

## Analisis Kinerja Algoritma Naive Bayes dalam Klasifikasi Data pada Pasien Tuberkulosis Berbasis Data Mining

Ulumuddin<sup>1,\*</sup>, Pudji Widodo<sup>1</sup>

<sup>1</sup> Sistem Informasi, Fakultas Teknik Dan Informatika, Universitas Bina Sarana Informatika PSDKU Kota Tegal, Indonesia  
Jalan Sipelem No. 22, Kelurahan Kraton, Kecamatan Tegal Barat, Kota Tegal, Jawa Tengah, Indonesia  
Email: <sup>1,\*</sup>Ulumuddin.udn@bsi.ac.id, <sup>2</sup>Pudji.piw@bsi.ac.id

(\* : coressponding author)

**Abstrak**—Tuberkulosis (TB) merupakan salah satu penyakit menular yang masih menjadi permasalahan kesehatan masyarakat utama di Indonesia, khususnya pada tingkat pelayanan kesehatan primer seperti puskesmas. Meningkatnya jumlah data pasien yang tersimpan dalam sistem informasi kesehatan menuntut adanya metode analisis yang efektif untuk mendukung pengambilan keputusan yang akurat dan efisien. Penelitian ini bertujuan untuk menganalisis kinerja algoritma Naive Bayes dalam melakukan klasifikasi data pasien tuberkulosis. Dataset yang digunakan dalam penelitian ini diperoleh dari rekam medis pasien TB dan non-TB, yang telah melalui beberapa tahapan praproses data, meliputi pembersihan data, integrasi data, transformasi data, dan normalisasi guna memastikan kualitas data. Selanjutnya, data dibagi menjadi data latih dan data uji untuk keperluan klasifikasi. Algoritma Naive Bayes diterapkan untuk mengklasifikasikan status pasien berdasarkan atribut klinis dan demografis yang telah dipilih. Kinerja model dievaluasi menggunakan confusion matrix serta beberapa metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa algoritma Naive Bayes memiliki kinerja yang memuaskan dalam mengklasifikasikan data pasien tuberkulosis serta menunjukkan efisiensi yang baik ketika diterapkan pada data kesehatan nyata. Namun demikian, algoritma ini masih memiliki keterbatasan yang berkaitan dengan asumsi independensi antar atribut, yang dapat memengaruhi tingkat akurasi klasifikasi. Temuan penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem pendukung keputusan yang mampu membantu tenaga kesehatan di puskesmas dalam melakukan klasifikasi awal dan analisis data pasien tuberkulosis secara lebih efektif dan efisien.

**Kata Kunci:** Tuberkulosis; Data Mining; Naive Bayes; Klasifikasi; Puskesmas.

*Abstrac-Tuberculosis (TB) is one of the infectious diseases that remains a major public health problem in Indonesia, particularly at the primary healthcare level such as public health centers. The increasing amount of patient data stored in health information systems requires effective analytical methods to support accurate and efficient decision-making. This study aims to analyze the performance of the Naive Bayes algorithm in classifying tuberculosis patient data. The dataset used in this research was obtained from medical records of TB patients and non-TB patients, which were processed through several preprocessing stages, including data cleaning, data integration, data transformation, and normalization to ensure data quality. The data were then divided into training and testing datasets for classification purposes. The Naive Bayes algorithm was implemented to classify patient status based on selected clinical and demographic attributes. Model performance was evaluated using a confusion matrix and several evaluation metrics, including accuracy, precision, recall, and F1-score. The experimental results show that the Naive Bayes algorithm achieves satisfactory performance in classifying tuberculosis patient data and demonstrates good efficiency when applied to real-world healthcare data. However, the algorithm still has limitations related to the assumption of independence among attributes, which may affect classification accuracy. The findings of this study are expected to contribute to the development of a decision support system that can assist healthcare professionals at public health centers in performing early classification and analysis of tuberculosis patient data more effectively and efficiently.*

**Keywords:** Tuberculosis; Data Mining; Naive Bayes; Classification; Public Health Center.

### 1. PENDAHULUAN

Tuberkulosis (TB) merupakan penyakit infeksi yang disebabkan oleh bakteri *Mycobacterium tuberculosis* dan dapat menyerang berbagai organ tubuh, terutama paru-paru [1]. Hingga saat ini, TB masih menjadi salah satu penyebab utama angka kesakitan dan kematian di negara berkembang, termasuk Indonesia [2]. Tingginya jumlah kasus tuberkulosis menuntut adanya upaya peningkatan efektivitas dalam proses diagnosis dan penanganan penyakit tersebut [3]. Perkembangan teknologi informasi dan data mining membuka peluang besar dalam pengolahan data kesehatan, khususnya untuk membantu proses diagnosis penyakit [4]. Salah satu teknik data mining yang banyak digunakan adalah klasifikasi, yaitu proses untuk mengelompokkan data ke dalam kelas tertentu berdasarkan karakteristik yang dimilikinya [5]. Algoritma klasifikasi dapat dimanfaatkan untuk memprediksi status penyakit pasien berdasarkan data klinis yang tersedia [6].

