

ANALISIS PERBANDINGAN ALGORITMA MACHINE LEARNING KLASIFIKASI UNTUK DETEKSI TINGKAT KEGANASAN PENYAKIT KANKER PAYUDARA

Muhammad Raffi¹, Indra Gunawan^{2*}

^{1,2} Prodi Informatika, Sekolah Tinggi Teknologi Ronggolawe, *Penulis Koresponden
e-mail:²igunstr@gmail.com

ABSTRACT

This research focuses on implementing Machine Learning (ML), an Artificial Intelligence (AI) branch, to enhance breast cancer detection. The Synthetic Minority Over-sampling Technique (SMOTE) method is implemented for the first time in this study. The research compares various ML algorithms to improve the accuracy and efficiency of diagnosing the malignancy level of breast cancer, using the Wisconsin Breast Cancer dataset. Support Vector Machine (SVM) is identified as the best-performing algorithm, demonstrating high accuracy, a high Area Under the Curve (AUC), and good precision. Experimental results show the highest accuracy with an AUC value close to perfection (0.99). Furthermore, the identification of 10 factors causing breast cancer malignancy provides valuable insights. Despite contributing significantly to developing more effective detection methods, the research has two main limitations: reliance on a single dataset and the potential for expanding experiments by testing more classification algorithms. In conclusion, this study supports efforts for more effective breast cancer detection, hoping its findings can be applied more broadly.

Keywords: breast cancer, classification, machine learning, SVM

INTISARI

Penelitian ini fokus pada penerapan Machine Learning (ML), sebuah cabang dari Artificial Intelligence (AI), untuk meningkatkan deteksi kanker payudara. Metode Synthetic Minority Over-sampling Technique (SMOTE) diimplementasikan untuk pertama kalinya dalam penelitian ini. Penelitian membandingkan berbagai algoritma ML guna meningkatkan akurasi dan efisiensi diagnosis tingkat keganasan kanker payudara, dengan menggunakan dataset Wisconsin Breast Cancer. Support Vector Machine (SVM) teridentifikasi sebagai algoritma terbaik, menunjukkan akurasi tinggi, AUC yang tinggi, dan presisi baik. Hasil eksperimen menunjukkan akurasi tertinggi dengan nilai AUC mendekati sempurna (0.99). Selain itu, identifikasi 10 faktor penyebab keganasan kanker payudara memberikan wawasan berharga. Meskipun penelitian ini memberikan kontribusi penting terhadap pengembangan metode deteksi yang lebih efektif, terdapat dua keterbatasan utama, yaitu pembatasan pada satu dataset dan potensi perluasan eksperimen dengan menguji lebih banyak algoritma klasifikasi. Kesimpulannya, penelitian ini mendukung upaya deteksi kanker payudara yang lebih efektif, dengan harapan hasilnya dapat diterapkan secara lebih umum setelah pengujian lebih lanjut.

Kata kunci: , kanker payudara, klasifikasi, machine learning, SVM

1. PENDAHULUAN

Machine Learning adalah cabang dari Ilmu Kecerdasan Buatan (*Artificial Intelligence*) yang bertujuan memberikan kemampuan pada mesin untuk belajar secara otomatis tanpa perlu diprogram secara eksplisit, dengan memanfaatkan dataset (Pratiwi, 2021). Penggunaan Machine Learning memiliki beberapa manfaat di bidang kesehatan (Sutarmi, 2023), (Indrayana, 2022). Dalam konteks kesehatan, kanker payudara menjadi fokus penting karena deteksi dini kunci untuk hasil pengobatan yang lebih baik. Algoritma Machine Learning digunakan untuk memproses data klinis kompleks dan memprediksi tingkat kanker secara akurat.

