



EVALUASI KINERJA ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI PENYAKIT KANKER PARU

Fadhilah Itsnani Nur Syamsiyah^{a*}, Hasbi Firmansyah^b

^a Fakultas Teknik dan Ilmu Komputer / Jurusan Informatika; fadhilahitsnani99@gmail.com, Universitas Pancasakti Tegal; Kota Tegal, Jawa Tengah 52121

^b Fakultas Teknik dan Ilmu Komputer / Jurusan Informatika; hasbifirmansyah@upstegal.ac.id, Universitas Pancasakti Tegal; Kota Tegal, Jawa Tengah 52121

* Penulis Korespondensi: Fadhilah Itsnani Nur Syamsiyah

ABSTRACT

One of the leading causes of death worldwide is lung cancer, with smoking being the biggest risk factor contributing to nearly 80% of cases. Exposure to carcinogens such as radon, asbestos, and air pollution also increases the risk of developing this disease. Using a dataset that includes various medical factors, this study attempts to apply the Naïve Bayes algorithm to lung cancer classification. Naïve Bayes was chosen because of its high accuracy in classifying difficult data. The results of the study show that Naïve Bayes has an accuracy rate of 92.93% in identifying lung cancer, which is influenced by the quantity of related features and data quality. This algorithm can help detect and prevent lung cancer early, thereby supporting the development of more effective strategies and early diagnosis. The results of this study are expected to be used as a reference in the development of data mining-based medical decision support systems, particularly for the early detection of lung cancer.

Keywords: Data Mining; Naïve Bayes; Classification; Cancer; Lung

Abstrak

Salah satu penyebab utama kematian di dunia adalah kanker paru-paru, dengan merokok sebagai faktor risiko terbesar yang berkontribusi pada hampir 80% kasus. Paparan terhadap zat karsinogen seperti radon, asbes, dan polusi udara juga meningkatkan risiko terkena penyakit ini. Dengan menggunakan dataset yang mencakup berbagai faktor medis, penelitian ini mencoba menerapkan algoritma *Naïve Bayes* pada klasifikasi kanker paru-paru. *Naïve Bayes* dipilih karena akurasi yang tinggi dalam mengklasifikasikan data yang sulit. Hasil penelitian menunjukkan bahwa *Naïve Bayes* memiliki tingkat akurasi 92,93% dalam mengidentifikasi kanker paru-paru, yang dipengaruhi oleh kuantitas fitur terkait dan kualitas data. Algoritma ini dapat membantu mendeteksi dan mencegah kanker paru-paru sejak dini, sehingga dapat mendukung pengembangan strategi dan diagnosis dini yang lebih efektif. Hasil penelitian ini diharapkan dapat digunakan sebagai acuan dalam pengembangan sistem pendukung keputusan medis berbasis *data mining*, khususnya untuk mendeteksi kanker paru-paru secepat mungkin.

Kata Kunci: Data Mining; Naïve Bayes; Klasifikasi; Kanker; Paru-paru

1. PENDAHULUAN

Kesehatan setiap orang sangat vital dan tak ternilai harganya karena itu adalah hal berharga yang tidak dapat dibeli oleh siapa pun. Salah satu organ vital dalam tubuh manusia yang secara signifikan mempengaruhi kesehatan adalah paru-paru. Paru-paru memiliki peran penting dalam system pernapasan manusia [1]. National Cancer Institute mendefinisikan kanker sebagai penyakit keturunan yang disebabkan oleh perubahan gen yang mengatur aktivitas sel, terutama kemampuan sel untuk membelah dan berkembang biak. [2].

The Global Burden of Cancer memperkirakan bahwa pada tahun 2020, terdapat sekitar 10 juta kematian terkait kanker dan 19,3 juta kasus baru penyakit tersebut. Kanker paru-paru merupakan penyebab kematian terbesar dengan 1,7 juta kematian, diikuti oleh kanker prostat (375 ribu) dan kanker kolorektal (935 ribu). Kanker paru-paru merupakan penyakit kanker yang paling umum di Indonesia, di mana penyakit tersebut menyebabkan sekitar 34.000 kasus baru dan 30.000 kematian pada tahun 2020. [2].

