

CESS

(Journal of Computer Engineering, System and Science)

Available online: <https://jurnal.unimed.ac.id/2012/index.php/cess>

ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Deteksi Kanker Paru Paru Berbasis *Mobile*: Integrasi *Machine Learning* dalam Aplikasi Nusaminer

Mobile-Based Lung Cancer Detection: Integrating Machine Learning in the Nusaminer Application

Deny Alfian¹, Fathimah Azzahro², Achmad Baroqah Pohan^{3*}, Besus Maulana Sulthon⁴, Yunita⁵

^{1,2,3,4,5}Sistem Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia
Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta

Email: ¹deny.al91@gmail.com, ²fathimahazzahra53@gmail.com, ³achmad.abq@bsi.ac.id,

⁴besus.bem@bsi.ac.id, yunita.ynt@bsi.ac.id

*Corresponding Author

ABSTRAK

Deteksi dini kanker paru-paru sangat penting untuk meningkatkan peluang kesembuhan. Penelitian ini membandingkan performa sembilan algoritma klasifikasi (*AdaBoost*, *ANN*, *Decision Tree*, *Gradient Boosting*, *KNN*, *Logistic Regression*, *Naïve Bayes*, *Random Forest*, dan *SVM*) menggunakan *RapidMiner* dan *NusaMiner*. Dataset berasal dari *Kaggle* dengan 309 data pasien dan 16 atribut terkait faktor risiko serta gejala klinis. Evaluasi dilakukan berdasarkan akurasi, *precision*, *recall*, *F1-score*, dan *AUC* pada rasio 70:30, 80:20, dan 90:10. Hasil menunjukkan *NusaMiner* memberikan akurasi lebih tinggi dan stabil dibanding *RapidMiner*. Algoritma *Gradient Boosting* dan *Random Forest* mencapai akurasi 100% pada beberapa rasio, sementara *Naïve Bayes* menunjukkan akurasi terendah di *RapidMiner*. Temuan ini menegaskan pentingnya pemilihan algoritma dan platform dalam meningkatkan efektivitas sistem deteksi kanker paru-paru berbasis *machine learning*.

Kata Kunci: kanker paru-paru; klasifikasi; prediksi; data mining; machine learning.

ABSTRACT

Early detection of lung cancer is crucial to enhance the chances of recovery. This research compares the performance of nine classification algorithms (*AdaBoost*, *ANN*, *Decision Tree*, *Gradient Boosting*, *KNN*, *Logistic Regression*, *Naïve Bayes*, *Random Forest*, and *SVM*) using *RapidMiner* and *NusaMiner*. The dataset comes from *Kaggle* with 309 patient data and 16 attributes related to risk factors and clinical symptoms. Evaluation was conducted based on accuracy, precision, recall, *F1-score*, and *AUC* at ratios of 70:30, 80:20, and 90:10. The results show that *NusaMiner* provides higher and more stable accuracy compared to *RapidMiner*. The



Gradient Boosting and Random Forest algorithms achieved 100% accuracy at several ratios, while Naïve Bayes showed the lowest accuracy in RapidMiner. These findings emphasize the importance of algorithm and platform selection in improving the effectiveness of machine learning-based lung cancer detection systems.

Keywords: *lung cancer; classification; prediction; data mining; machine learning.*

1. PENDAHULUAN

Kanker paru-paru adalah salah satu kanker paling mematikan, dengan 2,5 juta kasus baru dan 1,8 juta kematian global pada 2022 menurut WHO [1]. Di Indonesia, kanker ini menjadi penyebab kematian tertinggi akibat kanker, dengan 34.339 kematian dari 38.904 kasus baru [2]. Penyebab utamanya adalah merokok, serta polusi udara, bahan kimia, alkohol, genetika, dan obesitas [3], [4].

Deteksi dini kanker paru-paru sering terlambat, menyebabkan banyak kasus ditemukan pada stadium lanjut dengan peluang sembuh rendah [5]. Tingginya angka morbiditas dan mortalitas menekankan pentingnya pencegahan, deteksi dini, dan penanganan efektif [6].

Perkembangan teknologi, khususnya *Artificial Intelligence (AI)* dan *machine learning*, berperan penting dalam dunia medis dengan kemampuan mengenali pola kompleks pada data medis untuk memprediksi penyakit secara akurat. Teknologi ini telah dimanfaatkan dalam diagnosis pneumonia, tuberculosis, serta kanker paru-paru [7], [8], [9].

Andriska et al., telah melakukan penelitian deteksi kanker paru-paru menggunakan Google Colab, dengan model C4.5 menghasilkan nilai akurasi 87% [10]. Sementara itu Widyawati & Faradibah [6], menggunakan Phyton dengan model *Decision Tree*, *Naïve Bayes*, dan SVM nilai akurasi yang didapat 89% untuk *Decision Tree* dan *Naïve Bayes*, 87% untuk SVM. Sari et al., [11] menggunakan RStudio dengan model *Naïve Bayes* dan *Random Forest* menghasilkan 98,4% untuk *Random Forest* dan 95,1% *Naïve Bayes*. Sementara Rahmaeda et al., [4] melakukan penelitian serupa dengan Phyton model yang digunakan *Logistic Regression* dan SVM dengan akurasi 97,85% dan 95,70%. Penelitian-penelitian ini menunjukkan bahwa keberhasilan klasifikasi penyakit kanker paru-paru sangat dipengaruhi oleh pemilihan algoritma yang tepat, kualitas data yang digunakan, serta *platform* pemrosesan yang mendukung.

