*P(H)$. Oleh karena itu, bentuk penyederhanaannya adalah:

$$P(H|X) \propto P(X|H).P(H) \quad (5)$$

Di mana hasil perhitungan probabilitas posterior tidak perlu dibagi oleh (X), cukup dibandingkan secara relatif antar kelas untuk menentukan prediksi dengan nilai tertinggi.

$$\begin{aligned} P(\text{Pernapasan}|X) & \propto 0,71429 * \\ & 0,71429 * 0,3 * 0,74286 * 0,25714 \approx \\ & 0,00586 \end{aligned}$$

Berikut tabel rekap perbandingan nilai probabilitas *posterior* tiap kelas penyakit

Tabel 1 Nilai Probabilitas *Posterior* Tiap Kelas Penyakit

Kelas Penyakit	Probabilitas <i>Posterior</i>
Penyakit Pernapasan	0,00586
Penyakit Kulit	0
Endokrin & Metabolik	0,00002
Gangguan Psikologis	0,00006
Saluran Pencernaan	0,00033
Autoimun & Sendi	0,00003
Kardiovaskular	0,00005
Ginjal dan Saluran Kemih	0,00023
Infeksi & Tropis	0,00014
Hati & Sistem Hepatik	0,00004
Gangguan Saraf	0,00002
Kanker & Tumor	0
Lainnya	0

Dari tabel ini, dapat dilihat bahwa hasil probabilitas *posterior* tertinggi yaitu berada pada kelas penyakit pernafasan, maka hasil klasifikasi penyakit adalah penyakit pernafasan.

Evaluasi Sistem

Setelah seluruh data uji diproses dengan metode *Naive Bayes*, hasil prediksi dibandingkan dengan label penyakit sebenarnya dari dataset. Sistem berhasil mengklasifikasikan 9 dari 20 data uji dengan benar. Dari 20 data uji

yang diambil secara acak, adapun perolehan nilainya yaitu:

Prediksi benar= 9

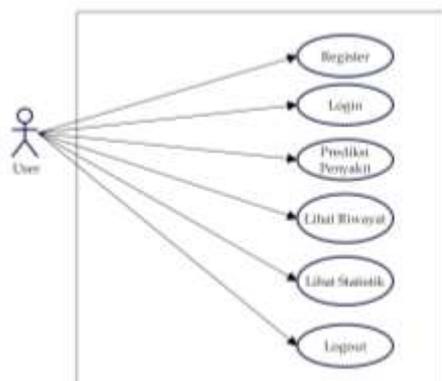
Prediksi salah= 11

- Akurasi= $9/20 = 0,45$
- Precision= $9/20 = 0,45$
- Recall= $9/20 = 0,45$
- F1-score= $2 \times (0.45 \times 0.45) / (0.45 + 0.45) = 2 \times 0.2025 / 0.9 \approx 0.45$

Meskipun performa sistem belum mencapai tingkat akurasi yang tinggi, hasil ini tetap menunjukkan bahwa sistem memiliki potensi dalam melakukan klasifikasi penyakit awal. Nilai metrik yang seimbang antara precision dan recall mencerminkan bahwa sistem cukup konsisten dalam melakukan prediksi benar terhadap data uji, walaupun masih terbatas.

Rancangan Sistem

Rancangan sistem disusun untuk menggambarkan hubungan antara pengguna dan fungsi sistem. Gambar 2 menunjukkan diagram use case yang menggambarkan interaksi antara pengguna (*user*) dan sistem klasifikasi penyakit.