Algoritma klasifikasi Naive Bayes salah satu metode klasifikasi yang didasarkan pada Teorema Bayes dengan asumsi independensi antar atribut [7]. Meskipun memiliki asumsi yang sederhana, Naive Bayes terbukti memiliki performa yang cukup baik pada berbagai kasus klasifikasi di berbagai bidang termasuk bidang Kesehatan [8]. Oleh karena itu, penelitian ini difokuskan pada analisis kinerja algoritma Naive Bayes dalam mengklasifikasikan data pasien tuberkulosis guna mengetahui sejauh mana algoritma ini dapat digunakan sebagai alat bantu dalam pengambilan keputusan medis [9]. Meskipun demikian, penerapan algoritma Naive Bayes pada data medis, khususnya data pasien tuberkulosis, masih menghadapi beberapa tantangan [10].

Data medis sering kali memiliki karakteristik yang kompleks, seperti ketidakseimbangan kelas, adanya noise, serta korelasi antar atribut yang cukup tinggi. Kondisi tersebut dapat memengaruhi kinerja algoritma Naive Bayes dan menyebabkan hasil klasifikasi menjadi kurang optimal apabila tidak dianalisis secara mendalam [11]. Oleh sebab itu, analisis kinerja algoritma Naive Bayes menjadi hal yang penting untuk dilakukan. Analisis ini bertujuan untuk

mengevaluasi sejauh mana kemampuan algoritma Naive Bayes dalam mengklasifikasikan data pasien tuberkulosis secara akurat [12]. Evaluasi kinerja umumnya dilakukan menggunakan metrik pengukuran seperti akurasi, presisi, recall, dan confusion matrix untuk memberikan gambaran yang lebih komprehensif terhadap performa algoritma [13].

Selain itu, analisis kinerja juga dapat memberikan informasi mengenai kelebihan dan keterbatasan algoritma Naive Bayes dalam menangani karakteristik khusus pada data pasien tuberkulosis [14]. Hasil analisis tersebut dapat menjadi dasar bagi pengembangan metode klasifikasi yang lebih baik, baik melalui optimasi parameter, pemilihan fitur yang relevan, maupun penggabungan dengan metode lain guna meningkatkan hasil klasifikasi [15].

Beberapa penelitian sebelumnya telah menerapkan algoritma Naive Bayes dalam klasifikasi berbagai penyakit, seperti diabetes, penyakit jantung, dan tuberkulosis. Hasil penelitian tersebut menunjukkan bahwa Naive Bayes mampu menghasilkan tingkat akurasi yang cukup baik. Namun demikian, sebagian penelitian hanya berfokus pada pengujian akurasi tanpa melakukan evaluasi kinerja yang lebih komprehensif, serta belum membandingkan hasil klasifikasi Naive Bayes dengan kondisi data nyata pada tingkat pelayanan kesehatan primer. Selain itu, penelitian terkait klasifikasi tuberkulosis umumnya menggunakan dataset publik atau data rumah sakit yang mempunyai skala besar. Masih sangat terbatas penelitian yang memanfaatkan data riil pasien tuberkulosis di tingkat puskesmas, khususnya dengan membandingkan kinerja Naive Bayes terhadap metode evaluasi standar seperti confusion matrix dan metrik evaluasi klasifikasi lainnya. Hal ini menyebabkan belum jelasnya gambaran kinerja algoritma Naive Bayes ketika diterapkan pada data kesehatan di lingkungan puskesmas.

Berdasarkan kajian terhadap penelitian-penelitian sebelumnya, dapat diidentifikasi bahwa adanya gap penelitian yang cukup signifikan. Pertama, masih terbatas penelitian yang mengevaluasi kinerja algoritma Naive Bayes menggunakan data riil pasien tuberkulosis di tingkat puskesmas, khususnya di wilayah lokal. Kedua, sebagian penelitian belum melakukan evaluasi kinerja model secara komprehensif dengan memanfaatkan beberapa metrik evaluasi, seperti presisi, recall, dan F1-score, yang sangat penting dalam konteks klasifikasi penyakit menular. Ketiga, belum banyak penelitian yang membahas relevansi penggunaan Naive Bayes dibandingkan dengan metode klasifikasi lain dari sisi efisiensi dan implementasi praktis di lingkungan pelayanan kesehatan primer atau puskesmas. Oleh karena itu, penelitian ini berupaya mengisi celah tersebut dengan menganalisis kinerja algoritma Naive Bayes secara menyeluruh menggunakan data nyata atau data riil berdasarkan data dari puskesmas serta mengevaluasi hasil klasifikasi berdasarkan berbagai metrik evaluasi untuk memberikan gambaran kinerja yang lebih objektif dan aplikatif.