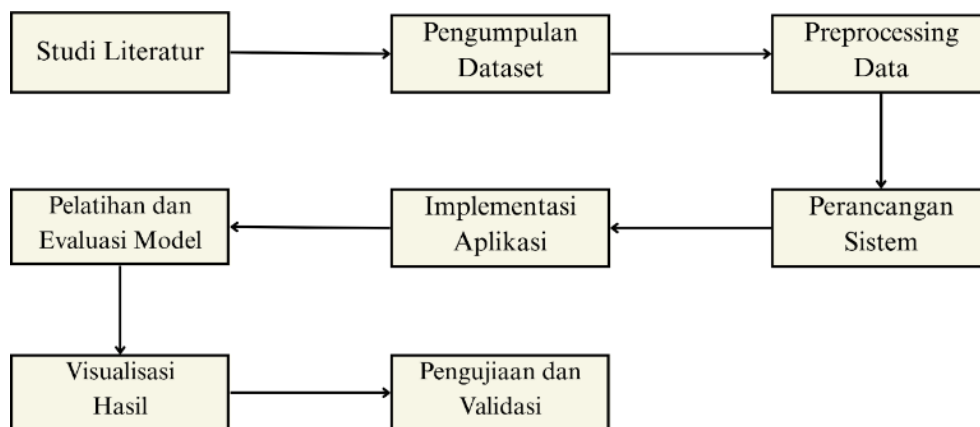
Penelitian ini bertujuan untuk mengkomparasi performa model prediksi kanker paru-paru menggunakan RapidMiner dan NusaMiner dengan sembilan algoritma klasifikasi, yaitu *AdaBoost*, *Artificial Neural Network (ANN)*, *Decision Tree*, *Gradient Boosting*, *K-Nearest Neighbors (KNN)*, *Logistic Regression*, *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine (SVM)*. Fokus penelitian meliputi analisis komparatif performa algoritma, evaluasi kemampuan generalisasi model berdasarkan variasi rasio data latih dan data uji, serta perancangan dan implementasi NusaMiner sebagai aplikasi *machine learning* berbasis *mobile* yang mendukung proses pelatihan, pengujian, evaluasi, dan prediksi secara terintegrasi untuk penggunaan yang lebih praktis.

Meskipun berbagai penelitian sebelumnya telah menerapkan *machine learning* untuk prediksi kanker paru-paru, sebagian besar masih terbatas pada jumlah algoritma yang sedikit, platform berbasis *desktop* atau *cloud*, serta evaluasi performa yang belum menekankan aspek

generalisasi dan implementasi praktis. Selain itu, perbandingan lintas-platform, khususnya antara platform visual data *mining* konvensional dan aplikasi *machine learning* berbasis *mobile*, masih jarang dilakukan. Oleh karena itu, penelitian ini memberikan kontribusi ilmiah melalui komparasi komprehensif sembilan algoritma klasifikasi pada dua platform berbeda, yaitu RapidMiner dan NusaMiner, guna menganalisis perbedaan performa, stabilitas, serta kemampuan generalisasi model pada lingkungan implementasi yang berbeda.

2. METODE PENELITIAN

Tahapan penelitian ini disusun secara sistematis mulai dari studi literatur hingga visualisasi hasil. Alur lengkap tahapan tersebut ditampilkan pada Gambar 1 berikut:



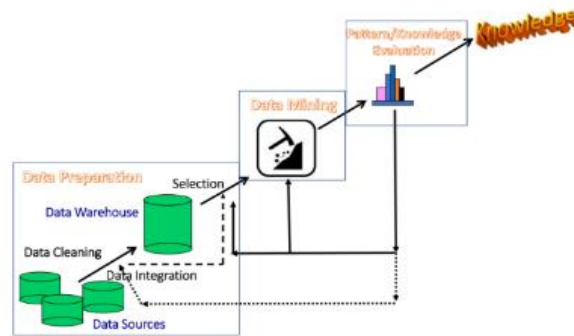
Gambar 1. Tahapan Penelitian

2.1. Studi Literatur

Paru-paru adalah organ utama pernapasan yang bertugas menukar oksigen dan karbon dioksida melalui alveoli, didukung oleh sistem peredaran darah, saraf, imun, dan otot [12]. Kanker paru-paru merupakan tumor ganas dari epitel saluran napas, terutama bronkus, yang berkembang akibat mutasi genetik dan menyebar melalui darah atau limfa (metastasis) serta membentuk pembuluh darah baru (angiogenesis) [2], [13], [14].