Di banyak negara, kanker adalah kondisi yang bisa berakibat fatal. Organisasi Kesehatan Dunia memperkirakan bahwa 9,6 juta orang di seluruh dunia kehilangan nyawa mereka karena kanker. Kanker paru-paru menyumbang 13% dari semua diagnosis kanker dan merupakan penyebab utama kematian ganas secara global. [3]. Salah satu jenis kanker paru-paru dihasilkan dari perubahan genetik pada sel epitel saluran napas yang mengakibatkan proliferasi sel yang tidak terkendali. Kanker ini dapat bermula dari organ paru-paru itu sendiri (primer) atau menyebar ke bagian tubuh lain (metastasis) [4]. Menurut Organisasi Kesehatan Dunia (WHO), kanker paru-paru merupakan salah satu penyebab utama kematian pada pria maupun wanita dari segala usia.

Berdasarkan data Globocan tahun 2018, proporsi kasus kanker paru-paru secara global mencapai 11,6%, sementara tingkat kematiannya yaitu sebesar 18,4%. Di Indonesia, kasus kanker paru-paru tercatat sebesar 8,6% atau sekitar 30.023 kasus, dengan tingkat kematian 12,6% atau setara dengan 26.095 orang. Kanker paru-paru juga berkontribusi terhadap sejumlah kematian.

Diagnosis penyakit telah memanfaatkan penambangan data dengan baik. Penambangan data adalah proses menganalisis data menggunakan komputer untuk menemukan pola, tren masyarakat, dan hukum mekanis tertentu. Koneksi antara data yang sebelumnya tidak terlihat ditemukan dengan menggunakan metode ini [2]. Metode klasifikasi adalah proses mengidentifikasi model atau fungsi yang menggambarkan atau membedakan kelas data. Metode ini membuat model yang dapat memisahkan data ke dalam kelas sesuai dengan aturan atau fungsi tertentu, memungkinkan objek yang tidak disebutkan namanya [8].

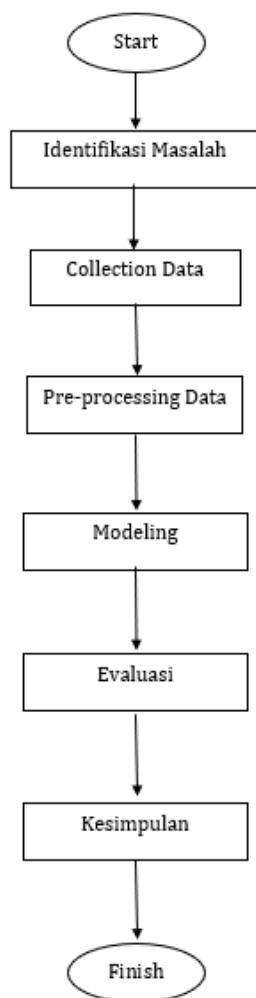
Sejarah ini menginformasikan peningkatan tantangan dalam penelitian ini, yang merupakan penilaian kinerja algoritma *Naïve Bayes* untuk mengklasifikasikan kanker paru-paru. Klasifikasi data yang akurat dan sangat akurat adalah tujuan dari implementasi metode *Naïve Bayes*. Studi ini juga menjelaskan potensi aplikasi teknologi pembelajaran mesin di bidang medis, khususnya dalam klasifikasi kanker paru-paru.

Di era digitalisasi, pentingnya teknologi berbasis data dalam bidang medis semakin meningkat. Pola tersembunyi dalam data medis yang sangat besar dapat ditemukan melalui teknik data mining, yang membantu pengambilan keputusan menjadi lebih baik. Penelitian terdahulu menunjukkan bahwa berbagai algoritma efektif dalam klasifikasi penyakit. Namun, sebagian besar penelitian berkonsentrasi pada algoritma seperti *Decision Tree* atau *k-Nearest Neighbor (k-NN)*. Oleh karena itu, penelitian ini mengevaluasi kinerja *Naïve Bayes* secara khusus dalam klasifikasi kanker paru-paru untuk mengisi kesenjangan penelitian.