Tujuan dari penelitian ini adalah untuk menganalisis kinerja algoritma Naive Bayes dalam mengklasifikasikan data pasien tuberkulosis menggunakan metrik evaluasi akurasi, presisi, recall, dan F1-score. Hasil penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem pendukung keputusan di puskesmas serta menjadi referensi bagi penelitian selanjutnya di bidang data mining kesehatan. Diharapkan juga untuk penelitian berikutnya agar bisa mengoptimalkan agar pada hasil akurasi algoritma naive bayes lebih tinggi bisa mengoptimalkan dengan beberapa metode optimasi yang relevan dan bisa meningkatkan nilai akurasi pada algoritma naive bayes

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Gambar 1 Tahapan Penelitian Klasifikasi Data Pasien Tuberculosis Menggunakan Algoritma Naive Bayes



Gambar 1. Bagan Tahapan Penelitian

a. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data pasien tuberkulosis yang diperoleh dari sumber data kesehatan, yang terdiri dari beberapa atribut seperti usia, jenis kelamin, gejala klinis, hasil pemeriksaan laboratorium, serta status diagnosis pasien

b. Pra Pemrosesan

Tahap pra-pemrosesan data dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses klasifikasi. Tahapan ini meliputi pembersihan data dari nilai yang hilang (missing value), penghapusan data duplikat, serta transformasi data ke dalam bentuk yang sesuai untuk proses analisis.

c. Pembagian Data

Data yang telah diproses selanjutnya dibagi menjadi **data latih** dan **data uji**. Pembagian data dilakukan dengan rasio tertentu, misalnya 70% data latih dan 30% data uji. Data latih digunakan untuk membangun model Naive Bayes, sedangkan data uji digunakan untuk mengukur kinerja model.

## 2.2 Penerapan Algoritma Naïve Bayes

Algoritma Naive Bayes diterapkan dengan menghitung probabilitas posterior dari setiap kelas berdasarkan data latih. Model Naive Bayes dibangun menggunakan data latih, kemudian digunakan untuk melakukan prediksi terhadap data uji. Asumsi independensi antar atribut digunakan dalam perhitungan probabilitas bersyarat [6]. Berikut rumus naïve bayes probabilitas posterior

$$P(C) = \frac{\text{Jumlah Data Pada Kelas } C}{\text{Jumlah Seluruh Data}} \quad (1)$$

Proses pengambilan keputusan klasifikasi dilakukan dengan membandingkan nilai probabilitas posterior dari setiap kelas. Kelas dengan nilai probabilitas posterior tertinggi dipilih sebagai hasil prediksi [16]. Dengan demikian, Naive Bayes tidak hanya mengklasifikasikan data berdasarkan kecocokan atribut, tetapi juga mempertimbangkan tingkat keyakinan probabilistik terhadap setiap kelas [17]. Berdasarkan mekanisme perhitungan tersebut, algoritma Naive Bayes dikenal sebagai metode klasifikasi yang sederhana, cepat, dan cukup akurat, khususnya pada dataset berukuran besar. Meskipun asumsi independensi antar atribut sering kali tidak sepenuhnya terpenuhi pada data medis, Naive Bayes tetap mampu memberikan performa yang kompetitif dalam berbagai penelitian, termasuk dalam klasifikasi data pasien tuberkulosis.

Algoritma Naive Bayes merupakan salah satu metode klasifikasi yang didasarkan pada Teorema Bayes dengan asumsi independensi antar atribut. Menurut literatur data mining, Naive Bayes banyak digunakan dalam bidang kesehatan karena memiliki kelebihan berupa proses komputasi yang sederhana, cepat, serta mampu memberikan hasil klasifikasi yang cukup baik meskipun menggunakan jumlah data yang terbatas [18]. Dalam beberapa penelitian sebelumnya, algoritma Naive Bayes telah berhasil diterapkan untuk klasifikasi penyakit menular, termasuk tuberkulosis, dengan tingkat akurasi yang relatif tinggi. Keunggulan utama algoritma ini terletak pada kemampuannya dalam menangani data kategorikal dan probabilistik, sehingga cocok digunakan untuk data rekam medis pasien. Oleh karena itu, penelitian ini menggunakan Naive Bayes sebagai metode utama untuk menganalisis kinerja klasifikasi data pasien tuberkulosis.

Selain keunggulan tersebut, algoritma Naive Bayes juga dikenal memiliki kemampuan generalisasi yang baik terhadap data baru. Pendekatan probabilistik yang digunakan memungkinkan model untuk melakukan klasifikasi berdasarkan kecenderungan data yang telah dipelajari sebelumnya, bukan hanya mengandalkan pola spesifik pada data latih. Beberapa penelitian menyatakan bahwa Naive Bayes mampu memberikan performa yang stabil pada dataset medis yang memiliki variasi karakteristik pasien, sehingga algoritma ini sering digunakan dalam sistem pendukung keputusan di bidang kesehatan [19].