Kecerdasan buatan (*Artificial Intelligence/AI*) merupakan teknologi yang memungkinkan sistem komputer meniru kemampuan kognitif manusia, seperti belajar dan mengenali pola [15], [16]. Salah satu cabangnya, *machine learning* (ML), memungkinkan sistem belajar dari data tanpa pemrograman langsung dengan membangun model prediktif [17], [18]. Dalam medis, AI digunakan untuk menganalisis data seperti rekam medis dan citra radiologi, membantu diagnosis lebih cepat dan akurat [7].

Knowledge Discovery in Databases (KDD), diperkenalkan oleh Fayyad (1996), adalah proses sistematis untuk menemukan pola bermakna dalam data besar yang valid, inovatif, dan dapat diinterpretasikan [19], [20]. KDD mencakup tahapan seperti: *Data cleaning*, *Data transformation*, dan *Data mining*. Tahapan KDD secara visual ditampilkan pada gambar 2.



Gambar 2. Tahapan proses KDD

Penelitian ini menggunakan pendekatan *supervised learning*, yaitu pelatihan model dengan data berlabel untuk memetakan *input* ke *output* secara akurat [17], [18]. Model divalidasi menggunakan data uji, penting dalam diagnosis medis. Algoritma yang digunakan meliputi *AdaBoost*, *Artificial Neural Network (ANN)*, *Decision Tree*, *Gradient Boosting*, *K-Nearest Neighbors (KNN)*, *Logistic Regression*, *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine (SVM)*, yang masing-masing memiliki karakteristik dan mekanisme klasifikasi yang berbeda dalam memprediksi kanker paru-paru.

2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs Kaggle, yang menyediakan *dataset* terbuka terkait deteksi kanker paru-paru. *Dataset* yang digunakan terdiri dari 309 *record* dan 16 atribut yang merepresentasikan informasi pasien, termasuk usia, jenis kelamin, kebiasaan merokok, riwayat kesehatan, serta status diagnosis kanker paru-paru. Pada tabel 2 terdapat penjelasan lebih lanjut mengenai atribut dari *dataset* yang digunakan

Tabel 1. Detail *Dataset*

Fitur	Tipe	Keterangan
<i>Gender</i>	Kategori	M/F atau 0/1
<i>Age</i>	Numerik	Usia responden
<i>Smoking</i>	Kategori	Status Merokok (Yes/No)
<i>Yellow Finger</i>	Kategori	Kuku menguning (Yes/No)
<i>Anxiety</i>	Kategori	Kecemasan (Yes/No)
<i>Peer Pressure</i>	Kategori	Tekanan dari teman (Yes/No)
<i>Chronic Disease</i>	Kategori	Riwayat penyakit kronis (Yes/No)
<i>Fatigue</i>	Kategori	Kelelahan (Yes/No)
<i>Allergy</i>	Kategori	Alergi (Yes/No)
<i>Wheezing</i>	Kategori	Mengi (Yes/No)
<i>Alcohol Consuming</i>	Kategori	Konsumsi alkohol (Yes/No)
<i>Coughing</i>	Kategori	Batuk (Yes/No)
<i>Shortness of Breath</i>	Kategori	Sesak napas (Yes/No)
<i>Swallowing Difficulty</i>	Kategori	Kesulitan menelan (Yes/No)
<i>Chest Pain</i>	Kategori	Nyeri dada (Yes/No)
<i>Lung Cancer</i>	Target	Label target: diagnosis kanker paru (Yes/No)

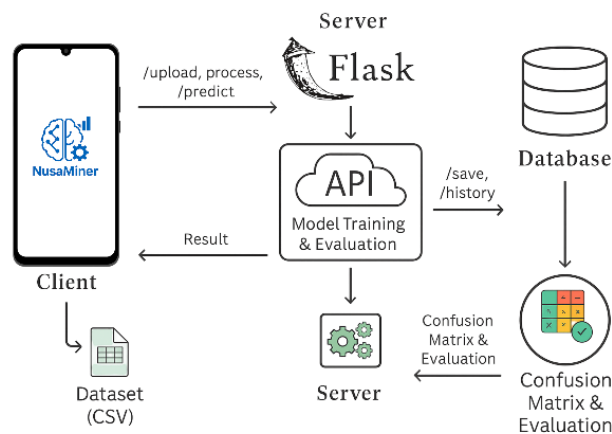
2.3. Preprocessing Data

Preprocessing data merupakan tahap penting dalam *Knowledge Discovery in Databases* (KDD) untuk memastikan kualitas data [21]. Langkah-langkah yang dilakukan mencakup:

1. *Data cleaning*: menghapus nilai kosong, duplikasi, dan *outlier* agar data bersih dan layak digunakan [22];
2. *Data transformation*: melakukan label *encoding* pada atribut kategorikal serta normalisasi fitur numerik menggunakan metode *min-max scaling* agar berada dalam rentang 0–1 [23];
3. *Splitting Data*, membagi data menjadi data latih dan data uji dengan rasio 70:30, 80:20, dan 90:10 untuk melihat pengaruh proporsi terhadap performa model.

