

Aplikasi *Machine Learning* Berbasis *Mobile* untuk Deteksi Dini Kanker Paru paru

¹Deny Alfian, ²Fathimah Azzahro, ³Achmad Baroqah Pohan, ⁴Besus Maula Sulthon

^{1,2,3,4}Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta

E-mail: ¹deny.al91@gmail.com, ²fathimahazzahra53@gmail.com,
³achmad.abq@bsi.ac.id, ⁴besus.bem@bsi.ac.id

ABSTRAK

Kanker paru-paru merupakan salah satu jenis kanker paling mematikan di dunia dengan angka morbiditas dan mortalitas yang tinggi. Deteksi dini memiliki peran penting untuk meningkatkan peluang kesembuhan, namun masih sering terkendala keterlambatan diagnosis. Penelitian ini bertujuan untuk melakukan komparasi performa sembilan algoritma klasifikasi dalam mendeteksi kanker paru-paru menggunakan aplikasi NusaMiner, 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). *Dataset* yang digunakan berasal dari Kaggle, terdiri atas 309 record dengan 16 atribut yang mencakup faktor risiko dan gejala pasien. Tahapan penelitian meliputi *preprocessing data* (*data cleaning*, *encoding*, normalisasi, dan pembagian data), pelatihan model, evaluasi performa menggunakan *Confusion Matrix*, *Classification Report*, *ROC Curve*, dan AUC, serta pengujian aplikasi dengan metode *blackbox testing*. Hasil penelitian menunjukkan bahwa NusaMiner mampu memberikan akurasi yang tinggi dan stabil pada berbagai rasio pembagian data, dengan algoritma *Gradient Boosting* dan *Random Forest* menempati posisi terbaik dengan akurasi hingga 100% pada beberapa skenario. Dengan demikian, NusaMiner berpotensi menjadi solusi inovatif untuk mendukung deteksi dini kanker paru-paru yang lebih efektif, efisien, dan mudah diimplementasikan dalam praktik pelayanan kesehatan.

Kata kunci : Kanker Paru-paru, Klasifikasi, Prediksi, *Data Mining*, *Machine Learning*

ABSTRACT

Lung cancer is one of the deadliest types of cancer in the world with high morbidity and mortality rates. Early detection plays a crucial role in increasing the chances of recovery; however, it is often hindered by delays in diagnosis. This study aims to compare the performance of nine classification algorithms in detecting lung cancer using the NusaMiner application, namely *AdaBoost*, *Artificial Neural Network* (ANN), *Decision Tree*, *Gradient Boosting*, *K-Nearest Neighbors* (KNN), *Logistic Regression*, *Naïve Bayes*, *Random Forest*, and *Support Vector Machine* (SVM). The dataset used comes from Kaggle, consisting of 309 records with 16 attributes that include risk factors and patient symptoms. The stages of the research include data preprocessing (*data cleaning*, *encoding*, normalization, and data splitting), model training, performance evaluation using *Confusion Matrix*, *Classification Report*, *ROC Curve*, and AUC, and application testing using the black-box testing method. The research results show that NusaMiner is capable of providing high and stable accuracy across various data split ratios, with the *Gradient Boosting* and *Random Forest* algorithms ranking the best with accuracy up to 100% in several scenarios. Thus, NusaMiner has the potential to be an

innovative solution to support more effective, efficient, and easily implementable early detection of lung cancer in health service practices.

Keyword : 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 (World Health Organization, 2024). Di Indonesia, kanker ini menjadi penyebab kematian tertinggi akibat kanker, dengan 34.339 kematian dari 38.904 kasus baru (Ferlay et al., 2021). Penyebab utamanya adalah merokok, serta polusi udara, bahan kimia, alkohol, genetika, dan obesitas (Kementerian Kesehatan, 2016; Rahmaeda & Prathivi, 2025).

Deteksi dini kanker paru-paru sering terlambat, menyebabkan banyak kasus ditemukan pada stadium lanjut dengan peluang sembuh rendah (Naseh Khudori & Syauqi Haris, 2024). Tingginya angka morbiditas dan mortalitas menekankan pentingnya pencegahan, deteksi dini, dan penanganan efektif (Widyawati & Faradibah, 2023).

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, tuberkulosis, serta kanker paru-paru (Alieksieieva et al., 2023; Farshchiha et al., 2025; Jasmine Pemeena Priyadarsini et al., 2023).

