

**STUDI KOMPARATIF ALGORITMA *MACHINE LEARNING* DALAM
KLASIFIKASI DAN PREDIKSI KANKER PARU-PARU**

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

ATMA AGILIA TRIWARDANI

NPM: 2113030005

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI

2025

Skripsi oleh :

ATMA AGILIA TRIWARDANI

NPM: 2113030005

Judul

**STUDI KOMPARATIF ALGORITMA *MACHINE LEARNING* DALAM
KLASIFIKASI DAN PREDIKSI KANKER PARU-PARU**

Telah disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri

Tanggal: 4 Juli 2025

Pembimbing I



Erna Daniati, M.Kom
NIDN. 0723058501

Pembimbing II



Anita Sari Wardani, M.Kom
NIDN. 0713018402

Skripsi oleh :

ATMA AGILIA TRIWARDANI

NPM: 2113030005

Judul

**STUDI KOMPARATIF ALGORITMA *MACHINE LEARNING* DALAM
KLASIFIKASI DAN PREDIKSI KANKER PARU-PARU**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 8 Juli 2025

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua Penguji : Erna Daniati, M.Kom

[.....]

2. Penguji 1 : Arie Nugroho, M.Kom

[.....]

3. Penguji 2 : Anita Sari Wardani, M.Kom

[.....]



Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer

Dr. Sulistiono, M.Si.

NIDN. 0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Atma Agilia Triwardani
Jenis Kelamin : Perempuan
Tempat/tgl. lahir : Kediri / 17 Februari 2003
NPM : 2113030005
Fak/Prodi. : FTIK / S1-Sistem Informasi

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 8 Juli 2025

Yang Menyatakan



ATMA AGILIA TRIWARDANI

NPM: 2113030005

MOTTO

”Kapan-kapan semoga kau berani. Hidup bukan saling mendahului. Bermimpilah sendiri-sendiri. Tak ada yang tahu kapan kau mencapai tuju. Dan percayalah bukan urusanmu untuk menjawab itu. Katakan pada dirimu besok mungkin kita sampai, besok mungkin tercapai.”

(Hindia – *Besok mungkin kita sampai*)

“Lakukan apa yang kau mau sekarang. Saat hatimu bergerak, jangan kau larang. Lagi pula hidup akan berakhir. Maka lakukan apa yang kau mau sekarang.”

(Hindia – *Berdansalah, Karir Ini Tak Ada Artinya*)

“Pada akhirnya ini semua, hanyalah permulaan”

(Nadin Amizah – *Beranjak Dewasa*)

ABSTRAK

Atma Agilia Triwardani : Studi Komparatif Algoritma *Machine Learning* Dalam Klasifikasi dan Prediksi Kanker Paru-Paru, Skripsi, SI, FTIK UN PGRI Kediri, 2025.

Kata Kunci : kanker paru-paru, *machine learning*, klasifikasi, data klinis.

Kanker paru-paru merupakan salah satu penyebab kematian tertinggi akibat penyakit tidak menular di dunia, termasuk di Indonesia. Tingginya angka mortalitas disebabkan oleh keterlambatan deteksi, sehingga deteksi dini menjadi krusial dalam meningkatkan efektivitas penanganan dan peluang kesembuhan pasien. Dalam era digital saat ini, pemanfaatan teknologi berbasis kecerdasan buatan seperti *machine learning* menjadi solusi potensial dalam mendeteksi kanker sejak dini melalui analisis data klinis dan identifikasi pola gejala penyakit secara otomatis.

Penelitian ini bertujuan untuk melakukan studi komparatif terhadap lima algoritma *machine learning* dalam klasifikasi dan prediksi kanker paru-paru, yaitu *Decision Tree (C4.5)*, *K-Nearest Neighbor (KNN)*, *Naive Bayes*, *Random Forest*, dan *Extreme Gradient Boosting (XGBoost)*. Pendekatan yang digunakan adalah metode kuantitatif dengan eksperimen komputasi berbasis *Python* pada platform *Google Colaboratory*. Dataset yang digunakan bersumber dari *Kaggle*, terdiri atas 1.157 data pasien dengan 16 atribut. Tahapan penelitian meliputi data selection, data preprocessing (*missing value*, normalisasi, eliminasi duplikat), *feature selection* menggunakan korelasi *Pearson*, transformasi data kategorikal, pelatihan model, dan evaluasi menggunakan akurasi, presisi, *recall*, *F1-score*, serta confusion matrix.

Hasil penelitian menunjukkan bahwa seluruh algoritma mampu mengklasifikasikan data dengan performa baik. Namun, algoritma *Naive Bayes* memperoleh hasil terbaik dengan akurasi 89,61%, presisi 94,19%, *recall* 85,26%, dan *F1-score* 89,47%. Ini membuktikan bahwa meskipun tergolong algoritma sederhana, *Naive Bayes* efektif untuk klasifikasi penyakit dengan kompleksitas data klinis tinggi.

