**JURNAL SISTEM INFORMASI TGD**
**Volume 5, Nomor 6, November 2026, Hal 1432-1443**
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*

# Implementasi Naive Bayes Dan K-Nearest Neighbors Untuk Diagnosis Dispepsia Akibat Infeksi Helicobacter Pylori

**Ray Charlia[1], Nurcahyo Budi Nugroho[2], Dedi Setiawan[3]**
[1,2,3] Sistem Informasi, Stmik Triguna Dharma
Email: [1]rayysin0309@gmail.com, [2]nurcahyobn@gmail.com, [3]setiawandedi07@gmail.com
Email Penulis Korespondensi: rayysin0309@gmail.com

**Abstrak**
Penyakit dispepsia akibat infeksi Helicobacter pylori merupakan salah satu gangguan saluran pencernaan yang sering dijumpai dan dapat berkembang menjadi komplikasi serius apabila tidak didiagnosis secara dini. Keterbatasan tenaga medis, lamanya proses pemeriksaan, serta tingginya biaya diagnosis konvensional menjadi kendala dalam penanganan penyakit ini. Oleh karena itu, diperlukan suatu sistem yang mampu membantu proses diagnosis awal secara cepat dan akurat. Penelitian ini bertujuan mengimplementasikan metode Naive Bayes dan K-Nearest Neighbors (KNN) untuk mendiagnosis penyakit dispepsia akibat infeksi Helicobacter pylori berdasarkan gejala yang dialami pasien. Tahapan penelitian meliputi pengumpulan dan preprocessing data, pembentukan model klasifikasi menggunakan kedua metode, pengujian model, serta evaluasi kinerja menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa kedua metode mampu mengklasifikasikan kondisi pasien berdasarkan data gejala dengan tingkat performa yang baik, namun terdapat perbedaan nilai pada setiap metrik evaluasi yang digunakan. Perbandingan hasil evaluasi digunakan untuk menentukan metode yang memberikan performa terbaik dalam proses diagnosis. Implementasi kedua algoritma ini diharapkan dapat menjadi alternatif pendukung keputusan bagi tenaga medis dalam melakukan diagnosis awal dispepsia akibat infeksi Helicobacter pylori, sekaligus meningkatkan efisiensi pelayanan kesehatan melalui proses identifikasi penyakit yang lebih cepat, objektif, dan akurat.
**Kata kunci:** Dispepsia, Helicobacter pylori, Naive Bayes, K-Nearest Neighbors, Klasifikasi.

***Abstract***
*Dyspepsia associated with Helicobacter pylori infection is one of the most common gastrointestinal disorders and may progress to serious complications if not diagnosed at an early stage. Limitations in the availability of medical professionals, lengthy examination procedures, and the high cost of conventional diagnostic methods remain significant challenges in managing this condition. Therefore, a system capable of supporting rapid and accurate preliminary diagnosis is required. This study aims to implement the Naive Bayes and K-Nearest Neighbors (KNN) methods to diagnose dyspepsia associated with Helicobacter pylori infection based on patients' clinical symptoms. The research stages include data collection, data preprocessing, classification model development using both algorithms, model testing, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The experimental results indicate that both methods are capable of classifying patients' conditions based on symptom data with satisfactory performance. However, differences were observed in the evaluation metrics achieved by each algorithm. A comparative analysis of the evaluation results was conducted to determine the method that provides the best performance for the diagnostic process. The implementation of these two algorithms is expected to serve as a decision support tool for healthcare professionals in performing the preliminary diagnosis of dyspepsia associated with Helicobacter pylori infection, while also improving healthcare service efficiency through a faster, more objective, and more accurate disease identification process.*
***Keywords****: Dyspepsia, Helicobacter pylori, Naive Bayes, K-Nearest Neighbors, Classification..*

## 1. PENDAHULUAN

Penyakit saluran pencernaan merupakan salah satu masalah kesehatan global yang berdampak signifikan terhadap kualitas hidup masyarakat di berbagai negara. Kondisi seperti dispepsia sering muncul dari gangguan pada lambung dan saluran cerna atas, dan menjadi penyebab umum kunjungan medis di klinik maupun rumah sakit. Infeksi oleh bakteri *Helicobacter pylori* (*H. pylori*) merupakan salah satu faktor utama penyebab gangguan tersebut, yang jika tidak terdiagnosis dan ditangani sejak dini dapat berujung pada komplikasi serius seperti tukak lambung dan kanker lambung. Pendekatan deteksi otomatis berdasarkan data klinis semakin dipandang sebagai solusi keterbatasan tenaga medis dalam diagnosis awal penyakit pencernaan [1].

Namun demikian, permasalahan nyata yang sering terjadi di lapangan adalah ketidakakuratan dan keterlambatan diagnosis, terutama di fasilitas kesehatan primer. Banyak pasien dengan gejala dispepsia tidak segera mendapatkan diagnosis yang tepat karena keterbatasan waktu konsultasi dokter, minimnya tenaga spesialis, serta ketidaktahuan masyarakat mengenai gejala infeksi *H. pylori.* Situasi ini diperburuk oleh fakta bahwa proses diagnosis tradisional sering kali memerlukan pemeriksaan laboratorium dan endoskopi yang memakan waktu dan biaya tinggi. Kondisi-kondisi tersebut dapat menyebabkan penanganan yang terlambat dan risiko komplikasi meningkat. Untuk menjawab tantangan tersebut, sistem pakar berbasis kecerdasan buatan dan pembelajaran mesin telah banyak dikembangkan untuk membantu diagnosis dini penyakit berbagai jenis. Dalam konteks sistem pakar diagnosa penyakit pencernaan, misalnya penelitian di Indonesia menunjukkan bahwa penerapan metode *Naive Bayes* dapat digunakan untuk mendiagnosis sindrom dispepsia berdasarkan kumpulan data gejala pasien, memberikan diagnosis yang sesuai dengan pakar medis mencapai tingkat akurasi yang memadai [2].

Dari sekian banyak metode pembelajaran mesin, metode *Naive Bayes* dan *K-Nearest Neighbors* (KNN) sering dipilih karena keunggulannya dalam klasifikasi data medis. *Naive Bayes* merupakan pendekatan probabilistik yang sederhana

namun efektif untuk masalah klasifikasi dengan asumsi fitur yang independen. Sementara itu, KNN bekerja berdasarkan kemiripan data dengan cara mengukur jarak antara data uji dan data latih di ruang fitur, sehingga mampu menangkap hubungan non-linier pada data gejala penyakit. Kedua metode ini terbukti memiliki performa baik dalam penelitian klasifikasi penyakit lainnya, seperti pada klasifikasi penyakit ginjal kronik menggunakan kedua algoritma tersebut dalam satu studi yang membandingkan performanya [3].