Penelitian terdahulu juga menunjukkan efektivitas algoritma *Naïve Bayes* dalam berbagai kasus klasifikasi medis. Yunianto et al. [10] menerapkan metode ini dengan kombinasi fitur *Gray Level Co-occurrence Matrix (GLCM)* untuk mendeteksi kanker paru-paru dan memperoleh tingkat akurasi yang tinggi. Naezer dan Supriyanto [11] membandingkan performa *Naïve Bayes* dengan *k-Nearest Neighbor (k-NN)* dan mendapati bahwa *Naïve Bayes* lebih stabil pada data berukuran kecil dengan distribusi variabel yang terbatas. Permana et al. [12] serta Juliani dan Soleh [13] menegaskan potensi pendekatan data mining berbasis pembelajaran mesin dalam diagnosis dini kanker paru-paru. Wulandari dan Perdana [14] juga menyoroti keunggulan *Naïve Bayes* dalam efisiensi pengolahan data medis berukuran besar dan kompleks. Berdasarkan studi-studi tersebut, penelitian ini memperkuat landasan penggunaan algoritma *Naïve Bayes* sebagai model yang efektif untuk klasifikasi penyakit kanker paru-paru.

2. METODOLOGI PENELITIAN

Penelitian ini menerapkan algoritma *Naïve Bayes* untuk mengklasifikasikan kanker paru-paru melalui beberapa tahap guna mengatasi masalah yang dihadapi. Data yang telah diformat akan diklasifikasikan oleh algoritma *Naïve Bayes* dengan mengikuti langkah-langkah yang telah ada. Proses klasifikasi data bergantung pada prediksi yang didasarkan pada data tersebut. Berikut adalah metode penelitian yang digunakan untuk klasifikasi kanker paru-paru.



Gambar 1. Metode Penelitian

Gambar 1 menampilkan alur metode penelitian yang mencakup tahapan pengumpulan data, pre-processing, pembagian data, penerapan algoritma *Naïve Bayes*, dan evaluasi hasil.

2.1. Identifikasi Masalah

Kanker paru-paru termasuk salah satu kondisi medis yang paling umum. Menurut Organisasi Kesehatan Dunia (WHO), penyakit ini diketahui menewaskan lebih banyak orang dibandingkan penyakit lain, termasuk wanita dewasa dan anak-anak. Di era teknologi maju saat ini, mendiagnosis atau mendeteksi tuberkulosis (TB) membutuhkan alat yang dapat menyimpan data medis pribadi seseorang, terutama yang berkaitan dengan fitur kesehatan, sehingga pengobatan dapat dimulai segera setelah penyakit terdeteksi [6].

Berdasarkan penelitian objektif, mengungkapkan bahwa ada hubungan antara merokok dengan peningkatan risiko kanker paru-paru. Data terkini menyatakan bahwa merokok merupakan faktor penyebab utama dengan jumlah sekitar 87% kematian akibat kanker paru-paru. Namun, upaya deteksi dini untuk penyakit ini terkadang terhambat oleh ketiadaan gejala awal yang jelas dan teknik skrining yang kurang memadai [5].

Penulis bertujuan untuk mengembangkan model pembelajaran mesin yang dapat memprediksi tingkat keparahan pasien (*Low, Medium, High*) berdasarkan risiko dan gejala yang termasuk dalam dataset dengan menggunakan data medis yang berisi komponen penting termasuk riwayat medis, temuan laboratorium, dan pencitraan medis. Diharapkan penelitian ini akan membantu otoritas kesehatan dan profesional medis mencegah dan mendeteksi kanker paru-paru sejak dini berdasarkan tingkat risiko.

2.2. Collection Data (Pengumpulan Data)

Dalam studi ini, penulis menggunakan dataset lung cancer yang diperoleh dari platform website Kaggle melalui tautan <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link>. Dataset mengenai penyakit kanker paru-paru ini memiliki jumlah sebanyak 26 atribut dalam format CSV (*Comma Separated Values*) berupa angka. Banyak faktor penting, termasuk ID pasien, usia, jenis kelamin, risiko pekerjaan, alergi debu, konsumsi alkohol, tingkat polusi udara, kerentanan genetik, dan kondisi paru-paru kronis, termasuk dalam dataset tersebut.