Penelitian ini menyimpulkan bahwa pendekatan *machine learning* dapat menjadi alat bantu diagnostik yang andal dalam mendeteksi kanker paru-paru. Namun, keterbatasan studi ini terletak pada penggunaan satu jenis dataset dan belum menguji *hyperparameter tuning* yang lebih dalam. Penelitian selanjutnya dapat mengembangkan integrasi model *ensemble*, validasi silang yang lebih komprehensif, serta uji coba pada dataset real-time dari rumah sakit untuk meningkatkan generalisasi dan akurasi prediksi.

KATA PENGANTAR

Puji Syukur Kami panjatkan kehadiran Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul "Studi Komparatif Algoritma *Machine Learning* Dalam Klasifikasi dan Prediksi Kanker Paru-Paru" ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Pendidikan, pada Jurusan Sistem Informasi Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri. Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus-tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri yang selalu memberikan dorongan motivasi kepada mahasiswa.
2. Dr. Sulistiono, M.Si., selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri, atas arahan dan dukungan selama proses studi.
3. Sucipto, M.Kom., selaku Ketua Program Studi Sistem Informasi yang selalu memberikan pengarahan dan motivasi dalam penyelesaian tugas akademik, termasuk dalam penulisan skripsi ini.
4. Erna Daniati, M.Kom dan Anita Sari Wardani, M.Kom., selaku dosen pembimbing yang telah membimbing, mengarahkan, serta memberikan masukan yang sangat berarti selama proses penyusunan skripsi ini.
5. Kepada bapakku Alm. Bapak Supadmo terimakasih atas segala pengorbanan, kasih sayang, serta segala bentuk tanggung jawab atas kehidupan layak yang telah diberikan semasa bapak hidup. Terimakasih telah menjadi alasan penulis untuk tetap semangat berjuang meraih gelar sarjana, walaupun berat sekali harus melewati kerasnya kehidupan tanpa di dampingi sosok bapak, rasa iri dan rindu sering kali membuat terjatuh. Semoga bapak bangga dengan usaha dan perjuanganku selama ini dan bahagia di surganya Allah, aamiin. Untuk pintu surgaku Ibu Sunarti terima kasih karena telah menjadi tempat berpulang yang paling sabar dan pendoa yang tak pernah lelah menyebut namaku dalam setiap harapan terbaik. Terimakasih juga atas segala materi yang telah diusahakan untuk aku berkuliah. Segala lelah, ragu, dan jatuh bangunku dalam

perjalanan ini selalu kalian sambut dengan keyakinan. Jika aku sampai di titik ini, itu karena pengorbanan dan cinta kalian yang tak bersyarat. Semoga kelak aku bisa membalas meski takkan pernah sebanding dengan yang telah kalian beri.

6. Untuk kakak-kakakku tersayang, Ayu Suryaning Wulan dan Fany Amawati terima kasih telah menjadi contoh yang baik penulis. Penulis menyadari bahwa sebagian besar hidupku adalah karena pengaruh kalian. Terima kasih karena banyak membantu aku disegala situasi dan kondisi walau tak lepas dari roastingan jahanam tentunya. Terima kasih sudah nungguin aku ngeprint H-1 sidang sampe malem karena aku panik filenya gabisa diprint (memorable). Aku sangat bersyukur punya kakak seperti kalian, jika ada kehidupan selanjutnya jadi kakak-kakakku lagi ya (janji pas remaja aku gak bandel).
7. Pasanganku Riko Aditia Riansyah terima kasih atas doa, dukungan, makanann dan liburan menyenangkan dari kamu yang tak pernah putus. Terlepas ke mana hidup membawa kita nanti, hadirmu sangat berarti. Terimakasih telah menjadi bagian dari perjalanan hidupku dan selalu mengusahakan yang terbaik, ssemoga doa-doa baik membersamai kita!
8. Kembaranku since 2018 Jihan Febriyanti, S.Pd yang telah membersamai aku sedari bangku menengah atas hingga saat ini dan selamanya. Terimakasih atas telinga dan mata yang selalu ready untuk mendengar dan membaca semua curhatanku dari yang gapenting sampe penting banget. Untuk ngga bosen ngajak aku ngopi walau aku selalu sambat gaada duit. Terima kasih sudah selalu men-*support* dan membersamai perjalanan ini, dengan segala tawa, air mata, dan kebodohan-kebodohan kecil yang selalu kita rayakan. Mari mewujudkan wishlist-wishlist yang menumpuk itu dan membicarakan (gibah) banyak hal untuk waktu yang sangat lama.
9. Sahabat karibku Okfi Alfi Nurlaili dan Niken Avilaria terimakasih karena selalu ngajak ngopi demi kewarasan penulis dalam berkuliah dan selalu jadi tempat yang nyaman untuk berkeluh kesah.
10. Sobat produksian Audi Nurtania, A.Md.Kom dan Yessica Pungki Alya Firdaus, S.Sn terimakasih untuk saran-saran dan masukan selama dibangku perkuliahan.

Walau kita ngga satu kampus terimakasih karena tetap menjaga pertemanan kita sampai saat ini.