Berdasarkan uraian di atas, penelitian ini difokuskan pada implementasi metode *Naive Bayes* dan *K-Nearest Neighbors* untuk diagnosis penyakit dispepsia akibat infeksi *Helicobacter pylori*. Target utama penelitian adalah menghasilkan model klasifikasi yang mampu membedakan status infeksi *H. pylori* dengan akurasi tinggi, sehingga dapat menjadi alat bantu diagnosis dalam sistem pakar kesehatan. Diharapkan penelitian ini tidak hanya memberikan rekomendasi metode terbaik antara kedua algoritma, tetapi juga menyediakan solusi teknologi informasi yang dapat membantu tenaga medis dan masyarakat dalam proses diagnosis lebih cepat dan efisien.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Metode penelitian merupakan rangkaian prosedur terstruktur yang menggunakan pendekatan ilmiah untuk memperoleh data yang valid dan relevan dalam menyelesaikan permasalahan. Dalam penelitian ini, metode yang digunakan adalah penelitian terapan (*applied research*) yang berfokus pada pengembangan sistem pakar untuk diagnosis penyakit dispepsia akibat infeksi *Helicobacter Pylori* menggunakan algoritma Naive Bayes dan K-Nearest Neighbors (KNN). Pendekatan ini dipilih karena bertujuan menghasilkan solusi praktis yang dapat langsung diimplementasikan dalam dunia kesehatan, khususnya untuk membantu tenaga medis dalam melakukan diagnosis secara lebih objektif dan konsisten.

Proses penelitian diawali dengan identifikasi masalah terkait penyakit dispepsia, dilanjutkan dengan studi literatur untuk memperoleh dasar teori yang relevan, serta pengumpulan data melalui metode seperti wawancara dengan dokter spesialis, observasi, dan pengambilan data historis pasien. Penelitian ini juga menggabungkan pendekatan kualitatif dan kuantitatif (*mixed methods*) untuk menghasilkan analisis yang lebih komprehensif dalam membandingkan akurasi antara metode Naive Bayes dan KNN.

Secara keseluruhan, metode penelitian ini disusun secara sistematis mulai dari identifikasi masalah hingga evaluasi hasil, sehingga diharapkan mampu menghasilkan sistem yang akurat, efisien, dan bermanfaat dalam membantu proses diagnosis penyakit dispepsia akibat infeksi Helicobacter Pylori.

### 2.2 *Data Collecting* (Teknik Pengumpulan Data)

Teknik pengumpulan data dalam penelitian ini memuat uraian mengenai karakteristik, kondisi, serta aktivitas yang berkaitan langsung dengan objek penelitian, yaitu proses diagnosis penyakit dispepsia akibat infeksi Helicobacter Pylori. Data dihimpun di Rumah Sakit Umum Hidayah yang berlokasi di Kelurahan Suka Makmur, Kec. Deli Tua, Kabupaten Deli Serdang dan difokuskan pada pengumpulan informasi yang mendukung pengembangan sistem pakar berbasis algoritma Naive Bayes dan KNN. Pengumpulan data dilakukan melalui beberapa metode utama sebagai berikut::

a. Observasi
   Metode pengumpulan data dilakukan melalui observasi lapangan dengan melakukan kunjungan langsung ke Rumah Sakit Umum Hidayah sebagai fasilitas pelayanan kesehatan yang menangani pasien dispepsia. Teknik ini bertujuan untuk memperoleh gambaran nyata mengenai proses konsultasi dan diagnosis penyakit dispepsia yang dilakukan oleh tenaga kesehatan. Melalui observasi, peneliti dapat memahami alur pelayanan mulai dari pencatatan data pasien, pemeriksaan gejala klinis, hingga penentuan diagnosa infeksi Helicobacter Pylori. Selain itu, peneliti juga mengamati penggunaan sistem pencatatan data pasien serta ketersediaan fasilitas pendukung rekam medis. Hasil observasi ini memberikan gambaran kondisi aktual di lapangan sehingga dapat dijadikan dasar dalam merancang sistem pakar berbasis Web yang sesuai dengan kebutuhan pengguna serta mampu meningkatkan efisiensi dan konsistensi dalam proses diagnosis..
b. Wawancara
   Pengumpulan data juga dilakukan melalui wawancara dengan tenaga ahli di bidang kesehatan, khususnya dokter di Rumah Sakit Umum Hidayah yaitu ibu dr. Issumi Maharani Tanjung. Wawancara dilakukan secara terstruktur untuk menggali informasi terkait gejala-gejala spesifik dispepsia akibat Helicobacter Pylori dan proses diagnosis yang dilakukan secara medis. Hasil wawancara ini sangat penting sebagai basis kasus (*knowledge base*) dalam memastikan bahwa sistem pakar yang dikembangkan sesuai dengan praktik medis yang sebenarnya serta mampu memberikan hasil diagnosis yang akurat dan dapat dipercaya.
c. Data Historis
   Selain observasi dan wawancara, penelitian ini juga menggunakan data historis berupa data gejala dan diagnosis pasien dispepsia yang diperoleh dari Rumah Sakit Umum Hidayah. Data ini digunakan sebagai *dataset training* dalam proses perhitungan algoritma Naive Bayes dan K-Nearest Neighbors (KNN). Data historis

JURNAL SISTEM INFORMASI TGD
Volume 5, Nomor 6, November 2026, Hal 1432-1443
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*

mencakup atribut-atribut gejala klinis serta hasil diagnosis akhir yang telah ditentukan oleh tenaga ahli. Dataset ini menjadi dasar dalam menghitung probabilitas pada Naive Bayes serta penentuan kedekatan jarak pada KNN, sehingga berperan penting dalam membandingkan tingkat akurasi kedua metode tersebut.

Tabel 1 Rekapitulasi Distribusi Penyakit pada Dataset

| Kode | Kategori Penyakit | Jumlah Kasus (N) | Persentase (%) |
|---|---|---|---|
| P01 | Dispepsia Akibat Infeksi H. Pylori | 15 | 25.0% |
| P02 | Dispepsia Fungsional | 15 | 25.0% |
| P03 | Gastritis Akibat H. Pylori | 10 | 16.7% |
| P04 | Tukak Lambung (Peptic Ulcer) | 10 | 16.7% |
| P05 | GERD (Gastroesophageal Reflux Disease) | 10 | 16.7% |
| **Total Dataset (Data Latih)** | | **60** | **100%** |

Tahap selanjutnya adalah menghitung frekuensi kemunculan setiap gejala (G) pada setiap kategori penyakit (P). Nilai frekuensi ini kemudian ditransformasikan menjadi nilai probabilitas *likelihood*. Untuk menghindari nilai probabilitas nol yang dapat membatalkan hasil perkalian, diterapkan teknik *Laplace Smoothing*. *Laplace Smoothing* (atau *Add-one Smoothing*) dilakukan dengan menambahkan nilai konstanta (k) pada pembilang dan menambahkan (k × jumlah kategori) pada penyebut. Rumus yang digunakan adalah:

$$P(Gi|Pj) = \frac{(count(Gi,Pj) + k)}{(count(Pj) + k * 2)} \quad (3.1)$$

Di mana *k* adalah parameter smoothing (dalam sistem ini menggunakan k=0.01) dan 2 merepresentasikan dua kemungkinan status gejala (Ya/Tidak). Perhitungan manual untuk gejala G01 pada penyakit P01 adalah sebagai berikut:

$$P(Gi|Pj) = \frac{(14 + 0.01)}{(\mathbf{15 + 0.02})} = 0.932 = 0.93 \quad (3.1)$$