Tabel 1. Tipe Data

No.	Nama Variabel	Tipe Data
1	index	Integer
2	Patient ID	Nominal
3	Age	Integer
4	Gender	Integer
5	Air Pollution	Integer
6	Alcohol use	Integer
7	Dust Allergy	Integer
8	Occupational Hazards	Integer
9	Genetic Risk	Integer
10	Chronic Lung Disease	Integer
11	Balance Diet	Integer
12	Obesity	Integer
13	Smoking	Integer
14	Passive Smoker	Integer
15	Chest Pain	Integer
16	Coughing of Blood	Integer
17	Fatigue	Integer
18	Weight Loss	Integer
19	Shortness of Breath	Integer
20	Wheezing	Integer
21	Swallowing Difficulty	Integer
22	Clubbing of Finger Nails	Integer
23	Frequent Cold	Integer
24	Dry Cough	Integer
25	Snoring	Integer
26	Level	Nominal

2.3. Pre-processing Data

Proses menyiapkan data mentah untuk langkah-langkah pemrosesan lebih lanjut dikenal sebagai pre-processing data. Untuk membuat data lebih terstruktur, fase pre-processing data menghilangkan data yang berisi nilai nol atau kosong dan melakukan transformasi dan pemurnian data. Tahap ini meningkatkan struktur data yang diproses dan merampingkan proses pemodelan.

Dataset ini mencakup 26 fitur, yang dibagi menjadi dua jenis data: integer dan objek. Data integer merujuk pada data numerik dalam bentuk bilangan bulat, baik positif, nol, maupun negatif, tanpa melibatkan desimal atau pecahan. Di sisi lain, entitas dunia nyata dengan atribut atau operasi tertentu direpresentasikan oleh tipe data objek. Misalnya, dalam dataset ini, Patient ID berfungsi sebagai identifikasi unik untuk membedakan satu pasien dari pasien lainnya, sementara Level menunjukkan kategori kanker paru-paru pada seorang pasien. Atribut yang dikecualikan dari analisis adalah Index dan Patient ID, karena kedua atribut tersebut tidak dapat menyajikan data yang signifikan dalam proses prediksi kanker paru-paru. "Level" adalah kolom yang harus digunakan sebagai label untuk model analisis dan prediksi *Rapid Miner*. Kolom ini menyimpan referensi utama untuk aktivitas prediksi, seperti kategorisasi atau tingkat keparahan kanker paru-paru. Pemilihan label akan sangat penting karena model akan menggunakan input dari fitur lain untuk mengantisipasi tujuan ini.

2.4. Split Data

Setelah tahap pre-processing, langkah selanjutnya adalah membagi data menjadi data latih dan data uji. Proses segmentasi kumpulan data untuk tujuan pengujian dan pelatihan model dikenal sebagai "pemisahan

data". Secara konvensional, dataset akan dibagi menjadi dua komponen utama yaitu data pelatihan, yang digunakan untuk melatih model, dan data pengujian, yang digunakan untuk mengevaluasi kinerja model setelah pelatihan. Perlu dicatat bahwa mekanisme pelatihan data secara fundamental berbeda dari proses pengujian. Dalam studi ini, penulis menerapkan rasio 80:20, dengan 80% dialokasikan untuk pelatihan dan 20% untuk pengujian. Rasio 80:20 ini dipilih karena memberikan keseimbangan antara jumlah data pelatihan yang cukup untuk membangun model serta menilai performa generalisasi model secara akurat. Selain itu, rasio 80:20 sering digunakan dalam studi machine learning karena menjaga keseimbangan representasi data di kedua subset sekaligus mengurangi risiko *overfitting* yang sering terjadi ketika nilai dataset pelatihan terlalu besar. Dengan mencegah *overfitting*, pendekatan ini menjamin model dapat beradaptasi dengan efektif terhadap data baru.



Gambar 2. Proses Split Data

Gambar 2 menunjukkan proses pembagian data dengan rasio 80:20 untuk membangun dan mengevaluasi kinerja model.

2.5. Naïve Bayes

Algoritma *Naïve Bayes* mengembangkan model prediktif untuk klasifikasi dengan memanfaatkan konsep probabilitas. Sebagai metode pembelajaran yang terkontrol, *Naïve Bayes* memerlukan dataset pelatihan untuk menghasilkan perkiraan selama fase pembelajaran. Pada tahap klasifikasi, nilai probabilitas untuk setiap kelas target dihitung berdasarkan input yang diberikan. Kelas target dengan probabilitas tertinggi kemudian ditetapkan sebagai kelas untuk data input. Keunggulan utama algoritma *Naïve Bayes* terletak pada kecepatan dan efisiensinya dalam memproses dataset berskala besar [7].