11. Sahabat perkuliahanku Elisa Tri Ammah A'fena, S.Kom yang tulus bersedia berteman dengan aku disaat aku berpikir tidak akan punya teman di bangku perkuliahan dan *emergency call* yang tidak pernah merasa terbebani walau diteror terus menerus. Terimakasih karena selalu mengingatkan penulis di segala urusan-urusan perkuliahan yang ribet itu. Terimakasih juga telah membersamai penulis selama berkuliah dan memberikan kritik saran untuk menyusun skripsi hingga tuntas. *Proud of u bestie see u on top!*
12. Teman-teman seperjuangan Sistem Informasi angkatan 21 terimakasih atas dukungan, bantuan, pengalaman, canda, tawa dan doa-doa baiknya. *See u on top guys, i'll always remember all of u!*
13. Saudariku Nur Habibah walaupun kamu nan jauh disana tapi terimakasih untuk segala bentuk dukungan, waktu yang selalu kamu luangkan untuk membaca dan memberi saran serta masukan atas curhatanku, semoga kita segera bertemu.
14. Ucapan terima kasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan Skripsi ini.
15. *Last but not least* untuk diriku sendiri, terima kasih telah bertahan sejauh ini, melewati berbagai ujian yang tak selalu mudah, namun tetap memilih untuk tidak menyerah. Dalam diam, kamu pernah menangis, merasa lelah, kecewa, bahkan ingin berhenti. Tapi di setiap keraguan itu, kamu selalu menemukan cara untuk bangkit meski perlahan, meski tertatih. Terima kasih telah tetap percaya, meskipun sering kali diliputi ragu. Terima kasih telah berani melangkah, meski tak selalu tahu apa yang menunggu di depan. Terima kasih telah terus belajar, tumbuh, dan membentuk versi terbaik dari dirimu sendiri hari demi hari. Perjalanan ini mungkin belum sempurna, tapi kamu patut berbangga karena telah sampai sejauh ini. Hari ini, pencapaian ini, adalah bukti bahwa kamu layak diperjuangkan oleh dirimu sendiri. Mari selalu mencintai dan merayakan diri sendiri lebih banyak lagi!

Disadari bahwa skripsi ini masih banyak kekurangan, maka diharapkan tegur sapa, kritik dan saran-saran, dari berbagai pihak sangat diharapkan. Akhirnya, disertai harapan semoga skripsi ini ada manfaatnya bagi kita semua, khususnya bagi dunia pendidikan, meskipun hanya ibarat setitik air bagi samudra yang luas.

Kediri, 4 Juli 2025

ATMA AGILIA TRIWARDANI

NPM: 2113030005

DAFTAR ISI

Halaman Sampul.....	i
Halaman Persetujuan	ii
Halaman Pengesahan	iii
Halaman Pernyataan	iv
Halaman Motto	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Identifikasi Masalah	3
C. Batasan Masalah.....	3
D. Rumusan Masalah	3
E. Tujuan Penelitian	4
F. Manfaat Penelitian.....	4
BAB II KAJIAN PUSTAKA	5
A. Kajian Teori	5
1. Kanker Paru-Paru	5
2. Algoritma Machine Learning.....	5
3. Data Mining	6
4. Knowledge Discovery in Database	6
5. Decision Tree : C4.5.....	7
6. K-Nearest Neighbors	8
7. Naive Bayes	9
8. Random Forest	9
9. Extreme Gradient Boosting.....	10
10. Confusion Matrix.....	11

11. Google Collaboratory	12
B. Kajian Hasil Penelitian Terdahulu	12
C. Kerangka Berpikir	24
D. Hipotesis	26
BAB III METODE PENELITIAN	27
A. Alur Penelitian	27
1. Kajian Pustaka.....	27
2. Identifikasi Masalah	27
3. Data Selection	28
4. Data Prepocecing.....	28
5. Transformation.....	29
6. Data mining.....	29
7. Evaluation	29
B. Tempat dan Waktu Penelitian.....	30
1. Tempat Penelitian.....	30
2. Waktu Penelitian	30
BAB IV HASIL DAN PEMBAHASAN.....	31
A. Data Selection	31
B. Data Prepocecing	32
C. Transformation.....	37
D. Data Mining	38
E. Evaluation.....	41
BAB V PENUTUP.....	50
A. Kesimpulan	50
B. Saran.....	51
DAFTAR PUSTAKA.....	52
LAMPIRAN	61

DAFTAR GAMBAR

Gambar 2.1 Alur KDD	7
Gambar 2.2 Kerangka Berpikir	24
Gambar 3.1 Alur Penelitian.....	27
Gambar 4.1 Kode Cek Missing Value	33
Gambar 4.2 Kode Cek Duplikat data.....	34
Gambar 4.3 Kode Labeling data.....	34
Gambar 4.4 Grafik Korelasi Pearson	35
Gambar 4.5 Kode Split Data	36
Gambar 4.6 Kode Transformation.....	37
Gambar 4.7 Kode Algoritma Decision Tree C4.5	38
Gambar 4.8 Kode Algoritma K-Nearest Neighbors (KNN)	39
Gambar 4.9 Kode Algoritma Naive Bayes.....	39
Gambar 4.10 Kode Algoritma Random Forest.....	40
Gambar 4.11 Kode Algoritma XGBoost.....	40
Gambar 4.12 Kode Evaluasi Model.....	41
Gambar 4.13 Confussion Matrix Decission Tree	43
Gambar 4.14 Confusion Matrix K-Nearest Neighbors.....	44
Gambar 4.15 Confusion Matrix Naive Bayes	45
Gambar 4.16 Confusion Matrix Random Forest	46
Gambar 4.17 Confusion Matrix XGBoost	47