Andriska et al., telah melakukan penelitian deteksi kanker paru-paru menggunakan Google Colab, dengan model C4.5 menghasilkan nilai akurasi 87% (Andriska et al., 2023). Sementara itu Widyawati & Faradibah (Widyawati & Faradibah, 2023) 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., (Sari et al., 2023) menggunakan RStudio dengan model *Naïve Bayes* dan *Random Forest* menghasilkan 98,4% untuk *Random Forest* dan 95,1% *Naïve Bayes*. Rahmaeda et al., juga (Rahmaeda & Prathivi, 2025) 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 NusaMiner dengan Sembilan algoritma klasifikasi: *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 ini meliputi: perbandingan akurasi model klasifikasi dalam prediksi kanker paru-paru, perancangan dan pengujian data medis secara efektif dalam aplikasi *mobile*, dan integrasi dan pengelolaan dataset medis berbasis Python agar aplikasi dapat digunakan secara praktis oleh tenaga medis.

2. LANDASAN TEORI

Paru-Paru

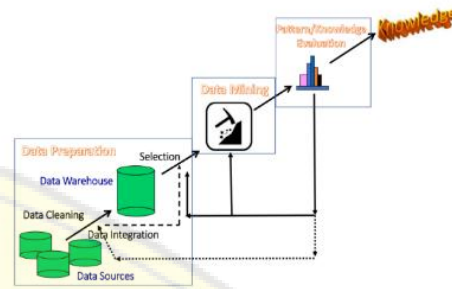
Paru-paru adalah organ utama pernapasan yang bertugas menukar oksigen dan karbon dioksida melalui alveoli, didukung oleh sistem peredaran darah, saraf, imun, dan otot (National Heart, Lung, 2024). Kanker paru-paru merupakan tumor ganas dari epitel saluran napas, terutama bronkus, yang berkembang akibat mutasi genetik dan menyebar melalui darah atau limfa (metastasis) serta membentuk pembuluh darah baru (angiogenesis) (Ferlay et al., 2021; Firyal Nabila et al., 2021; Sembiring et al., 2023).

Machine Learning

Kecerdasan buatan (AI) adalah rekayasa kecerdasan manusia pada sistem komputer untuk meniru proses berpikir, mengenali pola, belajar, dan beradaptasi (aici-umg.com, 2024; Wahyudi, 2023). Salah satu cabangnya, *machine learning* (ML), memungkinkan sistem belajar dari data tanpa pemrograman langsung dengan membangun model prediktif (Eriana, 2023; Goar & Yadav, 2024). Dalam medis, AI digunakan untuk menganalisis data seperti rekam medis dan citra radiologi, membantu diagnosis lebih cepat dan akurat (Alieksieieva et al., 2023).

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 (Palacios et al., 2021; Shu & Ye, 2023). KDD mencakup tahapan seperti: *Data cleaning*, *Data transformation*, dan *Data mining*. Tahapan proses KDD secara visual ditampilkan pada gambar 1.



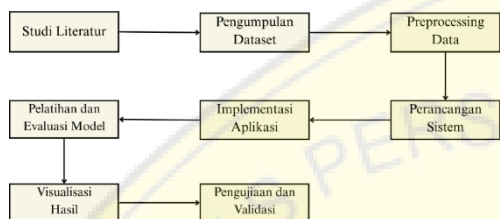
Gambar 1. Tahapan KDD

Algoritma Klasifikasi

- (a) *AdaBoost*: metode *boosting* yang meningkatkan akurasi dengan memberi bobot lebih besar pada data yang sulit diprediksi
- (b) *Artificial Neural Network*: jaringan saraf buatan dengan lapisan *input*, *hidden*, dan *output* yang mampu mengenali pola kompleks
- (c) *Decision Tree*: struktur pohon yang memetakan atribut data menjadi simpul keputusan hingga menghasilkan kelas akhir
- (d) *Gradient Boosting*: metode *boosting* bertahap yang memperbaiki kesalahan prediksi sebelumnya menggunakan gradien fungsi kerugian
- (e) *K-Nearest Neighbors*: mengklasifikasikan data baru berdasarkan kedekatan dengan sejumlah tetangga terdekat
- (f) *Logistic Regression*: memodelkan probabilitas suatu kelas menggunakan fungsi logistik (*sigmoid*)
- (g) *Naïve Bayes*: algoritma berbasis probabilitas Bayes dengan asumsi independensi antar fitur
- (h) *Random Forest*: metode *ensemble* yang menggabungkan banyak pohon keputusan dari subset data acak untuk meningkatkan akurasi
- (i) *Support Vector Machine*: mencari *hyperplane* optimal untuk memisahkan kelas, baik secara linier maupun non-linier.