Berikut adalah tabel hasil ekstraksi frekuensi dan probabilitas gejala yang telah diproses menggunakan *smoothing*:

Tabel 2 Analisis Frekuensi Gejala per Penyakit (Ekstraksi Probabilitas)

| Kode Gejala | P01 (N=15) | | P02 (N=15) | | P03 (N=10) | | P04 (N=10) | | P05 (N=10) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Freq | Prob | Freq | Prob | Freq | Prob | Freq | Prob | Freq | Prob |
| G01 | 14 | 0.93 | 15 | 1.00 | 10 | 1.00 | 10 | 1.00 | 2 | 0.20 |
| G02 | 13 | 0.87 | 0 | 0.00 | 10 | 1.00 | 10 | 1.00 | 0 | 0.00 |
| G03 | 10 | 0.67 | 5 | 0.33 | 10 | 1.00 | 10 | 1.00 | 0 | 0.00 |
| G04 | 6 | 0.40 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| G05 | 6 | 0.40 | 10 | 0.67 | 0 | 0.00 | 0 | 0.00 | 1 | 0.10 |
| G06 | 3 | 0.20 | 15 | 1.00 | 0 | 0.00 | 0 | 0.00 | 1 | 0.10 |
| G07 | 5 | 0.33 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| G08 | 1 | 0.07 | 10 | 0.67 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| G09 | 1 | 0.07 | 1 | 0.07 | 0 | 0.00 | 0 | 0.00 | 10 | 1.00 |
| G10 | 0 | 0.00 | 0 | 0.00 | 4 | 0.40 | 4 | 0.40 | 0 | 0.00 |
| G11 | 0 | 0.00 | 0 | 0.00 | 4 | 0.40 | 4 | 0.40 | 0 | 0.00 |
| G12 | 6 | 0.40 | 0 | 0.00 | 4 | 0.40 | 3 | 0.30 | 0 | 0.00 |
| G13 | 1 | 0.07 | 0 | 0.00 | 1 | 0.10 | 0 | 0.00 | 3 | 0.30 |
| G14 | 3 | 0.20 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| G15 | 0 | 0.00 | 14 | 0.93 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| G16 | 8 | 0.53 | 0 | 0.00 | 9 | 0.90 | 10 | 1.00 | 0 | 0.00 |
| G17 | 0 | 0.00 | 4 | 0.27 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| G18 | 2 | 0.13 | 0 | 0.00 | 3 | 0.30 | 0 | 0.00 | 10 | 1.00 |
| G19 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 10 | 1.00 |
| G20 | 1 | 0.07 | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 | 10 | 1.00 |
| G21 | 1 | 0.07 | 0 | 0.00 | 0 | 0.00 | 6 | 0.60 | 7 | 0.60 |
| G22 | 5 | 0.33 | 8 | 0.53 | 3 | 0.30 | 3 | 0.30 | 5 | 0.50 |

### 2.3 Penyakit Dispepsia

Dispepsia merupakan gangguan pada saluran pencernaan bagian atas yang ditandai dengan rasa tidak nyaman atau nyeri pada ulu hati. Kondisi ini sering disebut sebagai "maag" oleh masyarakat. Dispepsia bukan merupakan satu penyakit tunggal, melainkan kumpulan gejala yang dapat disebabkan oleh berbagai faktor seperti pola makan yang tidak teratur, stres, penggunaan obat antiinflamasi nonsteroid (OAINS), serta infeksi bakteri. Dispepsia merupakan sindrom atau kumpulan gejala yang sering dijumpai, ditandai dengan adanya nyeri atau tidak nyaman pada perut bagian atas atau ulu hati [4].

JURNAL SISTEM INFORMASI TGD
Volume 5, Nomor 6, November 2026, Hal 1432-1443
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*

Secara klinis, dispepsia dibagi menjadi dua jenis, yaitu dispepsia fungsional dan dispepsia organik. Dispepsia terjadi tanpa adanya kelainan struktural yang jelas pada organ pencernaan, sedangkan dispepsia organik disebabkan oleh gangguan nyata seperti gastritis, tukak lambung, atau infeksi bakteri tertentu. Salah satu penyebab utama dispepsia organik adalah infeksi bakteri *Helicobacter pylori* yang menyerang lapisan mukosa lambung. Diagnosis dispepsia umumnya dilakukan melalui wawancara medis terkait gejala yang dialami pasien, pemeriksaan fisik, serta pemeriksaan penunjang seperti tes laboratorium atau endoskopi [5].

### 2.4 Infeksi Bakteri

Infeksi bakteri merupakan kondisi ketika mikroorganisme patogen masuk dan berkembang biak di dalam tubuh manusia sehingga menimbulkan gangguan

kesehatan. Bakteri dapat menyerang berbagai organ tubuh, termasuk saluran pencernaan, sistem pernapasan, kulit, dan organ lainnya.

Pada sistem pencernaan, infeksi bakteri dapat menyebabkan peradangan pada lambung dan usus. Proses infeksi biasanya terjadi karena konsumsi makanan atau minuman yang terkontaminasi, kebersihan yang kurang baik, atau daya tahan tubuh yang menurun. Ketika bakteri masuk ke dalam tubuh, sistem imun akan bereaksi untuk melawan, namun dalam beberapa kondisi bakteri dapat bertahan dan menyebabkan peradangan kronis.

Salah satu bakteri yang paling dikenal dalam menyebabkan gangguan lambung adalah *Helicobacter pylori*, yang memiliki kemampuan bertahan dalam lingkungan asam lambung.

### 2.5 Helicobacter pylori

*Helicobacter pylori* adalah bakteri gram negatif berbentuk spiral yang hidup di lapisan mukosa lambung manusia. Bakteri ini memiliki kemampuan unik untuk bertahan dalam lingkungan asam lambung dengan cara menghasilkan enzim urease, yang dapat mengubah urea menjadi amonia sehingga menetralkan asam di sekitarnya.



Gambar 1 Fisik *Helicobacter pylori*

Infeksi *Helicobacter pylori* dapat menyebabkan peradangan kronis pada lambung (gastritis), tukak lambung, hingga meningkatkan risiko kanker lambung jika tidak ditangani dengan baik. Penularan bakteri ini umumnya terjadi melalui jalur oral-oral atau fekal-oral, terutama di lingkungan dengan sanitasi yang kurang baik.

Tidak semua individu yang terinfeksi *H. pylori* menunjukkan gejala. Namun, pada sebagian orang, infeksi ini dapat berkembang menjadi gangguan pencernaan yang serius. Oleh karena itu, deteksi dini sangat penting untuk mencegah komplikasi lebih lanjut.