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	11
Tabel 2.2 Kajian Penelitian Terdahulu	12
Tabel 3.1 Waktu Penelitian	30
Tabel 4.1 Dataset Lung Cancer	31
Tabel 4.2 Classification Report	42

DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan Skripsi dari SIAKAD2	63
Lampiran 2 Surat Keterangan Bebas Similarity dari PPI	64
Lampiran 3 Bukti Screenshot Loa Artikel/Screenshot Artikel Terbit	66

BAB I

PENDAHULUAN

A. Latar Belakang

Indonesia saat ini sedang menghadapi tantangan besar di sektor kesehatan akibat melonjaknya prevalensi penyakit tidak menular (PTM). PTM, yang juga dikenal sebagai penyakit degeneratif, menjadi isu utama kesehatan masyarakat pada abad ke-21 karena tingginya angka morbiditas dan mortalitas di tingkat global. Berdasarkan data dari World Health Organization, PTM mencakup penyakit seperti jantung, stroke, tekanan darah tinggi, diabetes, kanker, serta penyakit ginjal kronis. Diduga menyebabkan sekitar 36 juta kematian setiap tahun. dikarenakan kehidupan sehari-hari dan gaya hidup tidak sehat yang meningkatkan risiko munculnya penyakit tidak menular tersebut (Susanti et al., 2024).

Perkiraan secara menyeluruh menurut *International Agency for Research on Cancer* (IARC) menunjukkan bahwa pada tahun 2022 hampir 2,5 juta orang didiagnosis menderita kanker paru-paru dan lebih dari 1,8 juta orang meninggal karena penyakit tersebut. Kanker paru-paru termasuk jenis penyakit tidak menular yang sering dialami oleh anak-anak dan orang dewasa. Salah satu penyebab kanker paru-paru adalah paparan terlalu sering terhadap polusi udara yang buruk, paparan zat kimia bersifat karsinogen, paparan radiasi, beberapa infeksi virus dan individu yang memiliki riwayat kanker atau anggota keluarga mereka adalah risiko lain untuk kanker paru-paru. Gejala umum kanker paru-paru termasuk hemoptisi, suara serak, batuk, nyeri dada, abses paru-paru, dan sesak nafas (Septiani et al., 2022).

Mayoritas penderita kanker paru-paru adalah pria, sedangkan kanker paru pada wanita merupakan kasus keempat. *Bronchogenic carcinoma*, atau tumor ganas di bagian *epitel bronkus*, adalah awal bersemayamnya kanker paru paru dalam tubuh. Sampai saat ini, belum ada teknik penyaringan yang berhasil untuk mendeteksi kanker paru dengan cepat. Deteksi dini kanker paru-paru masih menjadi tantangan besar, sehingga penyakit ini sering kali baru diketahui ketika sudah memasuki stadium lanjut dan sulit diobati. Kebiasaan merokok serta pola hidup yang tidak sehat merupakan faktor pemicu utama, menjadikan kanker paru-paru

sebagai salah satu jenis kanker dengan angka kasus tertinggi ketiga di Indonesia (Septhya et al., 2023a).

Perkembangan teknologi membawa dampak yang signifikan dalam berbagai bidang termasuk bidang kesehatan. Konsep kecerdasan buatan yang meniru kecerdasan manusia untuk memecahkan berbagai masalah adalah *machine learning*. Cara kerja *machine learning* adalah dengan mengembangkan kemampuan pembelajaran mesin untuk menangani kumpulan data tertentu dan kemudian memberikan wawasan yang bermanfaat bagi orang yang menggunakannya. Kecerdasan buatan dan kumpulan data ini dimanfaatkan bidang kesehatan untuk mendeteksi kanker paru-paru pada pasien (Kharis et al., 2024).

Machine learning bagian dari kecerdasan buatan dimana algoritma dan metode tertentu digunakan untuk memprediksi, mengenala pola, dan mengklasifikasi. Analisis data yang dilakukan dikenal sebagai klasifikasi untuk membantu memprediksi label kelas mana yang harus diberikan kepada sampel. Berbagai bidang, seperti statistik, sistem pakar, dan pembelajaran mesin, telah mengusulkan berbagai metode klasifikasi. Algoritma *machine learning* termasuk *Decision Tree*, *K-Nearest Neighbor*, *Naive Bayes*, *Random Forest*, dan *Extreme Gradient Boosting* (Idris et al., 2024).

Deteksi dini sangat diperlukan untuk meningkatkan peluang kesembuhan. *Machine learning* dapat mendeteksi kanker paru-paru dalam data pasien dengan bantuan algoritma. Dengan teknologi yang semakin berkembang, *machine learning* diharapkan dapat memberikan hasil yang optimal serta dapat diandalkan sebagai solusi dari permasalahan penelitian ini. Hal tersebut memberikan harapan baru dalam mengurangi tingkat kematian akibat penyakit kanker paru-paru. Oleh karena itu, perlu dilakukan studi komparatif secara sistematis dan empiris untuk menganalisis performa berbagai algoritma *machine learning* dalam mengklasifikasikan jenis kanker paru-paru berdasarkan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*.