2.4. Perancangan Sistem

NusaMiner, dibangun berbasis *mobile* menggunakan Flutter (*frontend*) dan Python Flask (*backend*). Sistem ini dirancang dengan arsitektur *client-server*, memungkinkan pengguna mengunggah data pasien, memilih model, dan mendapatkan hasil klasifikasi secara otomatis. Dengan pendekatan ini, NusaMiner diharapkan dapat memberikan kontribusi terhadap Solusi deteksi dini kanker paru-paru yang lebih efisien, praktis, dan dapat diterapkan secara luas di fasilitas pelayanan kesehatan. Gambar 3 menunjukkan visualisasi arsitektur NusaMiner.



Gambar 3. Arsitektur NusaMiner

2.5. Implementasi

Penelitian ini mengimplementasikan prediksi kanker paru-paru menggunakan dua aplikasi yaitu: RapidMiner dan NusaMiner. Gambar 4 *icon* aplikasi yang digunakan.



Gambar 4. Icon RapidMiner dan NusaMiner

2.6. Pelatihan dan Evaluasi Model

Penelitian ini menggunakan sembilan algoritma klasifikasi *supervised learning* yaitu *AdaBoost*, *Artificial Neural Network*, *Decision Tree*, *Gradient Boosting*, *K-Nearest Neighbors*,

Logistic Regression, Naïve Bayes, Random Forest, dan Support Vector Machine. Model dilatih menggunakan data latih dan diuji pada data uji, dengan evaluasi performa menggunakan: (1) Akurasi: proporsi prediksi benar terhadap total prediksi yang dilakukan, perhitungan akurasi dapat dilihat pada persamaan 1; (2) *Precision*: ketepatan prediksi positif dari total prediksi positif yang dilakukan persamaan *precision* pada persamaan 2; (3) *Recall*: sensitivitas model terhadap data positif dengan benar, *recall* dapat dicari dengan persamaan persamaan 3; (4) *F1-Score*: keseimbangan *precision* dan *recall*, cara menghitungnya pada persamaan 4; (5) *Support*: merupakan jumlah kejadian sebenarnya dari setiap kelas dalam *dataset*; (6) *Confusion Matrix*: tabel untuk membandingkan label prediksi dan label aktual, visualisasi *confusion matrix* dapat dilihat pada gambar 5; (7) *ROC Curve*: mengukur kemampuan model membedakan kelas secara keseluruhan; dan (8) *AUC* : memberikan gambaran kinerja model.

$$Accuracy = \frac{Correct\ Classification}{Total\ Classification} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

	Predicted	
Actual	TN	FP
	FN	TP

Gambar 5. *Confusion Matrix*

Dengan TP (*True Positive*) data positif yang diklasifikasikan benar sebagai positif; TN (*True Negative*) data negatif yang diklasifikasikan benar sebagai negatif; FP (*False Positive*) data negatif yang salah diklasifikasikan sebagai positif; dan FN (*False Negative*) data positif yang salah diklasifikasikan sebagai negatif.

2.7. Pengujian dan Validasi Hasil

Pengujian dilakukan untuk memastikan fungsi utama aplikasi berjalan dengan baik dan sesuai tujuan. Validasi model dilakukan menggunakan data uji untuk mengukur kemampuan generalisasi model terhadap data baru [18].

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data yang digunakan pada penelitian ini menggunakan data sekunder dari situs penyedia dataset yaitu Kaggle (<https://www.kaggle.com/datasets/samuelotiattakorah/lung-cancer-data/data>) terdiri dari 309 *record* dengan 16 atribut, mencakup berbagai faktor resiko dan gejala. Atribut target adalah atribut LUNG_CANCER dengan nilai YES dan NO. Gambar 6 menampilkan *dataset* yang digunakan.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	GENDER	AGE	SMOKIN G	YELLOW FINGERS	ANXIETY	PEER_ PRESSUR	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUM	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHEST PAIN	LUNG_ CANCER
1	M	69	1	2	2	1	1	2	1	2	2	2	2	2	2	2 YES
2	M	74	2	1	1	1	2	2	2	1	1	1	2	2	2	2 YES
3	F	59	1	1	1	2	1	2	1	2	1	2	2	1	2	2 NO
4	M	63	2	2	2	1	1	1	1	1	2	1	1	2	2	2 NO
5	F	63	1	2	1	1	1	1	1	2	1	2	2	1	1	1 NO
6	F	75	1	2	1	1	2	2	2	2	1	2	2	2	1	1 YES
7	M	52	2	1	1	1	1	2	1	2	2	2	2	2	1	2 YES
8	F	51	2	2	2	2	1	2	2	1	1	1	2	2	2	1 YES
9	F	68	2	1	2	1	1	2	1	1	1	1	1	1	1	1 NO
10	M	53	2	2	2	2	2	1	2	1	2	1	1	2	2	2 YES
11	F	61	2	2	2	2	2	2	1	2	1	2	2	2	2	1 YES
12	M	72	1	1	1	1	2	2	2	2	2	2	2	2	1	2 YES
13	F	60	2	1	1	1	1	2	1	1	1	1	2	1	1	1 NO
14	M	58	2	1	1	1	1	2	2	2	2	2	2	1	2	2 YES
15	M	69	2	1	1	1	1	1	2	2	2	2	1	1	2	2 NO
16	F	48	1	2	2	2	2	2	2	2	1	2	2	2	2	1 YES
17	M	75	2	1	1	1	2	1	2	2	2	2	2	1	2	2 YES
18	M	57	2	2	2	2	2	1	1	1	2	1	1	2	2	2 YES
19	F	68	2	2	2	2	2	2	1	1	1	2	2	1	1	1 YES
20	F	68	2	2	2	2	2	2	1	1	1	2	2	1	1	1 YES

Gambar 6. Dataset diagnosis pasien kanker paru-paru

Data yang telah diunduh selanjutnya akan melalui *preprocessing*, pelatihan model dan evaluasi algoritma *machine learning* dengan menggunakan RapidMiner dan NusaMiner.