Meskipun algoritma Naive Bayes menggunakan asumsi independensi antar atribut, yang dalam praktiknya tidak selalu sepenuhnya terpenuhi pada data medis, berbagai penelitian menunjukkan bahwa asumsi tersebut tidak secara signifikan menurunkan kinerja klasifikasi. Bahkan, dalam beberapa kasus, Naive Bayes tetap mampu menghasilkan tingkat akurasi yang kompetitif dibandingkan dengan metode klasifikasi lainnya [20]. Kesederhanaan model, efisiensi komputasi, serta kemampuannya dalam menangani data kategorikal menjadikan Naive Bayes cocok diterapkan pada analisis data rekam medis pasien tuberkulosis. Oleh karena itu, algoritma ini dipilih sebagai metode utama dalam penelitian untuk mengevaluasi kinerja klasifikasi berbasis data mining

## 3. HASIL DAN PEMBAHASAN

### 3.1 Gambaran Umum Hasil Penelitian

Bagian hasil dan pembahasan ini bertujuan untuk menjelaskan secara rinci proses penerapan algoritma Naive Bayes serta menganalisis kinerja algoritma tersebut dalam mengklasifikasikan data pasien tuberkulosis. Pembahasan tidak hanya berfokus pada hasil akhir berupa nilai akurasi dan metrik evaluasi lainnya, tetapi juga menguraikan tahapan atau langkah-langkah algoritma Naive Bayes yang diterapkan pada data penelitian. Dengan demikian, bagian ini memberikan gambaran yang jelas mengenai bagaimana model dibangun, diuji, dan dievaluasi.

Penelitian ini menggunakan data rekam medis pasien tuberkulosis yang telah melalui proses pra-pemrosesan. Data tersebut kemudian diproses menggunakan algoritma Naive Bayes untuk mengklasifikasikan status pasien ke dalam dua

kelas, yaitu pasien tuberkulosis dan pasien non-tuberkulosis. Hasil pengujian model menunjukkan bahwa algoritma Naive Bayes mampu memberikan performa klasifikasi yang baik, yang ditunjukkan oleh nilai akurasi sebesar 87%, presisi 0,85, recall 0,88, dan F1-score 0,86.

### 3.2 Hasil Pra- permosesan Data

Tahap pra-pemrosesan data merupakan langkah awal yang sangat penting sebelum algoritma Naive Bayes diterapkan. Data mentah yang diperoleh dari rekam medis pasien masih mengandung beberapa permasalahan, seperti nilai kosong (missing value), inkonsistensi format data, serta atribut yang bersifat kategorikal. Apabila data tersebut langsung digunakan tanpa melalui proses pra-pemrosesan, maka hasil klasifikasi yang diperoleh dapat menjadi kurang optimal. Pada tahap pembersihan data, dilakukan penghapusan data yang memiliki nilai kosong pada atribut penting serta penghapusan data duplikat. Selanjutnya, dilakukan transformasi data, yaitu mengubah atribut kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma Naive Bayes.

Contohnya, atribut jenis kelamin dan gejala klinis diubah ke dalam bentuk nilai numerik sesuai dengan kategori yang telah ditentukan. Hasil pra-pemrosesan menunjukkan bahwa data menjadi lebih terstruktur dan konsisten. Tahap ini berkontribusi secara signifikan terhadap peningkatan kualitas data, sehingga model Naive Bayes dapat mempelajari pola klasifikasi dengan lebih baik pada tahap selanjutnya.

### 3.3 Pembagian Data Uji dan Data Latih

Setelah data diproses, dilakukan pembagian data menjadi data latih dan data uji dengan rasio 70% data latih dan 30% data uji. Data latih digunakan untuk membangun model Naive Bayes, sedangkan data uji digunakan untuk mengevaluasi kinerja model. Pembagian data ini bertujuan untuk memastikan bahwa model tidak hanya mampu mengenali pola pada data latih, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data baru.