2.6. Data Mining

Proses penyortiran data menggunakan metode statistik, matematika, dan teknologi untuk mengidentifikasi korelasi, pola, dan tren baru yang penting dikenal sebagai *Data Mining* [7]. *Data Mining* menurut Muflikhah (2018), adalah pemecahan rumit dari kumpulan data menjadi informasi yang sebelumnya tidak ditemukan dengan potensi implisit (tidak nyata atau jelas) [1].

Akibatnya, penambahan pengetahuan atau penemuan pengetahuan adalah nama lain untuk penambahan data. Penambahan data telah menjadi metode pilihan bagi banyak orang, meskipun ada perbedaan dalam nomenklatur dan makna. Nama lain dari *data mining* termasuk pengambilan informasi, pengenalan pola,

arkeologi data, analisis pola, dan ekstraksi pengetahuan. Kumpulan fakta atau informasi yang tidak berguna yang dicatat sebagai data. Oleh karena itu, proses yang menghasilkan hasil berbasis pengetahuan dapat dicirikan sebagai penambangan data [9].

2.7. Rapid Miner

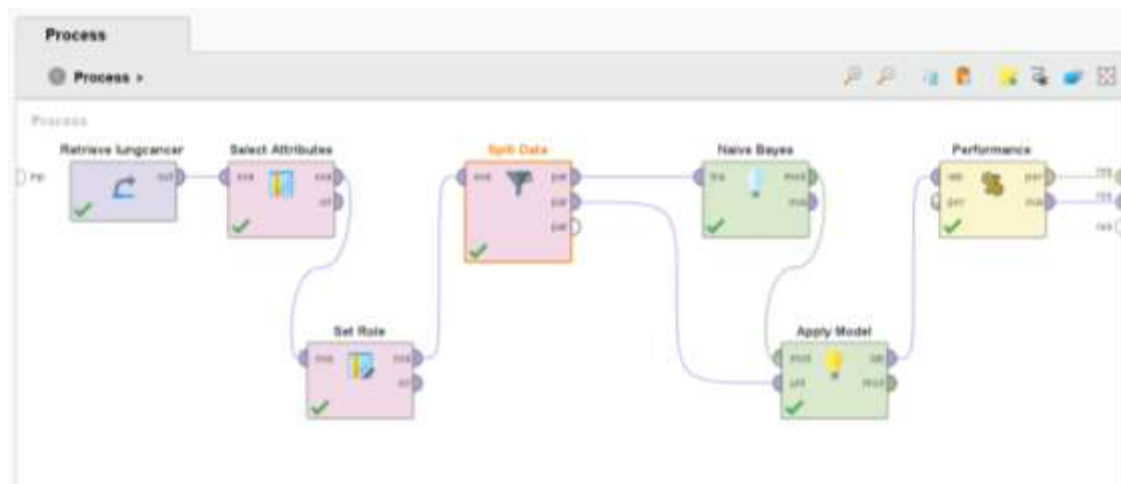
RapidMiner adalah platform analitik berbasis Java yang komprehensif, dikembangkan oleh perusahaan yang sama untuk mendukung aktivitas penambangan data, penambangan teks, analitik prediktif, dan analitik bisnis. Platform ini telah digunakan dalam berbagai aplikasi dan diakui sebagai pemimpin industri serta solusi mandiri dan sumber terbuka yang paling banyak dicari [6]. *RapidMiner* dipilih karena kemampuannya untuk menganalisis data melalui antarmuka visual, mengurangi kebutuhan penulisan kode, serta menyediakan dukungan efisien terhadap berbagai algoritma machine learning.

3. HASIL DAN PEMBAHASAN

Data dibagi menjadi 80% untuk data pelatihan serta 20% untuk data pengujian ketika diproses menggunakan algoritma *Naïve Bayes* dengan metode pemisahan data 80:20. Setelah pelatihan model, dilakukan evaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall* dan *F1-score*. Langkah-langkah dalam penerapan algoritma *Naïve Bayes* melalui aplikasi *RapidMiner* memanfaatkan data pelatihan yang tersedia di dalam dataset untuk mengklasifikasikan penyakit kanker paru-paru.