B. Identifikasi Masalah

Berdasarkan latar belakang diatas, masalah penelitian ini adalah deteksi dini kanker paru-paru sangat penting untuk mengurangi angka kematian akibat penyakit tersebut. Keterbatasan teknik skrining yang efektif untuk mendeteksi penyakit tersebut. Beberapa algoritma *machine learning* seperti *Decision Tree : C4.5*, *K-Nearest Neighbor*, *Naive Bayes*, *Random Forest*, dan *Extreme Gradient Boosting* digunakan untuk mendeteksi dan mengklasifikasi kanker paru-paru. Akan tetapi kurang bekerja dengan baik sehingga harus menganalisis kinerjanya untuk menemukan algoritma terbaik.

C. Batasan Masalah

Batasan masalah dari evaluasi yang dilakukan pada penelitian ini diantaranya :

1. Dataset yang dimanfaatkan dalam penelitian ini bersumber dari Kaggle <https://www.kaggle.com/datasets/chandanmsr/more-accurate-lung-cancer-dataset>
2. Ditemukan 3.000 data dengan 16 kolom berisi faktor-faktor yang dapat memicu pasien terkena kanker paru-paru.
3. Metode penelitian yang digunakan berupa algoritma klasifikasi yaitu termasuk *Decision Tree : C4.5*, *K-Nearest Neighbor*, *Naive Bayes*, *Random Forest*, dan *Extreme Gradient Boosting* untuk dianalisis tingkat keakuratannya.
4. Pemanfaatan *software* Google Colaboratory untuk mengolah data penelitian ini.

D. Rumusan Masalah

Dengan uraian diatas maka rumusan masalah akan diatasi adalah sebagai berikut:

1. Bagaimana algoritma *machine learning* dapat mengklasifikasi kanker paru-paru secara optimal dan efisien ?
2. Bagaimana metode klasifikasi kecerdasan buatan seperti *Decision Tree : C4.5*, *K-Nearest Neighbor*, *Naive Bayes*, *Random Forest*, dan *Extreme Gradient Boosting* bekerja ?

3. Algoritma *machine learning* mana yang terbaik untuk mendeteksi kanker paru-paru berdasarkan *classification report* dan *confusion matrix*?

E. Tujuan Penelitian

Penelitian ini dilakukan bertujuan untuk :

1. Untuk mengetahui algoritma *machine learning* mana yang secara optimal dan efisien mengklasifikasi kanker paru-paru.
2. Untuk mengetahui bagaimana metode klasifikasi kecerdasan buatan seperti *Decision Tree : C4.5, K-Nearest Neighbor, Naive Bayes, Random Forest, dan Extreme Gradient Boosting* bekerja.
3. Untuk mengetahui Algoritma *machine learning* mana yang terbaik untuk mengklasifikasi kanker paru-paru berdasarkan *classification report* dan *confusion matrix*.

F. Manfaat Penelitian

Nilai guna penelitian ini adalah penyediaan informasi baru tentang penggunaan algoritma pembelajaran mesin dalam mendeteksi kanker paru-paru. Hal ini juga membantu kita memahami cara kerja berbagai algoritma pembelajaran mesin saat menganalisis data medis, termasuk bagaimana kompleksitas algoritma, ukuran kumpulan data, dan kualitas data dapat memengaruhi akurasi deteksi. Selain itu, penelitian ini memberikan rekomendasi kepada institusi layanan kesehatan tentang pengintegrasian teknologi pembelajaran mesin dalam layanan sehari-hari untuk meningkatkan kualitas layanan kesehatan.

DAFTAR PUSTAKA

- Akbar, F., Saputra, H. W., Maulaya, A. K., Hidayat, M. F., & Rahmadden, R. (2022). Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 61–67. <https://doi.org/10.57152/MALCOM.V2I2.426>
- Amelia, D., Padilah, T. N., & Jamaludin, A. (2022). Optimasi Algoritma K-Means Menggunakan Metode Elbow dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat. *Jurnal Ilmiah Wahana Pendidikan*, 8(11), 207–215. <https://doi.org/10.5281/ZENODO.6831380>
- Amelia, Y. (2023). PERBANDINGAN METODE MACHINE LEARNING UNTUK MENDETEKSI PENYAKIT JANTUNG. *IDEALIS: InDonEsiA Journal Information System*, 6(2), 220–225. <https://doi.org/10.36080/IDEALIS.V6I2.3043>
- Ardyansyah, F., Daniati, E., & Ristyawan, A. (2024). Pemanfaatan Data Mining untuk Analisis Keputusan Perizinan Tenaga Kesehatan. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8(1), 52–62. <https://doi.org/10.29407/INOTEK.V8I1.4911>
- ASHURI, P. I., CAHYANI, I. A., & ADITYA, C. S. K. (2024). Klasifikasi Penyakit Stunting Menggunakan Algoritma Multi-Layer Perceptron. *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, 9(1), 52–63. <https://doi.org/10.26760/MINDJOURNAL.V9I1.52-63>
- Ayuningtyas, N., & Yustanti, W. (2024). Semi-Supervised Learning pada Pelabelan dalam Klasifikasi Multi-Label Data Teks. *Journal of Informatics and Computer Science (JINACS)*, 6(01), 240–248. <https://doi.org/10.26740/JINACS.V6N01.P240-248>
- Buana, I., & Harahap, D. A. (2022). ASBESTOS, RADON DAN POLUSI UDARA SEBAGAI FAKTOR RESIKO KANKER PARU PADA PEREMPUAN BUKAN PEROKOK. *AVERROUS: Jurnal Kedokteran Dan*