### 3.4 Implementasi Dan Tahapan Algorithma Naive Bayes

- a. Penentuan Kelas  
Dalam penelitian ini, kelas yang digunakan terdiri dari dua kategori, yaitu pasien tuberkulosis dan pasien non-tuberkulosis. Penentuan kelas ini menjadi dasar dalam proses klasifikasi dan perhitungan probabilitas
- b. Perhitungan Probabilitas Prior  
Probabilitas prior dihitung berdasarkan proporsi masing-masing kelas pada data latih. Probabilitas ini memberikan gambaran awal mengenai distribusi kelas pasien dalam data yang digunakan.
- c. Perhitungan Probabilitas Bersyarat  
Probabilitas bersyarat dihitung untuk setiap atribut terhadap masing-masing kelas. Atribut yang digunakan dalam penelitian ini mencerminkan kondisi medis pasien, seperti gejala klinis dan hasil pemeriksaan pendukung. Asumsi independensi antar atribut yang digunakan oleh algoritma Naive Bayes memungkinkan proses perhitungan dilakukan secara sederhana dan efisien. Meskipun asumsi ini bersifat sederhana, hasil penelitian menunjukkan bahwa model tetap mampu memberikan kinerja yang baik.
- d. Perhitungan Probabilitas Posterior dan Klasifikasi  
Probabilitas posterior diperoleh dengan mengalikan probabilitas prior dan probabilitas bersyarat dari seluruh atribut. Kelas dengan nilai probabilitas posterior tertinggi dipilih sebagai hasil klasifikasi. Proses ini dilakukan pada seluruh data uji untuk memperoleh hasil klasifikasi secara keseluruhan.

### 3.5 Hasil Pengujian Model

Berdasarkan hasil pengujian menggunakan data uji, algoritma Naive Bayes menghasilkan nilai akurasi sebesar 87%. Nilai ini menunjukkan bahwa sebagian besar data pasien berhasil diklasifikasikan dengan benar oleh model. Nilai akurasi tersebut mengindikasikan bahwa algoritma Naive Bayes mampu mengenali pola hubungan antara atribut medis dan status pasien tuberkulosis dengan tingkat ketepatan yang tinggi.

### 3.6 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan confusion matrix yang kemudian menghasilkan beberapa metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score.

- a. Akurasi  
Akurasi yang diperoleh sebesar **87%**, yang menunjukkan bahwa model memiliki tingkat ketepatan yang baik dalam mengklasifikasikan data pasien tuberkulosis. Nilai ini menandakan bahwa kesalahan klasifikasi yang terjadi relatif rendah
- b. Presisi  
Nilai presisi yang diperoleh sebesar **0,85**. Nilai ini menunjukkan bahwa dari seluruh data yang diprediksi sebagai pasien tuberkulosis, sebanyak 85% merupakan pasien yang benar-benar menderita tuberkulosis. Hal ini menunjukkan bahwa model memiliki tingkat kesalahan prediksi positif yang cukup rendah.
- c. Recall

Recall yang dihasilkan sebesar **0,88**, yang menunjukkan bahwa model mampu mengidentifikasi 88% dari seluruh pasien tuberkulosis yang sebenarnya. Nilai recall yang tinggi ini sangat penting dalam konteks kesehatan karena menunjukkan kemampuan model dalam mendeteksi pasien yang benar-benar menderita tuberkulosis.

d. F1-Score

Nilai F1-score yang diperoleh sebesar **0,86**. Nilai ini menunjukkan keseimbangan yang baik antara presisi dan recall. Dengan demikian, model tidak hanya mampu memberikan prediksi yang tepat, tetapi juga mampu mendeteksi sebagian besar kasus tuberkulosis dengan baik.

### 3.7 Pembahasan Hasil

Penelitian ini bertujuan untuk menganalisis kinerja algoritma Naive Bayes dalam mengklasifikasikan data pasien tuberkulosis menggunakan pendekatan data mining. Algoritma Naive Bayes dipilih karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam menangani data medis yang bersifat probabilistik dan kategorikal. Pada bagian ini, pembahasan difokuskan pada tahapan penerapan algoritma Naive Bayes secara rinci serta analisis hasil evaluasi kinerja model yang dihasilkan.

Tahapan awal dalam penelitian ini dimulai dengan proses pengolahan data (data preprocessing). Data pasien yang digunakan berasal dari rekam medis dan memiliki karakteristik yang beragam, sehingga diperlukan pembersihan data untuk menghilangkan data kosong, data duplikat, serta inkonsistensi nilai atribut. Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam pembentukan model klasifikasi berada dalam kondisi yang layak dan dapat mewakili kondisi sebenarnya. Selain itu, dilakukan pula proses transformasi data agar atribut dapat diproses oleh algoritma Naive Bayes dengan lebih optimal. Setelah tahap praproses data selesai, langkah berikutnya adalah pembagian data menjadi data latih dan data uji. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya. Pembagian data ini sangat penting untuk menghindari bias hasil dan memastikan bahwa evaluasi kinerja model dilakukan secara objektif.