### 2.6 Sistem Pakar

Sistem pakar merupakan aplikasi kecerdasan buatan yang menggabungkan basis pengetahuan dengan sistem inferensi untuk meniru cara pengambilan keputusan seorang ahli. Sistem ini dirancang untuk mentransfer pengetahuan yang dimiliki manusia ke dalam komputer, memungkinkan komputer untuk menyelesaikan masalah dengan pendekatan yang serupa dengan cara yang dilakukan oleh para pakar. Prinsip dasar dari sistem pakar adalah mentransfer pengetahuan yang dimiliki oleh seorang ahli ke dalam komputer, serta merancang sistem yang dapat membuat keputusan dan menarik kesimpulan berdasarkan pengetahuan tersebut. Keahlian manusia memiliki batasan, seperti dapat hilang karena kematian, pensiun, atau perpindahan tempat kerja. Dalam proses pengambilan keputusan, seorang pakar dapat dipengaruhi oleh berbagai faktor yang dapat memengaruhi hasil akhir keputusan yang diambil. Sebaliknya, sistem pakar memberikan hasil yang lebih konsisten dibandingkan dengan seorang ahli manusia. Selain itu, sistem ini dapat membuat keputusan dengan waktu yang lebih stabil dan dalam beberapa kasus, menghasilkan kesimpulan lebih cepat daripada seorang pakar [6].

Sistem merupakan kumpulan subsistem yang saling terhubung dan saling bergantung satu sama lain, yang bekerja secara bersama untuk mencapai tujuan dan sasaran yang telah ditentukan. Setiap sistem terdiri dari input, proses, output, dan umpan balik. Sedangkan, seorang pakar adalah individu yang memiliki pengetahuan, pengalaman, dan metode khusus, serta mampu menerapkannya untuk memecahkan masalah atau memberikan nasihat. Seorang pakar harus mampu menjelaskan dan mempelajari hal-hal baru yang relevan dengan masalah yang dihadapi, serta dapat menyusun kembali pengetahuan yang dimiliki dan memecahkan aturan-aturan yang ada untuk menentukan relevansi keahliannya.

1. Berdasarkan pendapat Sutojo (dalam *Jurnal MNEMONIC*), *expert system* adalah suatu sistem yang dirancang untuk menirukan keahlian yang dimiliki oleh seorang pakar dalam menyelesaikan atau memberikan jawaban terhadap suatu permasalahan[7].
2. Menurut Kusrini (dalam *Jurnal Core-IT: Jurnal Komputasi dan Teknologi Informasi*), *expert system* merupakan

sistem informasi yang memuat pengetahuan dari seorang pakar, sehingga dapat dimanfaatkan sebagai sarana konsultasi. Pengetahuan yang tersimpan dalam sistem tersebut menjadi dasar bagi *expert system* dalam memberikan jawaban atas pertanyaan yang diajukan [8].

3. Berdasarkan pendapat Giarratano dan Riley (dalam *Jurnal Diffusion*), *expert system* merupakan bentuk aplikasi dari *artificial intelligence* yang dirancang untuk memberikan solusi terhadap permasalahan yang kompleks [9].

Setelah keahlian-keahlian tersebut disimpan dalam basis pengetahuan dan program untuk mengakses basis data telah disiapkan, komputer perlu diprogram untuk menjalankan proses inferensi. Proses ini dilaksanakan oleh *inference engine*. Mayoritas sistem pakar komersial dibangun menggunakan sistem berbasis aturan (*rule-based systems*), di mana pengetahuan diorganisir dalam bentuk aturan-aturan yang umumnya mengikuti pola *IF-THEN*. Selain itu, kemampuan untuk memberikan rekomendasi juga merupakan fitur penting dalam sistem pakar, yang membedakannya dari sistem konvensional.

Mesin Inferensi merupakan inti dari aplikasi sistem pakar, di mana kemampuan seorang pakar disematkan dalam mesin ini. Tugas yang dilakukan oleh mesin inferensi bergantung pada pengetahuan yang terdapat dalam basis pengetahuan, yang telah diperoleh dari seorang pakar.

Pengetahuan merujuk pada informasi atau pemahaman yang dimiliki atau disadari oleh seseorang. Pengetahuan mencakup berbagai hal seperti deskripsi, hipotesis, konsep, teori, dan prinsip, meskipun tidak terbatas pada hal-hal tersebut. Pengetahuan dapat dianggap sebagai penyaringan atau inti dari informasi. Pengetahuan itu sendiri terbagi dalam beberapa kategori, yaitu:

1. *Procedural knowledge*, yang menjelaskan cara untuk melakukan sesuatu.
2. *Declarative knowledge*, yang memberikan jawaban atas pertanyaan yang bersifat benar atau salah.
3. *Tacit knowledge*, yaitu pengetahuan yang tidak dapat dijelaskan dengan bahasa.

### 2.7 *Naïve Bayes*

Naive Bayes adalah metode klasifikasi berbasis probabilistik yang sederhana, yang bekerja dengan menghitung kemungkinan dari suatu kelas berdasarkan frekuensi dan kombinasi nilai pada data yang tersedia. Algoritma ini menerapkan Teorema Bayes dan mengasumsikan bahwa setiap atribut bersifat independen atau tidak saling bergantung terhadap satu sama lain dalam hubungannya dengan variabel kelas [10].

Definisi lain menyebutkan bahwa metode Naïve Bayes adalah teknik klasifikasi yang didasarkan pada konsep probabilitas dan statistika, yang dikembangkan oleh ilmuwan asal Inggris, Thomas Bayes. Metode ini menggunakan Teorema Bayes yang dikombinasikan dengan asumsi "naive" atau sederhana, yaitu bahwa setiap atribut atau variabel dalam data dianggap saling bebas (independen).[11]. Metode ini dapat memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Salah satu keunggulan penggunaan metode Naïve Bayes dalam penelitian adalah kemampuannya untuk melakukan estimasi parameter klasifikasi hanya dengan data pelatihan yang relatif sedikit. Selain itu, metode ini seringkali menunjukkan kinerja yang sangat baik dalam berbagai situasi kompleks di dunia nyata, bahkan melebihi ekspektasi awal [12]. Teorema Naïve Bayes dirumuskan sebagai berikut :

$$\mathrm{P(c}|x) = \frac{P(x|C)P(c)}{P(x)} \quad \text{.....(2.1)}$$

Keterangan :

$x$ : Data dengan class yang belum diketahui
$c$ : Hipotesis data merupakan suatu class spesifik
$(c|x)$ : Nilai Probabilitas hipotesis C berdasarkan kondisi X
$(c)$ : Nilai Probabilitas pada hipotesis C
$(x|c)$ : Probabilitas X berdasarkan kondisi pada hipotesis C
$(x)$ : Probabilitas X

Untuk menjelaskan teorema *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut:

$$\mathrm{P(C}|F_1 \dots F_n)\frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \quad \text{.....(2.2)}$$

Pada persamaan diatas variabel C merepresentasikan kelas, sementara variabel $F1 \dots Fn$ mempresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi.