Gambar 2 Use Case Diagram

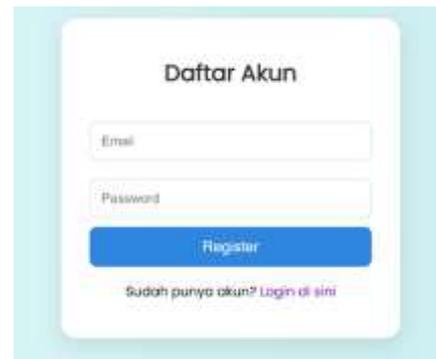
Implementasi Sistem

Implementasi sistem dilakukan dalam bentuk aplikasi web yang dapat diakses oleh pengguna melalui browser. Sistem ini memiliki beberapa fitur utama yang ditujukan untuk memfasilitasi proses pendaftaran, login, input data gejala dan profil pasien, prediksi penyakit, serta menampilkan riwayat dan statistik hasil prediksi.

Sistem terdiri dari beberapa antarmuka utama yang mempermudah pengguna dalam mengakses dan menjalankan proses klasifikasi penyakit:

Tampilan Halaman Register

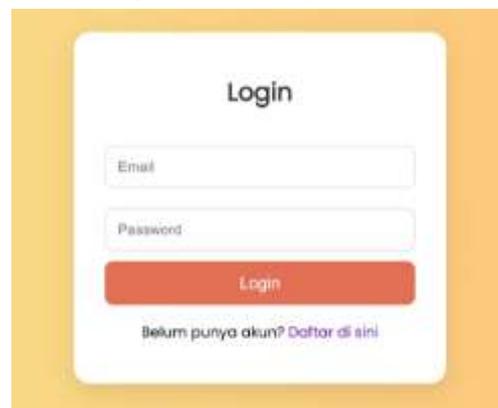
Tampilan ini digunakan oleh pengguna baru untuk melakukan pendaftaran akun agar dapat mengakses sistem. Pengguna diminta untuk mengisi *email* dan *password*.



Gambar 3 Halaman Register

Tampilan Halaman Login

Setelah mendaftar, pengguna dapat melakukan *login* untuk masuk ke dalam sistem. Pengguna perlu memasukkan *email* dan *password* yang telah terdaftar.



Gambar 4 Halaman Login

Tampilan Form Prediksi Penyakit

Setelah berhasil *login*, pengguna diarahkan ke halaman utama yaitu form prediksi. Di halaman ini, pengguna dapat memilih gejala yang dialami dan kelompok umur.



Gambar 5 Form Prediksi Penyakit

Tampilan Hasil Prediksi

Setelah mengisi data dan menekan tombol prediksi, sistem akan menampilkan hasil prediksi penyakit berdasarkan input yang diberikan. Selain itu, sistem juga memberikan saran penanganan awal.



Gambar 6 Tampilan Hasil Prediksi

Tampilan Riwayat Prediksi

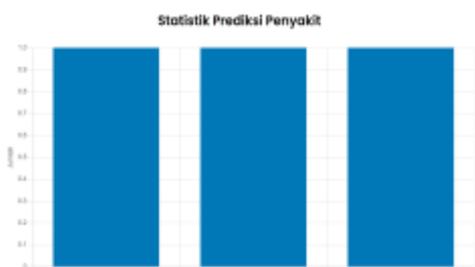
Pengguna dapat melihat riwayat prediksi yang telah dilakukan sebelumnya melalui menu "Riwayat". Riwayat ditampilkan dalam bentuk tabel.



No	Tanggal	Waktu	Lokasi	Status Riwayat	Riwayat Riwayat	Riwayat Riwayat	Riwayat Riwayat
1	2025-08-20	10:00	RS	Selesai	Penyakit Pernapasan	2025-08-20 10:00	2025-08-20 10:00
2	2025-08-20	10:05	RS	Selesai	Penyakit Pernapasan	2025-08-20 10:05	2025-08-20 10:05
3	2025-08-20	10:10	RS	Selesai	Penyakit Pernapasan	2025-08-20 10:10	2025-08-20 10:10
4	2025-08-20	10:15	RS	Selesai	Penyakit Pernapasan	2025-08-20 10:15	2025-08-20 10:15
5	2025-08-20	10:20	RS	Selesai	Penyakit Pernapasan	2025-08-20 10:20	2025-08-20 10:20

Gambar 7 Tampilan Riwayat Prediksi

Sistem juga menyediakan menu "Statistik" yang menampilkan visualisasi jumlah prediksi penyakit dalam bentuk grafik batang.