3.2. Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data bersih, terstruktur, dan siap dianalisis. Proses meliputi: (1) clean data untuk menangani missing values, namun tidak diperlukan karena dataset bebas nilai kosong; (2) *convert categorical* dengan label encoding (contoh: “Male”=1, “Female”=0); (3) *normalize* atribut AGE menggunakan *Min-Max Normalization*; dan (4) splitting data menjadi *training set* dan *testing set* dengan rasio 70:30, 80:20, atau 90:10.

3.3. Implementasi

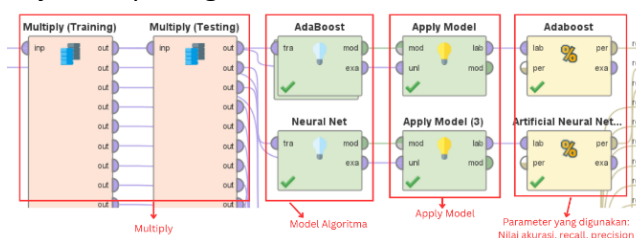
3.3.1. RapidMiner

Langkah *preprocessing* data menggunakan RapidMiner ditunjukkan pada gambar 7.



Gambar 7. Preprocessing Data pada RapidMiner

Operator *Multiply* digunakan agar dapat melakukan pengujian model lebih dari satu. Penambahan operator model dilakukan setelah *Multiply*, *Apply Model* agar model dapat digunakan, seperti ditunjukkan pada gambar 8.



Gambar 8. Penambahan algoritma menggunakan RapidMiner

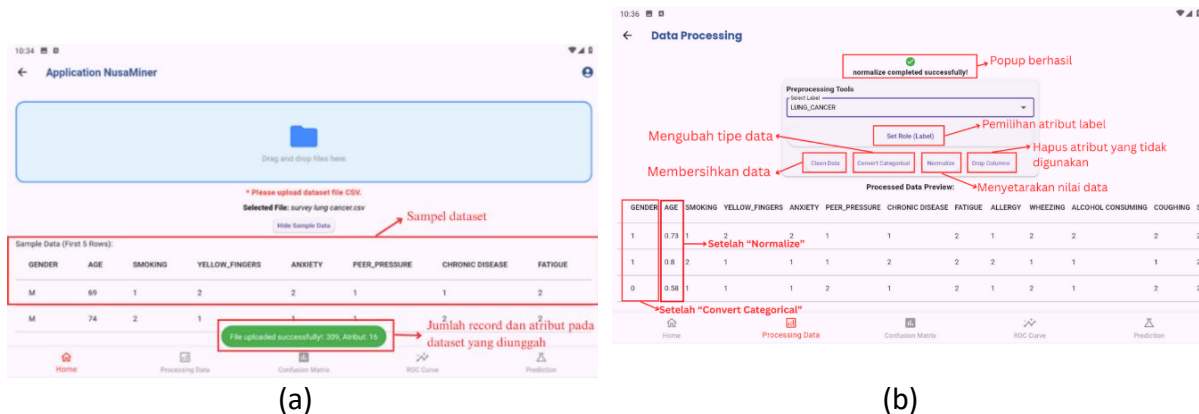
Hasil evaluasi model dapat diketahui dengan operator *Performance*, parameter yang digunakan adalah Akurasi, *recall*, dan *precision*. Hasil evaluasi dari Sembilan model ditampilkan pada gambar 9.

Model	Accuracy	Weighted Mean Recall	Weighted Mean Precision
Performance Vector (AdaBoost)	87.04%	66.35%	69.72%
Performance Vector (Artificial Neural Network)	86.28%	66.16%	77.60%
Performance Vector (Decision Tree (C))	83.31%	55.57%	40.80%
Performance Vector (Gradient Boosting)	84.11%	78.64%	69.55%
Performance Vector (K-Nearest Neighbor)	80.26%	66.95%	77.73%
Performance Vector (Logistic Regression)	80.57%	69.84%	62.10%
Performance Vector (Naive Bayes (G))	66.40%	73.04%	60.75%
Performance Vector (Random Forest (Z))	88.54%	67.04%	73.05%
Performance Vector (Support Vector Machine)	87.36%	51.61%	53.50%