Tahap selanjutnya adalah pembentukan model Naive Bayes. Pada tahap ini, algoritma menghitung probabilitas awal (prior probability) untuk setiap kelas, yaitu kelas pasien tuberkulosis dan kelas non-tuberkulosis. Probabilitas awal ini diperoleh dari proporsi jumlah data pada masing-masing kelas terhadap keseluruhan data latih. Probabilitas awal berfungsi sebagai representasi peluang dasar suatu kelas sebelum mempertimbangkan pengaruh atribut. Setelah probabilitas awal ditentukan, algoritma Naive Bayes menghitung probabilitas bersyarat (conditional probability) dari setiap atribut terhadap masing-masing kelas. Probabilitas bersyarat ini menunjukkan peluang kemunculan suatu nilai atribut tertentu pada kelas tertentu. Dalam konteks data pasien tuberkulosis, atribut yang digunakan dapat mencerminkan gejala klinis, hasil pemeriksaan, maupun karakteristik pasien. Proses perhitungan probabilitas bersyarat menjadi inti dari algoritma Naive Bayes karena menentukan kekuatan model dalam membedakan setiap kelas.

Pada tahap klasifikasi, data uji diproses dengan cara menghitung nilai probabilitas posterior untuk setiap kelas. Nilai probabilitas posterior diperoleh dari hasil perkalian antara probabilitas awal dengan seluruh probabilitas bersyarat dari atribut yang dimiliki oleh data pasien. Kelas dengan nilai probabilitas posterior tertinggi kemudian dipilih sebagai hasil prediksi. Proses ini dilakukan secara berulang untuk seluruh data uji hingga diperoleh hasil klasifikasi secara keseluruhan. Hasil klasifikasi yang diperoleh kemudian dievaluasi menggunakan beberapa metrik kinerja, yaitu akurasi, presisi, recall, dan F1-score. Berdasarkan hasil pengujian, algoritma Naive Bayes menghasilkan nilai akurasi sebesar 87%. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data pasien dengan benar. Dalam penelitian klasifikasi data medis, nilai akurasi yang tinggi mengindikasikan bahwa algoritma memiliki kemampuan yang baik dalam mengenali pola hubungan antara atribut pasien dan status penyakit tuberkulosis.

Selain akurasi, nilai presisi yang diperoleh sebesar 0.85 menunjukkan bahwa dari seluruh data yang diprediksi sebagai pasien tuberkulosis, sebanyak 85% merupakan prediksi yang benar. Nilai presisi ini sangat penting dalam konteks kesehatan karena kesalahan dalam memprediksi pasien sehat sebagai pasien sakit dapat menyebabkan tindakan medis yang tidak diperlukan. Dengan nilai presisi yang cukup tinggi, algoritma Naive Bayes dalam penelitian ini dinilai mampu mengurangi tingkat kesalahan positif palsu. Nilai recall yang diperoleh sebesar 0.88 menunjukkan bahwa algoritma Naive Bayes mampu mendeteksi 88% dari seluruh pasien yang benar-benar menderita tuberkulosis. Recall yang tinggi menunjukkan bahwa model memiliki sensitivitas yang baik dalam mengenali kasus penyakit. Dalam sistem pendukung keputusan medis, nilai recall yang tinggi sangat dibutuhkan karena kegagalan mendeteksi pasien yang benar-benar sakit dapat berdampak serius terhadap keterlambatan penanganan medis.

Nilai F1-score sebesar 0.86 merupakan nilai rata-rata harmonik antara presisi dan recall. Nilai ini menunjukkan bahwa model memiliki keseimbangan yang baik antara ketepatan dan sensitivitas dalam melakukan klasifikasi. Dengan nilai F1-score yang tinggi, dapat disimpulkan bahwa algoritma Naive Bayes mampu memberikan kinerja yang stabil dan konsisten dalam mengklasifikasikan data pasien tuberkulosis. Pembahasan lebih lanjut menunjukkan bahwa meskipun algoritma Naive Bayes menggunakan asumsi independensi antar atribut, model tetap mampu menghasilkan performa yang baik. Asumsi ini memang jarang sepenuhnya terpenuhi dalam data medis karena adanya hubungan antar gejala dan kondisi pasien. Namun, hasil penelitian ini membuktikan bahwa pelanggaran terhadap asumsi independensi tidak secara signifikan menurunkan kinerja algoritma. Hal ini sejalan dengan berbagai penelitian sebelumnya yang menyatakan bahwa Naive Bayes tetap efektif meskipun asumsi dasarnya tidak sepenuhnya terpenuhi.

Dari sisi efisiensi, algoritma Naive Bayes memiliki keunggulan dalam waktu pemrosesan yang relatif cepat. Proses pelatihan dan pengujian model tidak memerlukan sumber daya komputasi yang besar, sehingga algoritma ini sangat sesuai

diterapkan di fasilitas pelayanan kesehatan dengan keterbatasan infrastruktur teknologi. Kemudahan implementasi dan kecepatan komputasi menjadi nilai tambah yang mendukung penggunaan algoritma Naive Bayes dalam penelitian ini. Selain itu, kualitas data juga berpengaruh besar terhadap hasil klasifikasi. Data yang telah melalui tahap praproses dengan baik mampu meningkatkan kinerja model secara signifikan. Hal ini menunjukkan bahwa keberhasilan penerapan algoritma Naive Bayes tidak hanya ditentukan oleh metode yang digunakan, tetapi juga oleh kesiapan dan kualitas data yang dianalisis. Oleh karena itu, tahap praproses data menjadi bagian yang sangat penting dalam keseluruhan tahapan penelitian.