3. METODOLOGI

Tahapan penelitian ini disusun secara sistematis mulai dari studi literatur hingga visualisasi hasil. Alur lengkap tahapan tersebut ditampilkan pada gambar 2 berikut:



Gambar 2. Tahapan Penelitian

Penelitian ini menggunakan pendekatan *supervised learning*, yaitu pelatihan model dengan data berlabel untuk memetakan *input* ke *output* secara akurat (Eriana, 2023; Goar & Yadav, 2024). Model divalidasi menggunakan data uji, penting dalam diagnosis medis.

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 1 terdapat penjelasan lebih lanjut mengenai atribut dari *dataset* yang digunakan.

Tabel 1. Detail *Dataset*

No	Fitur	Keterangan
1	Gender	M/F atau 0/1
2	Age	Usia responden
3	Smoking	Status Merokok (Yes/No)
4	Yellow Finger	Kuku menguning (Yes/No)
5	Anxiety	Kecemasan (Yes/No)
6	Peer Pressure	Tekanan dari teman (Yes/No)
7	Chronic Disease	Riwayat penyakit kronis (Yes/No)
8	Fatigue	Kelelahan (Yes/No)

9	Allergy	Alergi (Yes/No)
10	Wheezing	Mengi (Yes/No)
11	Alcohol Consuming	Konsumsi alkohol (Yes/No)
12	Coughing	Batuk (Yes/No)
13	Shortness of Breath	Sesak napas (Yes/No)
14	Swallowing Difficulty	Kesulitan menelan (Yes/No)
15	Chest Pain	Nyeri dada (Yes/No)
16	Lung Cancer	Label target: diagnosis kanker paru (Yes/No)

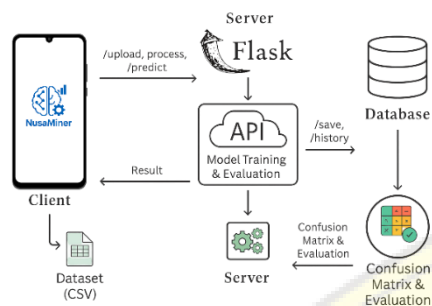
Preprocessing Data

Preprocessing data merupakan tahap penting dalam *Knowledge Discovery in Databases (KDD)* untuk memastikan kualitas data (Han et al., 2012). Langkah-langkah yang dilakukan mencakup:

- (a) *Data cleaning*: menghapus nilai kosong, duplikasi, dan *outlier* agar data bersih dan layak digunakan (Çetin & Yıldız, 2022)
- (b) *Data transformation*: melakukan label *encoding* pada atribut kategorikal serta normalisasi fitur numerik menggunakan metode *min-max scaling* agar berada dalam rentang 0–1 (Liu et al., 2024)
- (c) *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.

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

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:

- Akurasi: proporsi prediksi benar terhadap total prediksi yang dilakukan, perhitungan akurasi dapat dilihat pada persamaan 1
- Precision*: ketepatan prediksi positif dari total prediksi positif yang dilakukan persamaan *precision* pada persamaan 2
- Recall*: sensitivitas model terhadap data positif dengan benar, *recall* dapat dicari dengan persamaan persamaan 3
- F1-Score*: keseimbangan *precision* dan *recall*, cara menghitungnya pada persamaan 4
- Support*: merupakan jumlah kejadian sebenarnya dari setiap kelas dalam *dataset*
- Confusion Matrix*: tabel untuk membandingkan label prediksi dan label aktual, visualisasi *confusion matrix* dapat dilihat pada gambar 5
- ROC Curve*: mengukur kemampuan model membedakan kelas secara keseluruhan
- AUC*: memberikan gambaran kinerja model.

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

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

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

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

		Predicted	
Actual		TN	FP
		FN	TP

Gambar 4. *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.

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 (Goar & Yadav, 2024).

4. HASIL DAN PEMBAHASAN

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 5 menampilkan *dataset* yang digunakan.

GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEEL	CHRONIC_DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL	COUGHING	DIFFICULTY_OF_BREATH	SWALLOWING	CHEST_PAIN	LUNG_CANCER
M	65	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	74	2	1	1	1	1	1	1	1	1	1	1	1	1	1
F	60	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	63	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F	50	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F	50	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	50	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	52	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	52	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F	51	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	46	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	53	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	51	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	72	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	40	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	58	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	69	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	48	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	75	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	57	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	58	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Gambar 5. Dataset Diagnosis Pasien Kanker Paru-Paru

Hasil dan pembahasan berisi hasil analisis fenomena di wilayah penelitian yang relevan dengan tema kajian. Hasil penelitian hendaknya dibandingkan dengan teori dan temuan penelitian yang relevan)

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

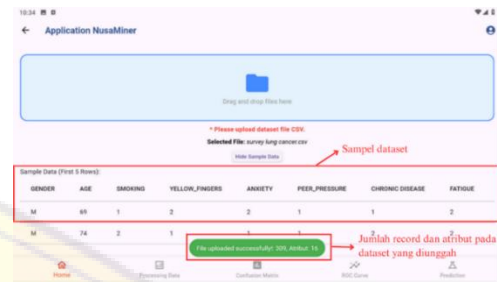
Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data bersih, terstruktur, dan siap dianalisis. Proses meliputi:

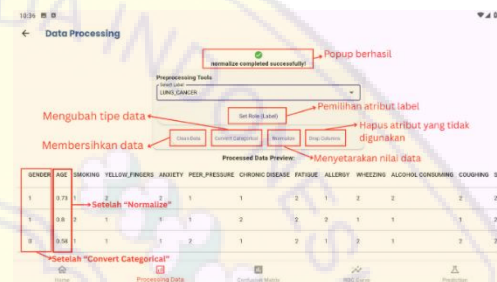
- Clean data* untuk menangani *missing values*, namun tidak diperlukan karena dataset bebas nilai kosong
- Convert categorical* dengan label encoding ("Male"=1, "Female"=0)
- Normalize* atribut AGE menggunakan Min-Max Normalization
- splitting data* menjadi *training set* dan *testing set* dengan rasio 70:30, 80:20, atau 90:10.

Implementasi

Preprocessing data menggunakan NusaMiner ditunjukkan pada gambar dibawah ini. Sebelum melakukan *preprocessing data* kita perlu *upload dataset* terlebih dahulu pada halaman *home*, seperti gambar 6. Setelah *dataset* di-*upload* maka masuk pada halaman *processing data* untuk melakukan *preprocessing*, seperti yang dijelaskan pada gambar 6.

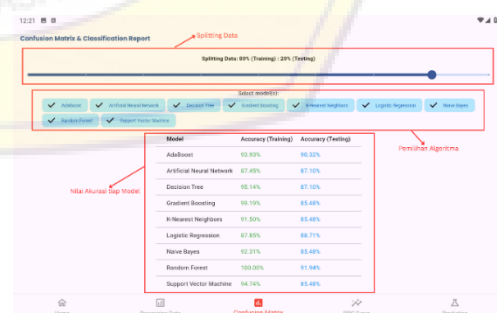


Gambar 6. Halaman utama untuk upload dataset

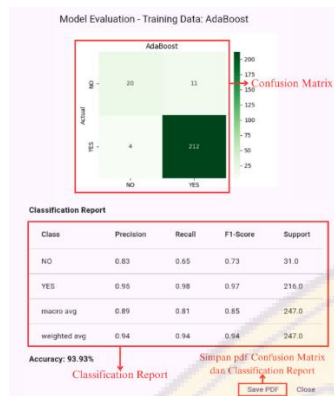


Gambar 7. Preprocessing pada NusaMiner

Pemilihan model algoritma dilakukan pada halaman *Confusion Matrix*, sebagaimana ditunjukkan pada gambar 8, 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 9.

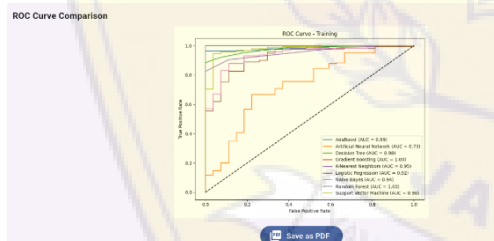


Gambar 8. Pemilihan algoritma, Splitting data, dan nilai akurasi



Gambar 9. Hasil evaluasi model NusaMiner

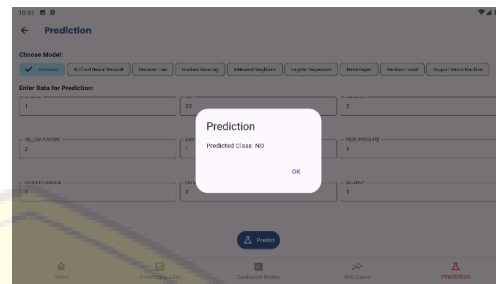
Pada halaman ROC dapat menyajikan ROC Curve dan AUC, menampilkan grafik *Receiver Operating Characteristic* (ROC) serta nilai *Area Under the Curve* (AUC) untuk mengevaluasi performa model dalam membedakan antara kelas. Grafik ini memberikan visualisasi *trade-off* antara *true positive rate* (TPR) dan *false positive rate* (FPR) pada berbagai *threshold*. Gambar 10 menunjukkan tampilan grafik ROC dan AUC.