Kesehatan Malikussaleh, 8(1), 1–16.
<https://doi.org/10.29103/AVERROUS.V8I1.7088>

Daniati, E. (2019). Decision support systems to determining programme for students using DBSCAN and naive bayes: Case study: Engineering faculty of universitas nusantara PGRI kediri. *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, 238–243. <https://doi.org/10.1109/ICAIIIT.2019.8834474>

Desiani, A., Maiyanti, S. I., Andriani, Y., Suprihatin, B., Amran, A., Marselina, N. C., & Salsabila, A. (2023). Perbandingan Klasifikasi Penyakit Kanker Paru-paru menggunakan Support Vector Machine dan K-Nearest Neighbor. *Jurnal PROCESSOR*, 18(1). <https://doi.org/10.33998/PROCESSOR.2023.18.1.700>

Dewi, F. P., Aryni, P. S., & Umaidah, Y. (2022). Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 7(2), 111–121. <https://doi.org/10.14421/jiska.2022.7.2.111-121>

Fadli, M. N., Damanik, I. S., & Irawan, E. (2021). Penerapan Metode Naive Bayes Dalam Menentukan Tingkat Kenyamanan Pada Rumah Sakit Terhadap Pasien. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 2(3), 117–122. <https://doi.org/10.30865/KLIK.V2I3.297>

Habibi, H. A. N. S., Nugroho, A., & Firliana, R. (2023). PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN K-NEAREST NEIGHBORS UNTUK ANALISIS SENTIMEN COVID-19 DI TWITTER. *JURNAL ILMIAH INFORMATIKA*, 11(01), 54–62. <https://doi.org/10.33884/JIF.V11I01.7069>

Idris, J. F., Ramadhani, R., & Mutoffar, M. M. (2024). KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING. *Jurnal Media Akademik (JMA)*, 2(2). <https://doi.org/10.62281/V2I2.145>

- Kharis, S. A. A., Zili, A. H. A., Fajar, F. I., Putri, A., & Arisanty, M. (2024). A Literature Review to Evaluate the Impact of Machine Learning and Artificial Intelligence for Lung Cancer Patient in COVID-19 Pandemic. *G-Tech: Jurnal Teknologi Terapan*, 8(1), 576–583. <https://doi.org/10.33379/GTECH.V8I1.3891>
- Kriswantara, B., & Sadikin, R. (2022). Machine Learning Used Car Price Prediction with Random Forest Regressor Model. *Journal of Information System, Informatics and Computing*, 6(1), 40–49. <https://doi.org/10.52362/JISICOM.V6I1.752>
- Kurnia, O. D., A’Fena, E. T., Ningrum, D. Y. A., Daniati, E., & Ristyawan, A. (2024). Analisis Perbandingan Algoritma Naïve Bayes dengan K-Nearest Neighbor (KNN) Pada Dataset Mobile Price Classification. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8(2), 1174–1183. <https://doi.org/10.29407/INOTEK.V8I2.5053>
- Kurniawan, D., Wahyudi, M., Pujiastuti, L., & Sumanto, S. (2024). Deteksi dan Prediksi Cerdas Penyakit Paru-Paru dengan Algoritma Random Fores. *Indonesian Journal Computer Science*, 3(1), 51–56. <https://doi.org/10.31294/IJCS.V3I1.6071>
- Kurniawan, R., Wintoro, P. B., Mulyani, Y., & Komarudin, M. (2023). IMPLEMENTASI ARSITEKTUR XCEPTION PADA MODEL MACHINE LEARNING KLASIFIKASI SAMPAH ANORGANIK. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(2). <https://doi.org/10.23960/JITET.V11I2.3034>
- Kurniawan, W., & Indahyanti, U. (2024). Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode XGBoost. *Indonesian Journal of Applied Technology*, 1(2), 18–18. <https://doi.org/10.47134/IJAT.V1I2.3045>
- Liang, S. (2021). Comparative Analysis of SVM, XGBoost and Neural Network on Hate Speech Classification. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5), 896–903. <https://doi.org/10.29207/RESTI.V5I5.3506>