Gambar 8 Tampilan Statistik Prediksi

Tampilan Logout

Logout merupakan fitur untuk keluar dari sistem dan mengakhiri sesi pengguna.



Gambar 9 Tampilan Halaman Logout

SIMPULAN

Penelitian ini menunjukkan bahwa algoritma *Naive Bayes* dapat diimplementasikan secara efektif dalam pengembangan sistem klasifikasi penyakit berbasis web dengan memanfaatkan kombinasi data gejala dan profil pasien. Meskipun performa klasifikasi belum mencapai tingkat akurasi tinggi, pendekatan ini tetap menunjukkan kemampuan dalam menghasilkan prediksi secara konsisten pada kondisi data yang terbatas. Kontribusi utama dari penelitian ini terletak pada integrasi metode probabilistik ke dalam platform digital yang mudah diakses dan responsif, menjadikannya relevan sebagai alat bantu diagnosis awal. Temuan ini memperkuat posisi *Naive Bayes* sebagai metode yang efisien untuk klasifikasi cepat dalam sistem pendukung keputusan medis, khususnya saat fitur gejala dan demografi masih terbatas. Penelitian ini turut memperluas literatur mengenai penerapan metode statistik dalam sistem informasi kesehatan berbasis web yang bersifat praktis dan terotomatisasi.

DAFTAR PUSTAKA

- A'yuniyah, Q., Tasia, E., Nazira, N., Pratama, P. F., Anugrah, M. R., Adhiva, J., & Mustakim, M. (2022). Implementasi algoritma Naïve Bayes Classifier (NBC) untuk klasifikasi penyakit ginjal kronik. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(1), 72. <https://doi.org/10.30865/json.v4i1.4781>
- Anggraini, S., Akbar, M., Wijaya, A., Syaputra, H., & Sobri, M. (2021). Klasifikasi gejala penyakit Coronavirus Disease 19 (COVID-19) menggunakan machine learning. *Journal of Software Engineering Ampera*, 2(1), 57–68. <https://doi.org/10.51519/journalsea.v2i1.105>
- Guines, A., Hutapea, A. N., & Kelmaskosu, R. M. (2022). Gambaran karakteristik klinis penderita Covid-19 di satu rumah sakit rujukan Covid-19 di Tangerang: Studi dokumentasi. *Journal of Bionursing*, 4(2), 92–103. <https://doi.org/10.20884/1.bion.2022.4.2.138>
- Kaggle. (2024). Disease Symptoms and Patient Profile Dataset. <https://www.kaggle.com/datasets/uom190346a/disease-symptoms-and-patient-profile-dataset>
- Pohan, K. P., & Chairunisah, C. (2024). Sistem pakar mendiagnosa penyakit kulit pada manusia menggunakan metode Naive Bayes Classifier berbasis web. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, 23(1), 204. <https://doi.org/10.53513/jis.v23i1.9521>
- Purba, B., & Syahputra, R. (2021). Implementasi metode Naive Bayes Classifier pada evaluasi kepuasan mahasiswa terhadap pembelajaran daring. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(1), 85–91. <https://doi.org/10.30743/infotekjar.v6i1.4352>
- Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan metode Naive Bayes untuk klasifikasi pelanggan. *Jurnal Teknologi Informasi dan Komunikasi (TIKOMSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.500>
- Riani, A., Susianto, Y., Rahman, N., & Ali, U. D. (2019). Implementasi data mining untuk memprediksi penyakit jantung menggunakan metode Naive Bayes. *JINITA: Jurnal Informatika dan Teknologi Informasi*, 1(01), 25–34. <https://doi.org/10.35970/jinita.v1i01.64>