Gambar 10. 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 11 menampilkan halaman *predict* dan hasil evaluasi dengan kategori *YES* dan *NO*.

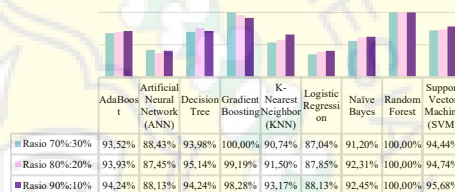


Gambar 11. Halaman Prediksi Manual

Pengujian Validasi Hasil

Berikut ini grafik Pengujian menggunakan NusaMiner.

Grafik Nilai Accuracy NusaMiner



Gambar 12. Hasil Pengujian NusaMiner

Pengujian Blackbox Testing

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 pada tabel 2.

Tabel 2. Pengujian Blackbox Testing

No	Skenario Pengujian – Test Case	Hasil Pengujian - Kesimpulan
1	Register – Input seluruh kolom	Akun berhasil didaftarkan – Berhasil
2	Login – Input: username dan password valid	Halaman utama ditampilkan – Berhasil
3	Update Profile – Update: fullname, address, dan phone	Notifikasi sukses muncul – Berhasil
4	Upload dataset : File CSV valid	Sampel dataset ditampilkan - Berhasil

5	<i>Clean Data – Dataset tanpa missing value</i>	Data tetap sama - Berhasil
6	<i>Convert Categorical – Dataset tanpa missing value</i>	Nilai Kategori menjadi angka – Berhasil
7	<i>Normalize – Kolom AGE</i>	Nilai berubah sesuai skala <i>min-max</i> – Berhasil
8	<i>Drop Column – Tidak ada kolom yang dihapus</i>	Tidak ada perubahan jumlah kolom – Berhasil
9	Pilih Algoritma – Memilih model dan rasio <i>split data</i>	Hasil Evaluasi ditampilkan – Berhasil
10	Prediksi manual – Label resiko ditampilkan	Prediksi berhasil ditampilkan - Berhasil

5. KESIMPULAN

Studi klasifikasi kanker paru-paru dengan sembilan algoritma *machine learning* menggunakan NusaMiner:

- NusaMiner menunjukkan akurasi tinggi dan stabil pada semua rasio pembagian data (70:30, 80:20, 90:10)
- Gradient Boosting* dan *Random Forest* menjadi yang terbaik dengan akurasi mencapai 100% pada beberapa rasio.

Secara umum, *Gradient Boosting* dan *Random Forest* merupakan pilihan unggulan untuk deteksi kanker paru-paru berbasis *machine learning*.

Rekomendasi Penelitian Selanjutnya:

- Tambahkan fitur interpretabilitas model (SHAP, LIME).
- Kembangkan fitur *clustering* dan regresi.
- Terapkan *cross-validation* untuk meningkatkan validitas.
- Uji coba pada *dataset* yang lebih besar dan beragam

6. UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang tulus kepada seluruh pihak yang telah berkontribusi dalam penyelesaian penelitian dan penulisan artikel ini. Terima kasih khusus ditujukan kepada para pembimbing dan rekan sejawat atas arahan, saran, serta dukungan yang berharga. Penulis juga mengakui peran penting keluarga yang senantiasa memberikan dorongan moril maupun

motivasi, sehingga penelitian ini dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

aici-umg.com. (2024). *Pengertian AI: Definisi dan Konsep Utama*. Aici-Umg.Com. https://aici-umg.com/article/pengertian-ai-definisi-dan-konsep-utama/?utm_source=chatgpt.com

Alieksieieva, H. M., Khomenko, V. H., & Khomenko, V. V. (2023). the Role of Artificial Intelligence in Enhancing Online Learning. *Innovate Pedagogy*, 2(63), 172–175. <https://doi.org/10.32782/2663-6085/2023/63.2.36>

Andriska, B., Permana, C., & Djamaluddin, M. (2023). *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. 6(2).