3.1. Implementasi Rapid Miner

Dalam proses implementasi menggunakan *RapidMiner*, dataset akan dikelola untuk menghasilkan tingkat akurasi berdasarkan kinerja data yang tersedia. Proses ini dimulai dengan operator *Retrieve*, yang mengimpor dataset dari file CSV. Selanjutnya, operator *Select Attributes* diterapkan untuk menentukan kolom mana yang akan dihapus. *Split Data* kemudian dimanfaatkan untuk membagi menjadi data pelatihan dan data pengujian. Akhirnya, menggunakan data pelatihan, model klasifikasi dikembangkan melalui operator *Naïve Bayes*. Model ini kemudian diterapkan pada data pengujian menggunakan operator *Apply Model* untuk menghasilkan prediksi. Terakhir, performa model dievaluasi dengan operator *Performance*, yang memberikan metrik evaluasi seperti akurasi untuk menilai kualitas prediksi.



Gambar 3. Proses Modeling

Gambar 3 menunjukkan tahapan pemodelan di *RapidMiner*, mulai dari pengambilan dataset, seleksi atribut, pembagian data, pelatihan model *Naïve Bayes*, hingga evaluasi kinerja menggunakan operator *Performance*.

3.2. Apply Model

Setelah model dikembangkan menggunakan data pelatihan, operator diterapkan untuk menggunakan model tersebut pada data baru yang belum pernah ditemui sebelumnya.

Row No.	Level	predicted...	confidence...	confidence...	confidence...	Index	Patient ID	Age	Gender
1	High	High	0	0.000	1.000	4	P101	40	1
2	Low	Low	0.997	0.010	0	5	P102	52	2
3	Medium	High	0.000	0.001	0.999	12	P103	35	2
4	Medium	High	0.000	0.001	0.999	18	P111	73	1
5	High	High	0	0.000	1.000	17	P112	34	1
6	High	High	0	0.000	1.000	20	P116	34	1
7	High	High	0	0.000	1.000	23	P118	38	2
8	Low	Low	0.999	0.010	0	33	P128	33	1
9	Low	Medium	0.000	1.000	0.000	36	P132	40	2
10	Low	Low	1.000	0.000	0	41	P135	26	2
11	Low	Low	1.000	0.000	0	42	P136	40	1
12	High	High	0	0.000	1.000	46	P14	34	1
13	Low	Low	1.000	0.000	0	52	P145	24	2

Gambar 4. Apply Model

Gambar 4 menunjukkan proses penerapan model *Naïve Bayes* yang telah dilatih terhadap data baru untuk menghasilkan prediksi tingkat risiko kanker paru-paru.

3.3. Evaluasi

PerformanceVector dalam *RapidMiner* merupakan output yang berisi daftar nilai kriteria kinerja dari model prediksi. Ini dihitung berdasarkan atribut label dan prediksi dari dataset input. *PerformanceVector* mencakup metrik seperti *accuracy*, *precision*, *recall*, dan kesalahan klasifikasi.

	true Low	true Medium	true High	class precision
pred. Low	29	0	0	100.00%
pred. Medium	1	27	0	96.43%
pred. High	0	6	36	85.71%
class recall	96.67%	81.82%	100.00%	

Gambar 5. Hasil Performance Accuracy

Hasil *Performance Accuracy* menunjukkan tingkat ketepatan model dalam klasifikasi kanker paru-paru berdasarkan dataset uji.

	true Low	true Medium	true High	class precision
pred. Low	29	0	0	100.00%
pred. Medium	1	27	0	96.43%
pred. High	0	6	36	85.71%
class recall	96.67%	81.82%	100.00%	

Gambar 6. Hasil Performance Kappa

Nilai *Kappa* menunjukkan tingkat kesepakatan antara hasil prediksi model dengan data aktual.

Dalam konteks studi ini, penulis menganalisis prediksi terhadap kanker paru-paru dengan menerapkan algoritma *Naïve Bayes*. Tingkat akurasi yang diperoleh dalam studi ini mencapai 92,93%, yang menunjukkan efektivitas algoritma *Naïve Bayes* dalam mengklasifikasikan kanker paru-paru.