Gambar 9. Hasil evaluasi model menggunakan RapidMiner

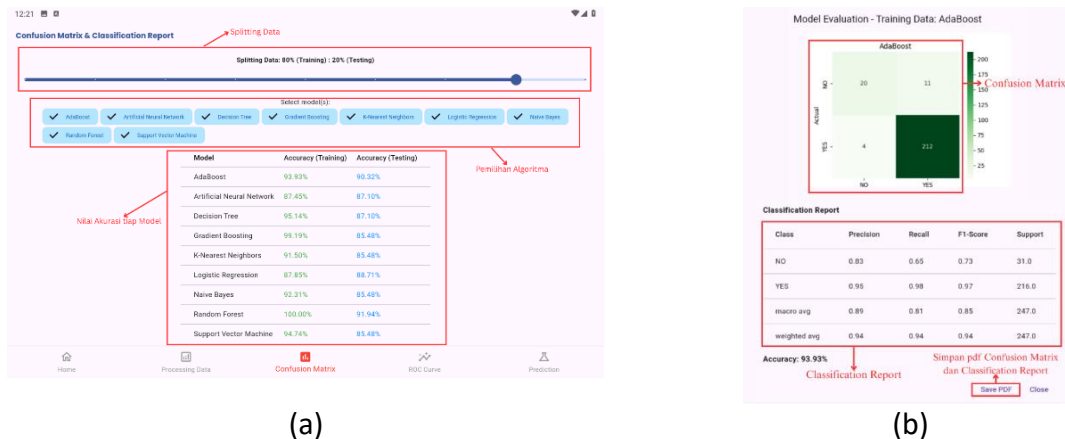
3.3.2.NusaMiner

Preprocessing data menggunakan NusaMiner ditunjukkan pada gambar dibawah ini. Sebelum melakukan *preprocessing* data kita perlu *upload dataset* terlebih dahulu pada halaman *home*, seperti gambar 10 (a). Setelah *dataset* di-*upload* maka masuk pada halaman *processing data* untuk melakukan *preprocessing*, seperti yang dijelaskan pada gambar 10 (b).



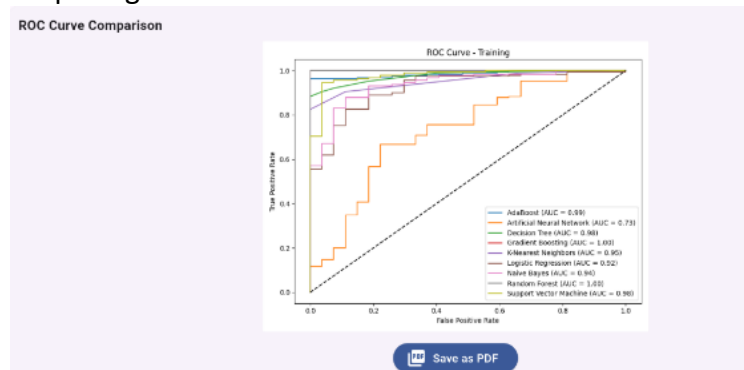
Gambar 10. (a) Halaman utama untuk *upload dataset* dan (b) *Preprocessing* menggunakan NusaMiner

Pemilihan model algoritma dilakukan pada halaman *Confusion Matrix*, sebagaimana ditunjukkan pada Gambar 11 (a), yang memungkinkan pengguna memilih lebih dari satu algoritma. Setelah algoritma dipilih, sistem secara otomatis melatih model menggunakan data *training* dan kemudian mengevaluasinya dengan data *testing* untuk mengukur kemampuan generalisasi. Hasil evaluasi disajikan dalam bentuk *Confusion Matrix* dan *Classification Report*, yang dapat diakses dengan memilih nilai akurasi pada model, sebagaimana ditampilkan pada Gambar 11 (b).



Gambar 11. (a) Pemilihan algoritma, *Splitting data*, dan nilai akurasi dan (b) Hasil evaluasi model NusaMiner

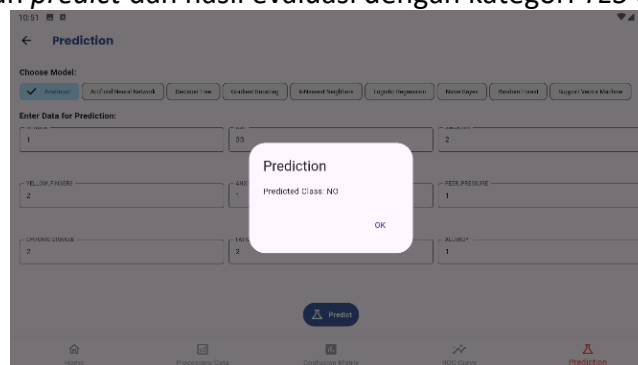
Pada halaman ROC dapat menyajikan ROC Curve dan AUC, menampilkan grafik *Receiver Operating Characteristic* (ROC) serta nilai *Area Under the Curve* (AUC) untuk mengevaluasi performa model dalam membedakan antara kelas. Grafik ini memberikan visualisasi *trade-off* antara *true positive rate* (TPR) dan *false positive rate* (FPR) pada berbagai *threshold*. Gambar 12 menunjukkan tampilan grafik ROC dan AUC.



Gambar 12. ROC Curve dan nilai AUC

Hasil evaluasi *Confusion Matrix*, *Classification Report* dan ROC Curve-AUC dapat di-download menjadi file pdf dengan memilih pada *save as PDF*.