Rumus tersebut menggambarkan bahwa probabilitas suatu sampel dengan karakteristik tertentu termasuk ke dalam kelas C (disebut posterior) ditentukan dengan mengalikan probabilitas awal kelas C sebelum mempertimbangkan data sampel (disebut prior) dengan probabilitas munculnya karakteristik-karakteristik tersebut dalam kelas C (likelihood), lalu dibagi dengan probabilitas keseluruhan munculnya karakteristik-karakteristik sampel tersebut di semua kelas (evidence). Oleh karena itu, rumus tersebut dapat disederhanakan dalam bentuk sebagai berikut:

$$Posterior = \frac{Prior\ x\ likelihood}{evidence} \quad \text{.....(2.3)}$$

JURNAL SISTEM INFORMASI TGD
**Volume 5, Nomor 6, November 2026, Hal 1432-1443**
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*

Nilai *evidence* pada rumus di atas bersifat konstan untuk semua kelas dalam satu sampel. Oleh karena itu, nilai *posterior* yang telah dihitung akan dibandingkan dengan nilai *posterior* dari kelas-kelas lainnya untuk menentukan ke kelas mana sampel tersebut akan diklasifikasikan. Adapun langkah-langkah dalam algoritma Naïve Bayes adalah sebagai berikut:

1. Siapkan dataset
2. Hitung Jumlah dan probabilitas.
3. Cari nilai *mean* dan *standar deviasi* dari masing-masing *parameter* yang merupakan data *numerik*
4. Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
5. Mendapatkan nilai dalam tabel *mean*, *standart deviasi* dan probabilitas.
6. Mendapatkan Solusi.

### 2.8 *K-Nearest Neightbor*

*K-Nearest Neighbor* (KNN) merupakan salah satu metode klasifikasi dalam data mining dan machine learning yang bekerja berdasarkan prinsip kedekatan jarak antar data. Metode ini mengklasifikasikan suatu objek baru dengan melihat sejumlah *k* data terdekat (nearest neighbors) dalam data latih, kemudian menentukan kelas berdasarkan mayoritas kelas dari tetangga terdekat tersebut. KNN termasuk dalam kategori *lazy learning algorithm* karena tidak membangun model secara eksplisit pada tahap pelatihan, melainkan melakukan proses komputasi saat proses klasifikasi dilakukan [13].

Menurut penelitian terbaru, KNN merupakan metode klasifikasi non-parametrik yang sederhana namun efektif dalam menangani berbagai permasalahan klasifikasi, terutama ketika distribusi data tidak diketahui secara pasti. Keunggulan utama metode ini adalah kemudahan implementasi dan kemampuan adaptif terhadap pola data yang kompleks.

Pemodelan *K-Nearest Neighbor* (KNN) menggunakan metrik jarak Euclidean untuk menentukan jarak antara data baru dan data yang sudah ada di dalam kumpulan data pelatihan, hal ini memungkinkannya untuk mengklasifikasikan data menurut mayoritas kategori tetangga terdekatnya [13]. Langkah-langkah untuk melakukan klasifikasi menggunakan *K-Nearest Neighbor* sebagai berikut:

1. Menentukan dataset yang berisi data latih, yaitu kumpulan data gejala dan diagnosis yang telah divalidasi oleh pakar sebagai basis kasus sistem.
2. Menerima input berupa gejala yang dipilih atau dimasukkan oleh pengguna sebagai data uji (kasus baru).
3. Menghitung nilai similaritas antara data uji dan seluruh data latih menggunakan rumus similaritas berbobot, di mana setiap gejala memiliki nilai bobot sesuai tingkat pengaruhnya terhadap jenis gangguan.
4. Menjumlahkan seluruh nilai kesesuaian gejala yang sama antara data uji dan data latih, kemudian membaginya dengan total bobot keseluruhan untuk memperoleh nilai similaritas dalam rentang 0 sampai 1.
5. Mengurutkan hasil nilai similaritas dari yang terbesar hingga yang terkecil untuk mengetahui data latih yang memiliki tingkat kemiripan tertinggi terhadap data uji.
6. Menentukan nilai k (jumlah tetangga terdekat) yang akan digunakan dalam proses klasifikasi.
7. Mengambil sebanyak k data dengan nilai similaritas tertinggi, kemudian menghitung jumlah kemunculan masing-masing kelas diagnosis.
8. Menentukan kesimpulan akhir berdasarkan kelas diagnosis yang paling banyak muncul (mayoritas) dari k tetangga terdekat sebagai hasil diagnosis awal sistem.

Pada proses perhitungan jarak data uji dan data latih, digunakan metrik jarak Euclidean dengan menerapkan Persamaan 2.1.

$$d(p,q) = \frac{\sum_{i=1}^{n} \mathrm{f(pi,qi)qX\ wi}}{\mathrm{wi}} \quad \text{.....(2.4)}$$

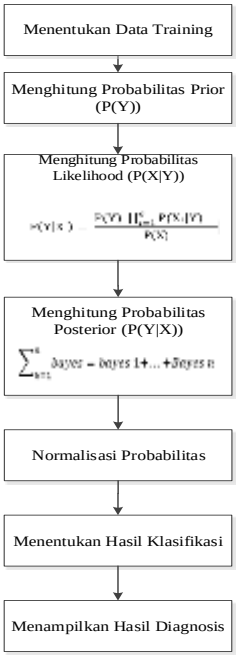
Keterangan:
p = data baru
q = data pada penyimpanan
n = jumlah atribut setiap kasus
i = atribut individu antara 1 sampai n
f = fungsi similarity atribut i antara kasus p dengan kasus q
w = bobot yang akan diberikan ke atribut i

## 3. HASIL DAN PEMBAHASAN

Algoritma sistem merupakan serangkaian langkah logis yang dilakukan oleh sistem dalam memproses data dan menghasilkan solusi terhadap suatu permasalahan. Dalam penelitian ini, algoritma yang digunakan adalah metode Naive Bayes yang diterapkan untuk melakukan klasifikasi penyakit dispepsia akibat infeksi Helicobacter Pylori berdasarkan data gejala yang dimasukkan ke dalam sistem. Metode Naive Bayes bekerja dengan menghitung probabilitas dari setiap kelas (jenis penyakit) berdasarkan atribut-atribut gejala yang dimiliki, seperti nyeri epigastrium, mual, muntah, perut kembung, dan gejala klinis lainnya. Berdasarkan perhitungan tersebut, sistem akan menentukan kategori penyakit dengan

**JURNAL SISTEM INFORMASI TGD**
**Volume 5, Nomor 6, November 2026, Hal 1432-1443**
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*

nilai probabilitas tertinggi sebagai hasil diagnosis. Berikut ini merupakan alur kerja dari metode Naive Bayes dalam sistem diagnosis penyakit dispepsia, yaitu sebagai berikut:




Gambar 2 Alur Kerja *Naïve Bayesian Classification*

1. Menentukan Dataset (*Data Training*)
   Tahap pertama adalah menyiapkan *data training* yang digunakan sebagai dasar perhitungan dalam metode Naive Bayes. Data ini berisi atribut gejala klinis serta label kelas berupa jenis penyakit dispepsia (H. Pylori, Dispepsia Fungsional, Gastritis, Tukak Lambung, GERD).
2. Menghitung Probabilitas Prior
   Pada tahap ini dihitung probabilitas awal (*prior*) dari masing-masing kelas penyakit berdasarkan jumlah kemunculan tiap kelas dalam *dataset* dibandingkan dengan total data.
3. Menghitung Probabilitas *Likelihood*
   Selanjutnya, sistem menghitung probabilitas kemunculan setiap gejala terhadap masing-masing kelas penyakit. Untuk menghindari nilai probabilitas nol, digunakan teknik *Laplace Smoothing*.
4. Menghitung Probabilitas Posterior
   Probabilitas *posterior* dihitung dengan mengalikan probabilitas *prior* dengan seluruh probabilitas *likelihood* dari gejala yang dipilih oleh pengguna.
5. Melakukan Normalisasi Probabilitas
   Nilai probabilitas yang diperoleh dari masing-masing kelas kemudian dinormalisasi agar total probabilitas menjadi 1, sehingga dapat dibandingkan secara proporsional.
6. Menentukan Hasil Klasifikasi
   Sistem akan membandingkan nilai probabilitas dari setiap kelas dan memilih kelas dengan nilai probabilitas tertinggi sebagai hasil akhir diagnosis penyakit.
7. Menampilkan Hasil Diagnosis
   Tahap terakhir adalah menampilkan hasil klasifikasi berupa kategori penyakit dispepsia beserta nilai tingkat kepercayaan (*confidence score*) sebagai pendukung keputusan medis.