Dalam beberapa penelitian yang telah dilakukan sebelumnya, oleh (Oktavianto & Handri, 2020) (Chazar & Erawan, 2020) yang melakukan analisis klasifikasi kanker payudara menggunakan algoritma *Naive Bayes*, membuat *Machine Learning* diagnosis kanker payudara menggunakan algoritma *Naive Bayes*, membuat *Machine Learning* diagnosis kanker payudara menggunakan algoritma *Support Vector Machine*. Penelitian tersebut membandingkan tingkat akurasi klasifikasi kanker payudara. Dalam penelitian (A.Vincent, 2022), dilakukan perbandingan analisis dari beberapa algoritma yang akan dioptimalkan dengan menggabungkan teknik *Naive Bayes* dan seleksi fitur

menggunakan teknik *SMOTE*. Tujuannya adalah untuk meningkatkan efisiensi pendeteksian tingkat keganasan pada kanker payudara.

Dalam hal ini, *Machine Learning* dapat digunakan untuk mengembangkan model klasifikasi yang dapat memprediksi tingkat keganasan kanker payudara. Algoritma *Machine Learning* yang digunakan untuk klasifikasi kanker payudara harus mampu mengenali pola-pola dalam data dan membedakan kanker payudara yang ganas dari yang tidak ganas dengan akurasi yang tinggi. Penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang performa algoritma *Machine Learning* yang berbeda dalam deteksi tingkat keganasan kanker payudara. Hasilnya diharapkan dapat membantu mengidentifikasi algoritma *Machine Learning* yang paling efektif dalam deteksi kanker payudara, dengan potensi meningkatkan akurasi dan efektivitas diagnosis serta pengobatan.

2. METODE PENELITIAN

2.1 Metode Kanker Payudara

Kanker payudara adalah jenis penyakit yang terjadi ketika sel-sel abnormal di jaringan payudara mulai tumbuh secara tidak terkendali dan dapat menyebar ke bagian tubuh yang lain. Kanker payudara biasanya berasal dari *epitel duktus* (sel yang membentuk saluran susu) atau *lobulus* (sel yang menghasilkan susu), dan dapat terjadi pada wanita maupun pria, meskipun lebih umum terjadi pada wanita. Berdasarkan data dari *The Global Cancer Observatory* pada tahun 2020, kanker payudara menduduki urutan pertama penyebab kematian karena kanker di Indonesia (Sesrianty & Amalia, 2023).

2.2 Algoritma *Machine Learning*

Machine Learning biasanya digunakan untuk mengajarkan mesin cara menangani data dengan lebih efisien. Dalam beberapa kasus, kita mungkin tidak dapat menginterpretasikan informasi yang diperoleh dari data. Dalam hal ini, kita menggunakan *Machine Learning*. Dengan banyaknya ketersediaan dataset, maka permintaan dalam penggunaan *Machine Learning* ini semakin meningkat. Banyak industri yang menggunakan *Machine Learning* untuk memperoleh data yang relevan. Kegunaan dari *Machine Learning* adalah untuk belajar dari data. Banyak penelitian telah dilakukan yang memungkinkan mesin dapat belajar secara mandiri tanpa memerlukan pemrograman yang eksplisit. Matematikawan dan *Programmer* sering menggunakan pendekatan untuk menemukan solusi pada masalah yang melibatkan dataset.

Untuk menyelesaikan masalah data, *Machine Learning* menggunakan berbagai jenis algoritma. Namun menurut ilmuwan data, tidak ada satu algoritma yang cocok untuk menyelesaikan semua jenis masalah yang ingin diselesaikan, jumlah variabel yang terlibat, jenis model yang paling sesuai, dan faktor-faktor lainnya (Batta, 2018).

a. Jenis-Jenis Algoritma

1. *Naïve Bayes*

Naive Bayes adalah sebuah metode pengklasifikasian probabilistik yang sederhana. Metode ini melakukan perhitungan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Metode *Naive Bayes* mengasumsikan bahwa semua atribut dalam setiap kategori tidak saling tergantung satu sama lain (independen) (Hanifah, 2024), (Devita et al., 2018).