Pada aplikasi NusaMiner dapat melakukan prediksi manual dengan masuk pada menu *Predict*, kolom yang disediakan akan menyesuaikan dengan *dataset* yang ada. Gambar 15 menampilkan halaman *predict* dan hasil evaluasi dengan kategori YES dan NO.



Gambar 13. Halaman prediksi manual

3.4. Pengujian Validasi Hasil

Berikut ini grafik Pengujian menggunakan RapidMiner dan NusaMiner



Gambar 14. Hasil pengujian RapidMiner dan NusaMiner

3.5. Pengujian Blackbox Testing NusaMiner

Pengujian aplikasi NusaMiner dilakukan dengan metode *blackbox testing* untuk memastikan fungsi berjalan sesuai spesifikasi. Fokus pengujian terletak pada interaksi pengguna tanpa menilai struktur kode. Skenario uji mencakup *input*, proses, dan *output* yang diharapkan. Hasil menunjukkan seluruh fitur utama, termasuk registrasi, *login*, unggah *dataset*, *preprocessing* (*set role*, *clean data*, konversi kategorikal, normalisasi), pemilihan algoritma, evaluasi model, prediksi gejala, serta pembaruan profil, berfungsi dengan baik. Hasil detail disajikan pada Tabel 2.

Tabel 2. Pengujian Blackbox testing

Skenario Pengujian	Test Case	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
Register	Input seluruh kolom	Ada notifikasi sukses	Akun berhasil didaftarkan	Berhasil
Login	Input: <i>username</i> dan <i>password valid</i>	Notifikasi berhasil masuk	Halaman utama ditampilkan	Berhasil
Update profile	Update: <i>fullname</i> , <i>address</i> , dan <i>phone</i>	Data berhasil diperbarui	Notifikasi sukses muncul	Berhasil
Upload dataset	File CSV <i>valid</i>	Menampilkan sampel <i>dataset</i>	Sampel <i>dataset</i> ditampilkan	Berhasil
Clean data	<i>Dataset</i> tanpa <i>missing value</i>	Proses berlanjut tanpa perubahan	Data tetap sam, tidak dibersihkan	Berhasil
Convert Categorical	<i>Dataset</i> berisi kolom kategorikal	Data diubah menjadi angka	Nilai kategori menjadi angka	Berhasil
Normalize	Kolom <i>AGE</i>	Nilai <i>AGE</i> disetaran menjadi skala 0-1	Nilai berubah sesuai skala <i>min-max</i>	Berhasil
Drop column	Tidak ada kolom yang dihapus	Semua kolom dipertahankan	Tidak ada perubahan jumlah kolom	Berhasil

Pilih algoritma	Memilih model dan rasio <i>split data</i>	<i>Confusion matrix</i> dan <i>classification report</i> ditampilkan	Hasil evaluasi ditampilkan	Berhasil
Prediksi manual	<i>Input</i> semua atribut pengujian	Label resiko ditampilkan	Prediksi berhasil ditampilkan	Berhasil

4. KESIMPULAN

Penelitian ini menunjukkan bahwa NusaMiner secara konsisten menghasilkan akurasi yang lebih tinggi dan stabil dibandingkan RapidMiner pada seluruh rasio pembagian data (70:30, 80:20, dan 90:10). Pada RapidMiner, algoritma SVM dan ANN memberikan performa terbaik dengan akurasi tertinggi sebesar 92,13%, sedangkan pada NusaMiner, *Gradient Boosting* dan *Random Forest* menjadi algoritma paling unggul dengan akurasi hingga 100% pada beberapa skenario pengujian. RapidMiner cenderung mengalami penurunan akurasi ketika proporsi data latih meningkat, sementara beberapa algoritma di NusaMiner, seperti *AdaBoost* dan *Naïve Bayes*, justru menunjukkan peningkatan performa. Secara keseluruhan, hasil penelitian ini menegaskan bahwa *Gradient Boosting* dan *Random Forest* merupakan algoritma yang paling andal untuk deteksi kanker paru-paru berbasis *machine learning*.

Penelitian selanjutnya disarankan untuk menambahkan fitur interpretabilitas model, menerapkan cross-validation, mengembangkan pendekatan clustering untuk analisis pola data tanpa label, serta menguji model pada dataset yang lebih besar dan beragam guna meningkatkan validitas dan generalisasi hasil.