Setelah seluruh data yang diperlukan untuk pengembangan sistem pakar terkumpul, tahap selanjutnya adalah melakukan identifikasi dan analisis data. Tujuan dari proses ini adalah untuk mengekstraksi pola probabilitas dari kumpulan data penyakit dispepsia yang telah diperoleh. Analisis dilakukan dengan menerapkan algoritma *Naive Bayes Classification*, di mana setiap atribut gejala yang dimasukkan oleh pengguna dievaluasi secara independen untuk menghitung kemungkinan (probabilitas) terhadap masing-masing kategori penyakit. Hasil dari perhitungan probabilitas ini akan menjadi dasar utama dalam penentuan diagnosis oleh sistem pakar. Dengan pendekatan ini, sistem mampu memberikan hasil klasifikasi secara cepat, objektif, dan konsisten dalam membantu proses diagnosis penyakit dispepsia.

Dalam penelitian ini, selain Naive Bayes, digunakan juga metode K-Nearest Neighbors (KNN) untuk melakukan klasifikasi penyakit dispepsia. KNN bekerja berdasarkan prinsip kemiripan fitur, di mana sistem akan mencari sejumlah *k* data pasien terdahulu (tetangga terdekat) yang memiliki kemiripan gejala paling tinggi dengan pasien baru. Hasil diagnosis ditentukan berdasarkan label penyakit yang paling dominan muncul di antara tetangga terdekat tersebut. Berikut ini merupakan alur kerja dari metode K-Nearest Neighbors (KNN) dalam sistem diagnosis penyakit dispepsia:

**JURNAL SISTEM INFORMASI TGD**
**Volume 5, Nomor 6, November 2026, Hal 1432-1443**
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*




Gambar 3 Kerangka Kerja Metode KNN

Adapun langkah-langkah algoritma penyelesaian menggunakan metode *K-Nearest Neighbor* (KNN) adalah sebagai berikut:

1. Menentukan *dataset* yang berisi data latih, yaitu kumpulan data gejala dan diagnosis penyakit dispepsia yang telah divalidasi oleh pakar di Rumah Sakit Umum Hidayah.
2. Menerima input berupa gejala yang dipilih atau dimasukkan oleh pengguna sebagai data uji.
3. Menghitung jarak antara data uji dengan seluruh data latih menggunakan rumus perhitungan jarak, seperti *Euclidean Distance*.
4. Mengurutkan hasil perhitungan jarak dari nilai terkecil hingga terbesar untuk mengetahui data yang paling dekat dengan data uji.
5. Menentukan nilai *k* (jumlah tetangga terdekat) yang akan digunakan dalam proses klasifikasi.
6. Mengambil sebanyak *k* data dengan jarak terdekat, kemudian menghitung jumlah kemunculan masing-masing kelas diagnosis.

Menentukan kesimpulan akhir berdasarkan kelas yang memiliki jumlah terbanyak (mayoritas) dari k tetangga terdekat sebagai hasil diagnosis awal.

Berikut adalah penjabaran langkah demi langkah berdasarkan studi kasus pasien dengan gejala: **G01, G02,** G03, G07, G16 dan G22.

1. Menentukan Dataset *(Data Training)*
   Tahap awal adalah mengidentifikasi data latih yang menjadi basis kasus sistem (Tabel 3.1) dan probabilitas yang diekstraksi (Tabel 3.3).
2. Menghitung *Probabilitas Prior*
   *Probabilitas prior* (P(Hi)) adalah probabilitas awal kemunculan suatu penyakit berdasarkan distribusi frekuensi pada dataset data latih (Tabel 3.2). Nilai ini dihitung dengan membagi jumlah kasus per penyakit dengan total seluruh kasus (N=60). Nilai probabilitas prior disajikan dalam tabel di bawah ini:

Tabel 3 Distribusi Nilai Probabilitas Prior

| **Kode** | **Nama Penyakit** | **Rumus (n/N)** | **P(Hi)** |
|---|---|---|---|
| P01 | Dispepsia H. Pylori | 15 / 60 | 0.25 |
| P02 | Dispepsia Fungsional | 15 / 60 | 0.25 |
| P03 | Gastritis H. Pylori | 10 / 60 | 0.1667 |
| P04 | Tukak Lambung | 10 / 60 | 0.1667 |
| P05 | GERD | 10 / 60 | 0.1667 |

3. Menghitung Probabilitas *Likelihood*
   Probabilitas *likelihood* merupakan probabilitas kemunculan sekumpulan gejala (E) jika diketahui penyakitnya (Hi). Penghitungan dilakukan dengan mengalikan probabilitas masing-masing gejala yang muncul pada pasien berdasarkan data pada Tabel 3.3. Rincian proses komputasi perkalian disajikan dalam tabel berikut:

Tabel 4 Rincian Komputasi Likelihood Per Kelas

| Penyakit | Perkalian (G01, G02, G03, G07, G16, G22) | Hasil |
|---|---|---|

JURNAL SISTEM INFORMASI TGD
**Volume 5, Nomor 6, November 2026, Hal 1432-1443**
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*

| | | |
|---|---|---|
| P01 | 0.93 × 0.87 × 0.67 × 0.33 × 0.53 | 0.0946 |
| P02 | 1.00 × 0.01 × 0.33 × 0.01 × 0.01 | 0.00000033 |
| P03 | 1.00 × 1.00 × 1.00 × 0.01 × 0.90 | 0.0090 |
| P04 | 1.00 × 1.00 × 1.00 × 0.01 × 1.00 | 0.0100 |
| P05 | 0.20 × 0.01 × 0.01 × 0.01 × 0.01 | 0.000000002 |

4. Menghitung Probabilitas *Posterior*
Probabilitas *posterior* diperoleh dari hasil perkalian probabilitas prior dengan likelihood. Nilai ini menunjukkan skor relatif untuk setiap kandidat diagnosis sebelum dilakukan normalisasi. Hasil perhitungan skor *posterior* untuk setiap kategori penyakit adalah sebagai berikut:

Tabel 5 Hasil Perhitungan Skor Posterior

| Penyakit | Perhitungan (Prior × Likelihood) | Posterior Score |
|---|---|---|
| P01 | 0.25 × 0.0946 | 0.02365 |
| P02 | 0.25 × 0.00000033 | 0.00000 |
| P03 | 0.1667 × 0.0090 | 0.00150 |
| P04 | 0.1667 × 0.0100 | 0.00167 |
| P05 | 0.1667 × 0.000000002 | 0.00000 |
| Total Evidence (Σ Score) | | 0.02682 |