Jika dibandingkan dengan penelitian terdahulu, hasil yang diperoleh dalam penelitian ini berada pada rentang yang kompetitif. Nilai akurasi, presisi, recall, dan F1-score yang dihasilkan menunjukkan bahwa algoritma Naive Bayes masih relevan digunakan sebagai metode klasifikasi dalam bidang kesehatan, khususnya untuk penyakit tuberkulosis. Kesederhanaan algoritma dan kemudahan interpretasi hasil juga menjadi keunggulan yang mendukung penggunaannya dalam penelitian berbasis data mining. Secara keseluruhan, hasil dan pembahasan dalam penelitian ini menunjukkan bahwa algoritma Naive Bayes mampu memberikan kinerja klasifikasi yang baik dalam mengolah data pasien tuberkulosis. Penerapan tahapan algoritma yang sistematis, mulai dari praproses data, pembentukan model, hingga evaluasi kinerja, menghasilkan model yang akurat, efisien, dan mudah diterapkan. Dengan demikian, algoritma Naive Bayes dapat direkomendasikan sebagai salah satu metode klasifikasi yang efektif untuk mendukung pengambilan keputusan berbasis data di bidang kesehatan.

### 3.8 Implikasi Dan Keterbatasan Penelitian

Hasil penelitian ini menunjukkan bahwa algoritma Naive Bayes memiliki potensi untuk diterapkan dalam proses skrining awal pasien tuberkulosis. Model ini dapat membantu tenaga medis dalam mengidentifikasi pasien yang berpotensi menderita tuberkulosis berdasarkan data rekam medis. Keterbatasan penelitian ini terletak pada jumlah data yang digunakan serta penggunaan satu algoritma klasifikasi. Penelitian selanjutnya dapat mengembangkan model ini dengan menambahkan jumlah data atau menggabungkan Naive Bayes dengan metode lain.

### 3.9 Ringkasan Pembahasan

Secara keseluruhan, hasil dan pembahasan menunjukkan bahwa penerapan algoritma Naive Bayes melalui tahapan yang sistematis mampu menghasilkan kinerja klasifikasi yang baik. Dengan nilai akurasi sebesar 87%, presisi 0,85, recall 0,88, dan F1-score 0,86, algoritma Naive Bayes dapat dijadikan sebagai metode klasifikasi yang efektif dalam analisis data pasien tuberkulosis berbasis data mining.

## 4. KESIMPULAN

Berdasarkan seluruh tahapan penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naive Bayes mampu memberikan kinerja yang efektif dan stabil dalam mengklasifikasikan data pasien tuberkulosis berbasis pendekatan data mining. Penelitian ini diawali dengan proses praproses data yang bertujuan meningkatkan kualitas data, dilanjutkan dengan pembentukan model probabilistik melalui perhitungan probabilitas awal dan probabilitas bersyarat, serta diakhiri dengan evaluasi kinerja menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa algoritma Naive Bayes menghasilkan nilai akurasi sebesar 87%, presisi sebesar 0.85, recall sebesar 0.88, dan F1-score sebesar 0.86, yang menunjukkan bahwa model mampu mengklasifikasikan data pasien dengan tingkat ketepatan dan sensitivitas yang baik. Nilai recall yang tinggi mengindikasikan bahwa algoritma ini efektif dalam mendeteksi pasien yang benar-benar menderita tuberkulosis, sehingga berpotensi mengurangi risiko keterlambatan diagnosis, sementara nilai presisi yang cukup tinggi menunjukkan kemampuan model dalam menekan kesalahan klasifikasi positif palsu. Meskipun algoritma Naive Bayes dibangun berdasarkan asumsi independensi antar atribut, hasil penelitian membuktikan bahwa asumsi tersebut tidak secara signifikan memengaruhi kinerja klasifikasi pada data medis yang digunakan. Selain itu, kesederhanaan model, efisiensi komputasi, serta kemudahan implementasi menjadikan algoritma Naive Bayes layak diterapkan dalam pengolahan data pasien tuberkulosis. Dengan demikian, penelitian ini menegaskan bahwa algoritma Naive Bayes masih relevan dan dapat dijadikan sebagai metode alternatif yang andal dalam pengembangan sistem pendukung keputusan berbasis data mining di bidang kesehatan