Çetin, V., & Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale University Journal of Engineering Sciences*, 28(2), 299–312. <https://doi.org/10.5505/pajes.2021.62687>

Eriana, E. S. (2023). Artificial Intelligence – AI. *Encyclopedia of Digital Agricultural Technologies*, 84–84. https://doi.org/10.1007/978-3-031-24861-0_300007

Farshchiha, S., Asoudeh, S., Kuhshuri, M. S., Eisaeid, M., Azadie, M., & Hesarak, S. (2025). *A Comprehensive Analysis on*

Machine Learning based Methods for Lung Cancer Level Classification.

- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, 149(4), 778–789. <https://doi.org/10.1002/ijc.33588>
- Firyal Nabila, S., Setiawan Hendyca Putra, D., Farlinda, S., Tri Ardianto, E., Kesehatan, J., & Negeri Jember, P. (2021). *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*. 2(2), 244–254.
- Goar, V., & Yadav, N. S. (2024). Foundations of machine learning. *Intelligent Optimization Techniques for Business Analytics, MI*, 25–48. <https://doi.org/10.4018/979-8-3693-1598-9.ch002>
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques: Concepts and Techniques (3rd Edition). In *Data Mining*. Morgan Kaufmann.
- Jasmine Pemeena Priyadarsini, M., Kotecha, K., Rajini, G. K., Hariharan, K., Utkarsh Raj, K., Bhargav Ram, K., Indragandhi, V., Subramaniaswamy, V., & Pandya, S. (2023). Lung Diseases Detection Using Various Deep Learning Algorithms. *Journal of Healthcare Engineering*, 2023. <https://doi.org/10.1155/2023/35636> 96
- Kementerian Kesehatan, R. (2016). Komite Penanggulangan Kanker Nasional. *Panduan Penatlaksanaan Kanker Serviks*, 10–11.
- Liu, Y., Stein Dani, V., Beerepoot, I., & Lu, X. (2024). Turning Logs into Lumber: Preprocessing Tasks in Process Mining. *Lecture Notes in Business Information Processing*, 503 LNBIP, 98–109. https://doi.org/10.1007/978-3-031-56107-8_8
- Naseh Khudori, A., & Syauqi Haris, M. (2024). Implementasi Decision tree Untuk Prediksi Kanker Paru-Paru. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, 9(1), 94–106. <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- National Heart, Lung, and B. I. (2024). *Lungs*. U.S. Department of Health and Human Services.
- Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile. *Entropy*, 23(4), 1–23. <https://doi.org/10.3390/e23040485>
- Rahmaeda, S., & Prathivi, R. (2025). Komparasi Metode SVM dan Logistic Regression untuk Klasifikasi Hipotesa Penyakit Kanker Paru Paru Berdasarkan Gejala Awal. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, 6(1), 210–218. <https://doi.org/10.30645/kesatria.v6i1.562>
- Sari, L., Romadloni, A., & Listyaningrum, R. (2023). Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Infotekmesin*, 14(1), 155–

162.
<https://doi.org/10.35970/infotekmesin.v14i1.1751>
- Sembiring, Y. E., Effendi, W. I., Dillon, J. J., Soebroto, H., Winarno, D. J. S., Puruhito, P., Putra, I. G. A. M. A., Sebayang, A. N. O., & Negoro, S. P. W. B. (2023). Lung Cancer: A Literature Review. *Jurnal Respirasi*, 9(3), 246–251.
<https://doi.org/10.20473/jr.v9-i.3.2023.246-251>
- Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110(April 2022), 102817.
<https://doi.org/10.1016/j.ssresearch.2022.102817>
- Wahyudi, T. (2023). Studi Kasus Pengembangan dan Penggunaan Artificial Intelligence (AI) Sebagai Penunjang Kegiatan Masyarakat Indonesia. *Indonesian Journal on Software Engineering (IJSE)*, 9(1), 28–32.
<https://ejournal.bsi.ac.id/ejurnal/index.php/ijse>
- Widyawati, D., & Faradibah, A. (2023). Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine. *Indonesian Journal of Data and Science*, 4(2), 80–89.
<https://doi.org/10.56705/ijodas.v4i2.76>
- World Health Organization. (2024). Press Release No.345: Global cancer burden growing, amidst mounting need for services. *International Agency for Research on Cancer*, 345, 185–187.
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