5. Menghitung *Normalisasi Probabilitas*
Tahap normalisasi bertujuan untuk memastikan bahwa total probabilitas dari seluruh kandidat diagnosis berjumlah 1 (100%). Hal ini dilakukan dengan membagi skor posterior masing-masing penyakit dengan total seluruh skor posterior (*Total Evidence*). Hasil normalisasi disajikan dalam tabel berikut:

Tabel 6 Hasil Normalisasi Probabilitas

| Kode | Nama Penyakit | Perhitungan Normalisasi | Hasil Akhir |
|---|---|---|---|
| P01 | Dispepsia H. Pylori | 0.02365 / 0.02682 | 0.9614 |
| P02 | Dispepsia Fungsional | 0.00000 / 0.02682 | 0.0177 |
| P03 | Gastritis H. Pylori | 0.00150 / 0.02682 | 0.0168 |
| P04 | Tukak Lambung | 0.00167 / 0.02682 | 0.0040 |
| P05 | GERD | 0.00000 / 0.02682 | 0.0001 |

6. Kesimpulan Akhir Diagnosis
Berdasarkan hasil normalisasi, sistem mengambil kategori penyakit dengan nilai probabilitas tertinggi sebagai hasil diagnosis final. Kesimpulan diagnosis beserta tingkat kepercayaannya disajikan pada tabel di bawah ini:

Tabel 7 Kesimpulan Akhir Diagnosis (NB)

| Parameter | Hasil Diagnosa |
|---|---|
| Penyakit Terdeteksi | Dispepsia Akibat Infeksi Helicobacter Pylori (P01) |
| Tingkat Kepercayaan | 96.14% |

### 3.1 Penerapan Metode KNN

Penerapan metode K-Nearest Neighbors (KNN) dalam diagnosis dispepsia didasarkan pada prinsip kemiripan (*similarity*). Sistem mencari kasus-kasus lama dalam *knowledge base* yang memiliki manifestasi gejala paling serupa dengan pasien baru (data uji). Berikut adalah simulasi perhitungan detail menggunakan data uji dengan gejala: **G01, G02,** G03, G07, G16, dan G22.

1. Representasi Data Uji (Vektor Biner)
Data uji dikonversi menjadi vektor biner 22 dimensi (G01-G22), di mana nilai 1 mewakili gejala yang dialami pasien dan nilai 0 mewakili gejala yang tidak dialami.

Tabel 8 Vektor Biner Data Uji (Kasus Baru)

| G01 | G02 | G03 | G04 | G05 | G06 | G07 | G08 | G09 | G10 | G11 | G12 | G13 | G14 | G15 | G16 | G17 | G18 | G19 | G20 | G21 | G22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |

2. Menghitung Jarak *Euclidean*
Jarak antara data uji (X) dan setiap data latih (Y) dihitung menggunakan rumus *Euclidean Distance*. Berikut adalah ekspansi perhitungan lengkap untuk 22 gejala antara Data Uji dengan Kasus 1 (Ahmad):

**JURNAL SISTEM INFORMASI TGD**
**Volume 5, Nomor 6, November 2026, Hal 1432-1443**
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
*https://ojs.trigunadharma.ac.id/index.php/jsi*

$$d(Uji, K01) = \sqrt{\begin{matrix}(G01: 1-1)^2 + (G02: 1-1)^2 + (G03: 1-1)^2 + (G04: 0-0)^2 + (G05: 0-0)^2 + \\ (G06: 0-0)^2 + (G07: 0-1)^2 + (G08: 0-0)^2 + (G09: 0-0)^2 + (G10: 0-0)^2 + \\ (G11: 0-0)^2 + (G12: 0-1)^2 + (G13: 0-0)^2 + (G14: 0-0)^2 + \\ (G15: 0-0)^2 + (G16: 1-1)^2 + (G17: 0-0)^2 + (G18: 0-0)^2 + \\ (G19: 0-0)^2 + (G20: 0-0)^2 + (G21: 0-0)^2 + (G22: 1-0)^2\end{matrix}}$$

$$d(Uji, K01) = \sqrt{\begin{matrix}0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + (-1)^2 + 0^2 + 0^2 \\ + 0^2 + 0^2 + (-1)^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + \\ 0^2 + 0^2 + 1^2\end{matrix}}$$

$d(Uji, K01) = \sqrt{3}$=1.732

3. Hasil Perhitungan Jarak (Seluruh Data Latih)
Berikut adalah hasil perhitungan jarak Euclidean antara data uji terhadap seluruh 60 data latih dalam sistem:

Tabel 9 Hasil Perhitungan Jarak Euclidean (N=60)

| ID | Nama Pasien (Data Latih) | Diagnosis | Jarak Euclidean | Ranking |
|---|---|---|---|---|
| K01 | Ahmad | P01 | 1.732 | 2 |
| K02 | Siti | P01 | 2.449 | - |
| K03 | Budi | P01 | 0.000 | 1 |
| K04 | Dewi | P01 | 2.828 | - |
| K05 | Rudi | P01 | 2.000 | - |
| K06 | Ratna | P01 | 2.449 | - |
| K07 | Hendra | P01 | 2.000 | - |
| K08 | Lisa | P01 | 2.449 | - |
| ... | ... | ... | ... | ... |
| K60 | Eko Prasetyo | P05 | 3.000 | - |

4. Penentuan Tetangga Terdekat (K=3)
Berdasarkan Tabel 3.11, sistem mengambil tiga kasus dengan jarak terkecil sebagai referensi klasifikasi, Setelah jarak diurutkan, tiga kasus teratas (jarak terkecil) dipilih untuk menentukan hasil diagnosis akhir melalui mekanisme voting.

Tabel 10 Analisis Tiga Tetangga Terdekat (K=3)

| Peringkat | ID Kasus | Nama Pasien | Jarak Euclidean | Diagnosis Kasus Latih |
|---|---|---|---|---|
| 1 | K03 | Budi | 0.000 | P01 |
| 2 | K01 | Ahmad | 1.732 | P01 |
| 3 | K46 | Sulistiya | 1.738 | P04 |

5. Mekanisme Voting Mayoritas

Tahap akhir adalah menghitung frekuensi kemunculan kategori diagnosis pada K tetangga terdekat. Diagnosis dengan jumlah kemunculan terbanyak dipilih sebagai hasil akhir.

Tabel 11 Analisis Voting Mayoritas

| Kategori Diagnosis | Jumlah Voting (K=3) | Persentase |
|---|---|---|
| Dispepsia Akibat Infeksi H. Pylori (P01) | 2 | 66.67% |
| Diagnosis Lainnya | 1 | 33.33% |

6. Mekanisme esimpulan Akhir Diagnosis (KNN)
Berdasarkan perolehan suara mayoritas (3 dari 3 tetangga terdekat), maka data uji diklasifikasikan ke dalam kategori Dispepsia Akibat Infeksi *Helicobacter Pylori* (P01).