Model dievaluasi dengan menggunakan pembagian data sebesar 80:20. Penelitian menunjukkan bahwa algoritma *Naïve Bayes* mencapai tingkat akurasi sebesar 92,93 persen. Angka ini sedikit lebih unggul dibandingkan dengan hasil penelitian Yuniyanto et al. [10], yang mencatat akurasi 91,5 persen melalui kombinasi fitur GLCM, sehingga membuktikan bahwa metode ini memiliki daya saing yang kuat dalam konteks diagnosis kanker paru-paru. Nilai kappa yang diperoleh, yakni 0,893, mencerminkan tingkat kesepakatan yang sangat kuat antara prediksi model dan data observasi, sehingga keandalan model dapat dinilai sangat memadai. Kinerja optimal tercapai pada kategori "*Low*", sedangkan kategori "*Medium*" dan "*High*" menunjukkan akurasi yang lebih bervariasi dan cenderung lebih rendah. Kondisi ini kemungkinan besar dipicu oleh ketidakseimbangan distribusi data di antara kelas-kelas tersebut, yang berpotensi mempengaruhi hasil klasifikasi secara menyeluruh.

Hasil penelitian ini konsisten dengan beberapa studi terkini yang menerapkan algoritma *Naïve Bayes* dalam bidang medis. Yuniyanto et al. [10] melaporkan tingkat akurasi yang serupa dalam klasifikasi kanker paru-paru menggunakan optimasi fitur, sementara Nasution et al. [15] serta Hidayat dan Santosa [16] membuktikan bahwa *Naïve Bayes* tetap kompetitif dalam klasifikasi berbasis teks dan data medis. Kombinasi teknik preprocessing dan pemilihan atribut relevan terbukti berpengaruh besar terhadap peningkatan performa model. Dengan demikian, hasil penelitian ini sejalan dengan tren penggunaan *Naïve Bayes* dalam diagnosis penyakit kompleks dan memperkuat keandalannya sebagai metode klasifikasi berbasis probabilistik.

4. KESIMPULAN DAN SARAN

Dari hasil evaluasi penelitian, dapat disimpulkan bahwa klasifikasi algoritma *Naïve Bayes* menggunakan metode *Split Data* dapat memprediksi kelas dengan akurasi yang baik, yaitu sebesar 92,93%. Nilai kappa sebesar 0,893 juga menunjukkan tingkat kesepakatan yang tinggi antara hasil prediksi dengan nilai aktualnya. Metode ini berkinerja sangat baik dalam memprediksi kelas "*Low*" dengan presisi dan penarikan 100%. Meskipun performa di kelas "*Medium*" dan "*High*" secara keseluruhan masih cukup baik, masih ada peluang untuk berkembang. Berdasarkan penelitian dari proses evaluasi, algoritma *Naïve Bayes* telah terbukti efektif dalam menangani dataset medis berskala sedang yang mencakup fitur-fitur bersifat independen, sehingga sangat sesuai untuk diterapkan dalam sistem pendukung keputusan medis.

Penelitian ini adalah yang pertama membantu membuat model kategorisasi untuk kasus kanker paru-paru. Namun, penelitian lebih lanjut diperlukan untuk mengatasi keterbatasan model dan meningkatkan kinerja keseluruhannya. Diharapkan pada penelitian selanjutnya untuk dapat fokus pada penanganan ketidakseimbangan data, seleksi fitur, dan penggunaan algoritma klasifikasi yang lain. Penelitian selanjutnya direkomendasikan untuk melakukan perbandingan kinerja algoritma *Naïve Bayes* dengan metode alternatif seperti *Random Forest*, *Support Vector Machine*, atau *Decision Tree*, dengan tujuan mencapai model klasifikasi yang lebih optimal.