## REFERENSI

- [1] J. S. Sipayung, W. Hidayat, and E. M. Silitonga, "Faktor Risiko yang Memengaruhi Kejadian Tuberkulosis ( TB ) Paru di Wilayah Kerja Puskesmas Perbaungan Risk Faktors Affecting the Incident of Pulmonary Tuberculosis ( TB ) in the Working Area of Perbaungan Public Health Center," vol. 15, no. 2, 2023.
- [2] F. K. Masyarakat *et al.*, "https://doi.org/10.36729," vol. 7, pp. 78–88, 2022.
- [3] T. Aprilia, T. Informatika, and U. S. Sri, "Klasifikasi Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor dan Metode Naive Bayes," vol. 4, no. 2, pp. 156–163, 2024, doi: 10.54259/satesi.v4i2.3167.
- [4] G. Satya Nugraha, M. Nurkholis Abdullah, and M. Innuddin, "Komparasi Akurasi Metode Correlated Naive Bayes Classifier Dan Naive Bayes Classifier Untuk Diagnosis Penyakit Diabetes."
- [5] N. Bayes, C. Dan, and R. Forest, "Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine ( SVM ), Sentiment Analysis Of The Sirekap 2024 Playstore Application Post-Presidential Election With Comparison Of Support Vector Machine ( SVM ), Naive Bayes Classifier , And Random Forest Methods .," vol.

- 11, no. 3, pp. 661–670, 2025.
- [6] S. Pemanfaatan, T. Data, U. Analisis, D. Kesehatan, and D. I. Klinik, “Jurnal Abdimas Saintika,” pp. 181–186.
- [7] A. A. Ningtyas, A. Solichin, and R. Pradana, “Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritme Sentiment Analysis Of Youtube Comments On Prediction Of Economic Recession In 2023 Using The Naïve Bayes,” vol. 20, no. 1, pp. 9–16, 2023.
- [8] I. Kononenko, “Machine Learning for Medical Diagnosis : History , State of the Art and Perspective Historical overview,” pp. 1–25.
- [9] D. D. Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat ( Dpr ) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” vol. 10, no. 1, pp. 34–40, 2022.
- [10] D. A. Faroek, M. Yusuf, and G. Syatauw, “A s t p p c 2024 t m a n b c,” vol. 17, no. 2, pp. 216–226, 2024.
- [11] L. R. Krosuri, R. Satish, S. Fan, and J. Yao, “Extraction Sentiment Analysis Using naive Bayes Algorithm and Reducing Noise Word applied in Indonesian Language Extraction Sentiment Analysis Using naive Bayes Algorithm and Reducing Noise Word applied in Indonesian Language”, doi: 10.1088/1757-899X/835/1/012051.
- [12] Y. Azhar, A. K. Firdausy, and P. J. Amelia, “Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke,” vol. 5, no. 2, pp. 191–197, 2022.
- [13] F. M. Julianto, A. T. Zy, and E. Rilvani, “Sentiment Analysis on Canva Reviews Using Naive Bayes Method,” *Int. J. Informatics Comput.*, vol. 7, no. 1, 2025, doi: 10.35842/ijicom.
- [14] G. H. Hilmawan, U. S. April, and K. Sumedang, “Literatur Review : Efektifitas Penerapan Metode,” vol. 3, no. 6, 2025.
- [15] T. Pantai, K. Jepara, M. A. Anwar, H. Mulyo, and T. Tamrin, “Optimalisasi Algoritma Naive Bayes Dengan Teknik Ensemble Dalam Analisis Sentimen,” *J. Minfo Polgan*, vol. 13, no. 1, 2024, doi: 10.33395/jmp.v13i1.14014.
- [16] U. N. Putra, S. Media, A. Sentimen, and N. Bayes, “Systematic Literature Review ( Slr ): Analisis Sentimen Pemilihan Calon Presiden 2024 Menggunakan Metode,” 2024.
- [17] F. Rambu, B. Kahi, A. C. Talakua, and R. T. Abineno, “Analisis Sentimen Masyarakat Di Twitter Terhadap Pemerintahan Anies Baswedan Menggunakan Metode Naive Bayes Classifier,” vol. 13, no. April, pp. 324–336, 2024.
- [18] R. Nurzuli, “Lung Diseases Classification Using the Naïve Bayes Algorithm,” vol. 7, no. 2, 2025, doi: 10.35842/ijicom.
- [19] “The Indonesian Journal of Health Promotion MPPKI Media Publikasi Promosi Kesehatan Indonesia Analisis Implementasi Strategi Promosi Kesehatan dalam Pencegahan Penyakit Tuberkulosis (TB) (Studi Kasus di Wilayah Kerja Puskesmas Kalumata Kota Ternate),” 2022, doi: 10.31934/mppki.v2i3.
- [20] Y. A. Rizky, A. Aziz, and W. Harianto, “Implementasi Naive Bayes Dengan Menggunakan Metode Laplace Smoothing,” *RAINSTEK J. Terap. Sains dan Teknol.*, vol. 6, no. 3, pp. 164–172, Sep. 2024, doi: 10.21067/jtst.v6i3.9132.