DAFTAR PUSTAKA

- [1] World Health Organization, "Prees Release No.345: Global cancer burden growing, amidst mounting need for services," *Int. Agency Res. Cancer*, no. 345, pp. 185–187, 2024, [Online]. Available: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services%0Ahttps://www.iarc.who.int/wp-content/uploads/2024/02/pr345_E.pdf
- [2] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview," *Int. J. Cancer*, vol. 149, no. 4, pp. 778–789, 2021, doi: 10.1002/ijc.33588.
- [3] R. Kementerian Kesehatan, "Komite Penanggulangan Kanker Nasional," *Pandu. Penatlaksanaan Kanker Serviks*, pp. 10–11, 2016.
- [4] S. Rahmaeda and R. Prathivi, "Komparasi Metode SVM dan Logistic Regression untuk Klasifikasi Hipotesa Penyakit Kanker Paru Paru Berdasarkan Gejala Awal," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 6, no. 1, pp. 210–218, 2025, doi: 10.30645/kesatria.v6i1.562.
- [5] A. Naseh Khudori and M. Syauqi Haris, "Implementasi Decision tree Untuk Prediksi Kanker Paru-Paru," *J. Ris. Sist. Inf. Dan Tek. Inform. (JURASIK)*, vol. 9, no. 1, pp. 94–106, 2024, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [6] D. Widyawati and A. Faradibah, "Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine," *Indones. J. Data Sci.*, vol. 4, no. 2, pp. 80–89, 2023, doi: 10.56705/ijodas.v4i2.76.

- [7] H. M. Aliksieieva, V. H. Khomenko, and V. V. Khomenko, "the Role of Artificial Intelligence in Enhancing Online Learning," *Innov. Pedagog.*, vol. 2, no. 63, pp. 172–175, 2023, doi: 10.32782/2663-6085/2023/63.2.36.
- [8] S. Farshchiha, S. Asoudeh, M. S. Kuhshuri, M. Eisaeid, M. Azadie, and S. Hesarak, "A Comprehensive Analysis on Machine Learning based Methods for Lung Cancer Level Classification," 2025.
- [9] M. Jasmine Pemeena Priyadarsini *et al.*, "Lung Diseases Detection Using Various Deep Learning Algorithms," *J. Healthc. Eng.*, vol. 2023, 2023, doi: 10.1155/2023/3563696.
- [10] B. Andriska, C. Permana, and M. Djamaluddin, "Penerapan Python Dalam Data Mining Untuk Prediksi Kanker Paru Organisasi Kesehatan Dunia (WHO) menyatakan bahwa kanker merupakan kelompok penyakit yang berasal dari hampir seluruh organ tubuh dimana sel-sel yang terdapat pada organ tubuh tersebut tumb," vol. 6, no. 2, 2023.
- [11] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, 2023, doi: 10.35970/infotekmesin.v14i1.1751.
- [12] and B. I. National Heart, Lung, "Lungs," U.S. Department of Health and Human Services.
- [13] Y. E. Sembiring *et al.*, "Lung Cancer: A Literature Review," *J. Respirasi*, vol. 9, no. 3, pp. 246–251, 2023, doi: 10.20473/jr.v9-i.3.2023.246-251.
- [14] S. Firyal Nabila, D. Setiawan Hendyca Putra, S. Farlinda, E. Tri Ardianto, J. Kesehatan, and P. Negeri Jember, "J-REMI : Jurnal Rekam Medik Dan Informasi Kesehatan Analisis Faktor Risiko Pada Penyakit Karsinoma Paru (C34) Pasien Rawat Inap di Rumah Sakit Baladhika Husada Jember," vol. 2, no. 2, pp. 244–254, 2021.
- [15] aici-umg.com, "Pengertian AI: Definisi dan Konsep Utama," aici-umg.com. [Online]. Available: https://aici-umg.com/article/pengertian-ai-definisi-dan-konsep-utama/?utm_source=chatgpt.com
- [16] T. Wahyudi, "Studi Kasus Pengembangan dan Penggunaan Artificial Intelligence (AI) Sebagai Penunjang Kegiatan Masyarakat Indonesia," *Indones. J. Softw. Eng.*, vol. 9, no. 1, pp. 28–32, 2023, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/ijse>
- [17] E. S. Eriana, "Artificial Intelligence – AI," *Encycl. Digit. Agric. Technol.*, pp. 84–84, 2023, doi: 10.1007/978-3-031-24861-0_300007.
- [18] V. Goar and N. S. Yadav, "Foundations of machine learning," *Intell. Optim. Tech. Bus. Anal.*, no. M1, pp. 25–48, 2024, doi: 10.4018/979-8-3693-1598-9.ch002.
- [19] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Soc. Sci. Res.*, vol. 110, no. April 2022, p. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.
- [20] C. A. Palacios, J. A. Reyes-Suárez, L. A. Bearzotti, V. Leiva, and C. Marchant, "Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile," *Entropy*, vol. 23, no. 4, pp. 1–23, 2021, doi: 10.3390/e23040485.
- [21] J. Han, M. Kamber, and J. Pei, *Data Mining : Concepts and Techniques : Concepts and Techniques (3rd Edition)*. Morgan Kaufmann, 2012.
- [22] V. Çetin and O. Yıldız, "A comprehensive review on data preprocessing techniques in data analysis," *Pamukkale Univ. J. Eng. Sci.*, vol. 28, no. 2, pp. 299–312, 2022, doi:

10.5505/pajes.2021.62687.

- [23] Y. Liu, V. Stein Dani, I. Beerepoot, and X. Lu, "Turning Logs into Lumber: Preprocessing Tasks in Process Mining," *Lect. Notes Bus. Inf. Process.*, vol. 503 LNBIP, pp. 98–109, 2024, doi: 10.1007/978-3-031-56107-8_8.