- Maciej Serda, Becker, F. G., Cleary, M., Team, R. M., Holtermann, H., The, D., Agenda, N., Science, P., Sk, S. K., Hinnebusch, R., Hinnebusch A, R., Rabinovich, I., Olmert, Y., Uld, D. Q. G. L. Q., Ri, W. K. H. U., Lq, V., Frxqwu, W. K. H., Zklfk, E., Edvhg, L. V, ... فاطمی, ح. (2014). KLASIFIKASI JENIS BIMBINGAN DAN KONSELING SISWA SMKN 1 KEDIRI MENGGUNAKAN NAIVEBAYES CLASSIFIER DAN NEAREST NEIGHBOR. *Nusantara of Engineering (NOE)*, 1(2), 343–354. <https://doi.org/10.2/JQUERY.MIN.JS>
- Mahanum, T. K. (2021). Tinjauan Kepustakaan. *ALACRITY : Journal of Education*, 1(2), 1–12. <https://doi.org/10.52121/ALACRITY.V1I2.20>
- Nugroho, A. (2022). Analisa Splitting Criteria Pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan. *JSITIK: Jurnal Sistem Informasi Dan Teknologi Informasi Komputer*, 1(1), 41–49. <https://doi.org/10.53624/JSITIK.V1I1.154>
- Nurhidayati, L., Umaidah, Y., & Enri, U. (2024). Analisis Sentimen Isu Childfree Di Media Sosial Twitter Menggunakan Algoritma Support Vector Machine. *Jurnal Ilmiah Wahana Pendidikan*, 10(4), 422–430. <https://doi.org/10.5281/ZENODO.10521284>
- Nurjanah, I., Karaman, J., Widaningrum, I., Mustikasari, D., & Sucipto, S. (2023). Penggunaan Algoritma Naïve Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa. *Explorer*, 3(2), 77–87. <https://doi.org/10.47065/EXPLORER.V3I2.766>
- Permana, A. H., Umbara, F. R., & Kasyidi, F. (2024). Klasifikasi Penyakit Jantung Tipe Kardiovaskular Menggunakan Adaptive Synthetic Sampling dan Algoritma Extreme Gradient Boosting. *Building of Informatics, Technology and Science (BITS)*, 6(1), 499–508. <https://doi.org/10.47065/BITS.V6I1.5421>
- Pratama, D. A., Mutaqin, I. R., & Manuela, K. R. (2023). Analisis Terjadinya Kanker Paru-Paru Pada Pasien Menggunakan Decision Tree: Penerapan Algoritma C4.5 Dan RapidMiner Untuk Menentukan Risiko Kanker Pada

- Gejala Pasien. *Jurnal Teknik Mesin, Industri, Elektro Dan Informatika*, 2(4), 156–170. <https://doi.org/10.55606/JTMEI.V2I4.3004>
- Putra, T. D., Utami, E., & Kurniawan, M. P. (2022). Klasifikasi penderita kanker Paru Paru Menggunakan Algoritma Artificial Neural Network (ANN). *Explore*, 12(2), 139–143. <https://doi.org/10.35200/EX.V12I2.67>
- Ridwan, achmad. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 15–21. <https://doi.org/10.47970/SISKOM-KB.V4I1.169>
- Rininda, G., Santi, I. H., & Kirom, S. (2023). PENERAPAN SVM DALAM ANALISIS SENTIMEN PADA EDLINK MENGGUNAKAN PENGUJIAN CONFUSION MATRIX. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3335–3342. <https://doi.org/10.36040/JATI.V7I5.7420>
- Risqi, A., & Nasution, S. (2021). Identifikasi Permasalahan Penelitian. *ALACRITY : Journal of Education*, 1(2), 13–19. <https://doi.org/10.52121/ALACRITY.V1I2.21>
- Rofiani, R., Oktaviani, L., Vernanda, D., & Hendriawan, T. (2024). Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5. *Jurnal Tekno Kompak*, 18(1), 126–139. <https://doi.org/10.33365/JTK.V18I1.3525>
- Sanjaya, R., Tohidi, E., Wahyudi, E., & Kec Kesambi Kota Cirebon, M. (2024). ANALISIS SENTIMEN TERHADAP BERHENTINYA TIKTOKSHOP PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 507–514. <https://doi.org/10.36040/JATI.V8I1.8443>
- Sari, P. K., & Suryono, R. R. (2024). KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS

SENTIMEN METAVERSE. *Jurnal Mnemonic*, 7(1), 31–39.
<https://doi.org/10.36040/MNEMONIC.V7I1.8977>

Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023a). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19.
<https://doi.org/10.57152/MALCOM.V3I1.591>

Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023b). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19.
<https://doi.org/10.57152/MALCOM.V3I1.591>

Septiani, I. W., Fauzan, Abd. C., & Huda, M. M. (2022). Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 556–566.
<https://doi.org/10.30865/JSON.V3I4.4055>

Shafa, B., Handayani, H. H., Lestari, S. A. P., & Cahyana, Y. (2024). Prediksi Kanker Paru dengan Normalisasi menggunakan Perbandingan Algoritma Random Forest, Decision Tree dan Naïve Bayes. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(3), 1057–1070.
<https://doi.org/10.51454/DECODE.V4I3.779>

Sidik, A. D., & Ansawarman, A. (2022). Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning. *Formosa Journal of Multidisciplinary Research*, 1(3), 559–568. <https://doi.org/10.55927/FJMR.V1I3.745>