**3.2 Perbandingan Hasil Naive Bayes Dan KNN**

Tahap akhir dari metodologi penelitian ini adalah melakukan komparasi hasil diagnosis antara metode Naive Bayes dan K-Nearest Neighbors (KNN). Perbandingan ini bertujuan untuk memvalidasi konsistensi sistem pakar dalam memproses data uji yang sama menggunakan dua pendekatan algoritma yang berbeda.

Tabel 12 Perbandingan Hasil Diagnosis Kedua Metode

| Parameter Perbandingan | Metode Naive Bayes | Metode K-Nearest Neighbors (KNN) |
|---|---|---|
| Hasil Diagnosis | Dispepsia Akibat Infeksi H. Pylori (P01) | Dispepsia Akibat Infeksi H. Pylori (P01) |

| | | |
|---|---|---|
| Dasar Keputusan | Probabilitas Posterior Tertinggi | Voting Mayoritas Tetangga Terdekat (K=3) |
| Nilai Metrik | Tingkat Kepercayaan: 96.14% | Voting: 66.67% (2/3) |
| Status Konsistensi | KONSISTEN | |

## 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa data gejala penyakit dispepsia akibat infeksi Helicobacter pylori berhasil dianalisis dan disusun menjadi dataset yang dapat digunakan sebagai dasar dalam proses klasifikasi diagnosis. Metode Naive Bayes dan K-Nearest Neighbors (KNN) berhasil diimplementasikan pada sistem untuk melakukan diagnosis berdasarkan gejala yang dimasukkan pengguna, sehingga mampu menghasilkan klasifikasi dengan performa yang baik. Selain itu, sistem diagnosis berbasis web yang dibangun menggunakan framework Laravel dan database MySQL telah berfungsi dengan baik dalam membantu proses diagnosis awal secara cepat, mudah, dan informatif. Hasil evaluasi menunjukkan bahwa kedua metode memiliki kemampuan yang baik dalam mengklasifikasikan kondisi pasien, sehingga dapat dimanfaatkan sebagai pendukung keputusan bagi tenaga medis maupun masyarakat dalam melakukan diagnosis awal dispepsia yang berkaitan dengan infeksi Helicobacter pylori.

## UCAPAN TERIMAKASIH

Ucapan terima kasih yang sebesar-besarnya disampaikan kepada pihak-pihak yang telah memberikan dukungan dan kontribusi yang sangat berarti dalam terlaksananya penelitian ini. Pertama, penulis menyampaikan terima kasih kepada Bapak Nurcahyo Budi Nugroho, S.Kom.,M.Kom selaku dosen pembimbing I yang telah memberikan bimbingan, arahan, serta dukungan ilmiah selama proses penyusunan penelitian ini. Pengetahuan dan pengalaman yang beliau bagikan sangat membantu dalam pengembangan konsep serta metodologi penelitian. Penulis juga mengucapkan terima kasih kepada Bapak D Dedi Setiawan,S.Kom.,M.Komselaku dosen pembimbing II atas saran, masukan, dan dukungan yang diberikan sehingga penelitian ini dapat diselesaikan dengan baik. Selain itu, penulis turut mengapresiasi bantuan dan dukungan dari berbagai pihak yang telah memberikan motivasi, ide, maupun kontribusi lainnya demi kelancaran penelitian ini. Semoga segala bantuan dan kebaikan yang diberikan mendapatkan balasan yang setimpal serta bermanfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang teknologi informasi dan pengolahan data.

## DAFTAR PUSTAKA

[1] S. Zhu, X. Tan, H. Huang, Y. Zhou, and Y. Liu, “Heliyon Data-driven rapid detection of Helicobacter pylori infection through machine learning with limited laboratory parameters in Chinese primary clinics,” *Heliyon*, vol. 10, no. 15, p. e35586, 2024, doi: 10.1016/j.heliyon.2024.e35586.

[2] V. A. Duha and A. F. Rozi, “SISTEM PAKAR DIAGNOSA SINDROM DISPEPSIA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER ( STUDI KASUS : PUSKESMAS DEPOK 3 ),” vol. 11, no. 3, pp. 887–897, 2023.

[3] V. Wulandari, W. J. Sari, and Z. Alfian, “Implementation of Naïve Bayes Classifier and K-Nearest Neighbor Algorithms for Chronic Kidney Disease Classification Implementasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor untuk Klasifikasi Penyakit Ginjal Kronik,” vol. 4, no. April, pp. 710–718, 2024.

[4] A. Yonata, P. Sagita, N. Islamy, and L. Pura, “Dispepsia : Klasifikasi , Faktor Risiko , Patofisiologi dan Tatalaksana Dyspepsia : Classification , Risk Factors , Pathophysiology and Management,” *JK Unila*, vol. 9, pp. 1–6, 2025.

[5] A. J. Sidik, “Diagnosis dan Tata Laksana Dispepsia Infeksi Dengue Sekunder : Pato siologi , Diagnosis , dan Implikasi Klinis,” *K-313/ vo*, vol. 51, no. 3, pp. 140–144, 2024.

[6] M. Zakaria, E. Affandi, and D. Helsa, “Sistem Pakar Dalam Mendiagnosa Penyakit Blast Pada Tanaman Padi Dengan Metode Teorema Bayes,” vol. 3, no. September, pp. 738–747, 2024.

[7] T. Rachman, “SISTEM PAKAR DIAGNOSA PENYAKIT MATA MENGGUNAKAN METODE FORWARD CHAINING (Studi kasus RUMAH SAKIT UMUM DAERAH LAWANG),” *J. Mnemon.*, vol. 4, no. 2, pp. 38–44, 2021.

[8] M. K. Muafi, M. K. Andi Wijaya, and V. A. Aziz, “SISTEM PAKAR MENDIAGNOSA PENYAKIT MATA PADA MANUSIA MENGGUNAKAN METODE FORWARD CHAINING,” *Core-IT J. Komputasi dan Teknol. Inf.*, vol. 1, no. 1, 2020.

[9] H. Mas’ud, “IMPLEMENTASI METODE FORWARD CHAINING DALAM SISTEM PAKAR UNTUK MENDETEKSI KERUSAKAN JARINGAN LOCAL AREA NETWORK (LAN),” *Diffusion*, vol. 4, no. 1, pp. 161–167, 2024.

[10] H. F. Putro, R. T. Vulandari, and W. L. Y. Saptomo, “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan,” *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.

[11] N. Chatrina Siregar, R. Ruli, A. Siregar, ; M Yoga, and D. Sudirman, “Implementasi Metode Naive Bayes

Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ)," *J. Teknol. Aliansi Perguru. Tinggi BUMN*, vol. 3, no. 1, pp. 102–110, 2020.

[12] Muqorobin and M. Bagoes Pakarti, "Sistem Prediksi Lama Studi Kuliah Menggunakan Metode Naive Bayes," *J. Inform. Komput. dan Bisnis*, vol. 2, no. 1, pp. 117–129, 2022.

[13] M. Nanja and Purwanto, "METODE K-NEAREST NEIGHBOR BERBASIS FORWARD SELECTION UNTUK PREDIKSI HARGA KOMODITI LADA," *J. Pseudocode*, vol. 9, no. 2, pp. 1848–1853, 2025.