DAFTAR PUSTAKA

- [1] R. A. Karima dan Z. Fatah, "Implementasi Metode K-Nearest Neighbor untuk Klasifikasi Penyakit Paru-Paru pada Anak," *Jurnal Teknologi dan Sistem Informasi*, vol. 1, no. 6, pp. 10–17, 2024.
- [2] L. Sari, A. Romadloni, dan R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Jurnal Sains dan Teknologi Informasi*, vol. 14, no. 1, pp. 155–162, 2023.
- [3] D. Septhya, R. Rahmawati, dan M. N. Fauzi, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *Jurnal Informatika dan Komputer*, vol. 3, no. 1, pp. 15–19, 2023.
- [4] I. Buana dan D. A. Harahap, "Asbestos, Radon, dan Polusi Udara sebagai Faktor Risiko Kanker Paru pada Perempuan Bukan Perokok," *Jurnal Kesehatan Indonesia*, vol. 8, no. 1, pp. 1–16, 2022.
- [5] M. A. M. Mustofa, S. R. Pratama, dan F. A. Harun, "Penggunaan Algoritma KNN dalam Deteksi Awal Kanker Paru-Paru Menggunakan Data Medis," *Jurnal Informatika dan Komputer (JIK)*, vol. 8, pp. 485–493, 2024.

- [6] H. A. Fadhlurrahman, R. S. Dewi, dan A. T. Siregar, “Peptida Anti Kanker Membranolitik (ACPS) pada Kanker Paru-Paru Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor,” *Jurnal Bioinformatika Indonesia*, vol. 1, no. 3, pp. 1–11, 2023.
- [7] M. A. Jabbar, T. R. Rahman, dan S. S. Sherekar, “Komparasi Algoritma Decision Tree, Naïve Bayes, dan K-Nearest Neighbors dalam Klasifikasi Kanker Payudara,” *Jurnal Sistem Informasi dan Komputer*, vol. 14, no. 3, pp. 258–270, 2022.
- [8] E. Wulandari dan A. Perdana, “Klasifikasi Kanker Paru-Paru Menggunakan Metode Naïve Bayes,” *I-Robot: International Research on Big-Data and Computer Technology*, vol. 6, no. 2, pp. 20–24, 2022.
- [9] A. Y. Permana, H. N. Fazri, M. F. N. Athoilah, M. Robi, dan R. Firmansyah, “Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest,” *Jurnal Ilmiah Teknik Informatika dan Komunikasi (JUITIK)*, vol. 3, no. 2, pp. 27–41, 2023.
- [10] M. Yuniarto, S. Soeparmi, C. Cari, F. Anwar, D. N. Septianingsih, T. D. Ardyanto, dan R. F. Pradana, “Klasifikasi Kanker Paru-Paru Menggunakan Naïve Bayes dengan Variasi Filter dan Ekstraksi Ciri GLCM,” *Indonesian Journal of Applied Physics*, vol. 11, no. 2, pp. 256–264, Okt. 2021.
- [11] M. Naezer dan R. Supriyanto, “Analisis Kinerja Algoritma Naïve Bayes dan k-NN untuk Memprediksi Penyakit Kanker Paru,” *Jurnal Ilmiah Komputasi*, vol. 22, no. 2, pp. 213–224, Jun. 2023.
- [12] A. Y. Permana, H. N. Fazri, M. F. N. Athoilah, M. Robi, dan R. Firmansyah, “Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest,” *Jurnal Ilmiah Teknik Informatika dan Komunikasi (JUITIK)*, vol. 3, no. 2, pp. 27–41, Jul. 2023.
- [13] D. Juliani dan M. Soleh, “Implementasi Machine Learning untuk Klasifikasi Penyakit Kanker Paru Menggunakan Metode Naïve Bayes dengan Tambahan Fitur Chatbot,” *Jurnal IPTEK (JII)*, vol. 8, no. 2, pp. 12–17, Ags. 2024.
- [14] E. Wulandari dan A. Perdana, “Klasifikasi Kanker Paru-Paru Menggunakan Metode Naïve Bayes,” *I-Robot: International Research on Big-Data and Computer Technology*, vol. 6, no. 2, pp. 20–24, 2022.
- [15] Y. R. Nasution, Suhardi, dan I. H. Satrio, “Penerapan Algoritma Klasifikasi Naïve Bayes untuk Analisis Sentimen tentang Pemilu 2024,” *Jurnal Elektronika dan Komputer (ELKOM)*, vol. 17, no. 2, pp. 495–502, Des. 2024.
- [16] R. Hidayat dan L. Santosa, “Penerapan Model Klasifikasi Metode Naïve Bayes untuk Prediksi Penyakit Berdasarkan Data Medis,” *JURSISTEKNI*, vol. 4, no. 1, pp. 1–8, Jan. 2022.