Simanjuntak, A. Y., Simatupang, I. S. eptian S., & Anita, A. (2022). IMPLEMENTASI DATA MINING MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER UNTUK DATA KENAIKAN PANGKAT DINAS

KETENAGAKERJAAN KOTA MEDAN. *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, 5(1), 85–91. <https://doi.org/10.54314/JSSR.V5I1.804>

Sirait, M. T. T. B., Fathonah, N. S., & Fauzan, M. N. (2024). PEMANFAATAN ALGORITMA ADASYN DAN SUPPORT VECTOR MACHINE DALAM MENINGKATKAN AKURASI PREDIKSI KANKER PARU-PARU. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 8773–8778. <https://doi.org/10.36040/JATI.V8I5.10752>

Suhirman, S. (2023). Data Warehouse dan Mining. *GAES - PACE Book Publisher*, 14–25. <https://digitalpress.gaes-edu.com/index.php/gaespace/article/view/122>

Suriani, N., Risnita, & Jailani, M. S. (2023). Konsep Populasi dan Sampling Serta Pemilihan Partisipan Ditinjau Dari Penelitian Ilmiah Pendidikan. *IHSAN: Jurnal Pendidikan Islam*, 1(2), 24–36. <https://doi.org/10.61104/IHSAN.V1I2.55>

Susanti, N., Nuraida, A., Amanda, I. A., & Khairunnisa, K. (2024). PENCEGAHAN DAN PENANGGULANGAN PENYAKIT TIDAK MENULAR PADA REMAJA. *Jurnal Kesehatan Tambusai*, 5(2), 4223–4233. <https://doi.org/10.31004/JKT.V5I2.28636>

Tarigan, L. R. A., & Dahlan, D. (2024). OPTIMALISASI FITUR DENGAN FORWARD SELECTION PADA ESTIMASI TINGKAT PENYAKIT PARU-PARU MENGGUNAKAN ALGORITMA KLASIFIKASI RANDOM FOREST. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 10341–10348. <https://doi.org/10.36040/JATI.V8I5.11064>

Tholib, A., Hidayat, M. N. F., yono, S., Wulanningrum, R., & Daniati, E. (2023). Comparison of C4.5 and Naive Bayes for Predicting Student Graduation Using Machine Learning Algorithms. *International Journal of Engineering and Computer Science Applications (IJECSA)*, 2(2), 71–78. <https://doi.org/10.30812/IJECSA.V2I2.3364>

- Trisnawati, A., Surani, D., Kurniawan, B. S., & Fidriyanto, A. (2024). ANALYSIS OF MATERIAL UNDERSTANDING USING GOOGLE COLABORATORY IN CLASS X INFORMATICS SUBJECTS AT SMAN 5 SERANG. *Cakrawala Pedagogik*, 8(1), 38–45. <https://doi.org/10.51499/CP.V8I1.627>
- Wahid, M. A. R., Nugroho, A., & Anshor, A. H. (2023). Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier. *Bulletin of Information Technology (BIT)*, 4(1), 63–74. <https://doi.org/10.47065/BIT.V4I1.501>
- Wibisono, A. B., & Fahrurrozi, A. (2020). PERBANDINGAN ALGORITMA KLASIFIKASI DALAM PENGKLASIFIKASIAN DATA PENYAKIT JANTUNG KORONER. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 24(3), 161–170. <https://doi.org/10.35760/TR.2019.V24I3.2393>
- Widiarahman, E., Irawan, B., Bahtiar, A., No, J. P., Majasem, B., Kesambi, K., & Cirebon, K. (2024). IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENGELOMPOKAN TINGKAT GAJI DI INDONESIA BERDASARKAN PROVINSI. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2824–2829. <https://doi.org/10.36040/JATI.V8I3.9565>
- Widiarni, A., & Mustakim, M. (2021). Penerapan Algoritma Support Vector Regression untuk Prediksi Jumlah Pasien Covid-19 di Provinsi Riau. *Building of Informatics, Technology and Science (BITS)*, 3(2), 71–78. <https://doi.org/10.47065/BITS.V3I2.1004>
- Wijaya, C., Irsyad, H., & Widhiarso, W. (2020). KLASIFIKASI PNEUMONIA MENGGUNAKAN METODE K-NEAREST NEIGHBOR DENGAN EKSTRAKSI GLCM. *Jurnal Algoritme*, 1(1), 33–44. <https://doi.org/10.35957/ALGORITME.V1I1.431>
- Wiwi, W., & Maulani, Y. (2024). ANALISIS PERBANDINGAN KUALITAS PEWARNAAN PAPANICOLAOU DAN GIEMSA PADA BILASAN BRONKUS KARSINOMA PARU. *Plenary Health : Jurnal*

Kesehatan Paripurna, 1(3), 430–433.

<https://doi.org/10.37985/PLENARYHEALTH.V1I3.646>

Chandan, M. S. R. (2021). *More Accurate Lung Cancer Dataset* [Dataset]. Kaggle.

<https://www.kaggle.com/datasets/chandanmsr/more-accurate-lung-cancer-dataset>