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data class spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)
- P(H) : Probabilitas hipotesis H (*prior probability*)
- P(X|H) : Probabilitas X berdasarkan kondisi H (*likelihood probability*)
- P(X) : Probabilitas X (*evidence probability*) (Anis Zubair, 2018)

2. *K-Nearest Neighbor (KNN)*

Metode *K-Nearest Neighbor (KNN)* merupakan sebuah teknik pengklasifikasian yang bersifat *supervised*, yang berarti ia memerlukan data training untuk dapat mengklasifikasikan objek berdasarkan jarak

terdekatnya. Prinsip kerja KNN adalah dengan mencari jarak terdekat antara data yang sedang dievaluasi dengan k tetangga terdekat dalam data pelatihan (Devita et al., 2018).

$$D(X, Y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2}$$

Keterangan :

- D : Jarak antara dua titik x dan y
- X : Data uji spesifik
- Y : Sampel data
- n : Dimensi data (Septian et al., 2019)

3. Decision Tree

Decision Tree adalah suatu struktur yang mirip dengan *flowchart*, di mana setiap node internal (node yang bukan *leaf* atau node terluar) digunakan untuk menguji variabel atribut, dan setiap cabangnya mewakili hasil dari pengujian tersebut. Sementara itu, node terluar (*leaf*) dalam *Decision Tree* menjadi label atau klasifikasi akhir (Nurul, 2024) (Sutoyo, 2018).

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2$$

Keterangan :

- S : Himpunan kasus
- n : Jumlah partisi S
- Pi : Proporsi Si terhadap S (Cynthia, 2018)

4. Random Forest

Random Forest adalah sebuah model *ensemble* yang terdiri dari banyak model *Decision Tree*, baik untuk regresi maupun klasifikasi. Model ini menggunakan metode *bootstrap aggregating (bagging)* dan *random feature selection*. Dalam metode ini, setiap model *Decision Tree* dibangun secara independen dengan menggunakan subset acak dari data pelatihan (*bootstrap*) dan subset acak dari fitur (*random feature selection*) (Purwa, 2019).

$$Entropy(Y) = \sum_i p(c|Y) \log^2 p(c|Y)$$

Keterangan :

- Y : Himpunan kasus
- P(c|Y) : Proporsi nilai Y terhadap kelas c (Sandag, 2020)

5. Super Vector Machine (SVM)

Super Vector Machine (SVM) adalah sebuah algoritma *Machine Learning* yang beroperasi berdasarkan prinsip *Structural Risk Minimization (SRM)* dengan tujuan mencari *hyperplane* terbaik yang dapat memisahkan dua kelas pada ruang input (Rahutomo et al., 2018).

$$D_{ij} = y_i y_j (K(\vec{x}_i \cdot \vec{x}_j) + \lambda^2)$$

Keterangan:

- D_{ij} : Jarak antara dua titik x dan y
- y_i : Data uji spesifik
- y_j : Sampel data
- λ : Dimensi data (Septian et al., 2019)
- $(K(\vec{x}_i \cdot \vec{x}_j))$: Fungsi kernel (Tineges et al., 2020)

6. Neural Network

Neural network merupakan sebuah model algoritma yang terinspirasi dari cara kerja neuron dalam otak manusia. Mirip dengan otak manusia, setiap neuron dalam *neural network* saling terhubung dan mengalirkan informasi antara satu dengan yang lainnya (Ridwan et al., 2020).

$$input_j = \sum_{i=1}^n o_i w_{ij} + \theta_j$$

Keterangan:

- o_i : Output simpul i dari layer sebelumnya
- w_{ij} : Bobot relasi dari simpul i pada layer sebelumnya ke simpul j
- θ_j : Bias (Rohman & Rocham, 2019)

2.3 Klasifikasi

Asal-usul kata "klasifikasi" berasal dari bahasa Latin "*classis*" yang berarti mengelompokkan benda-benda yang serupa dan membedakan benda-benda yang berbeda. Secara sederhana, klasifikasi dapat diartikan sebagai proses pengelompokan atau penggolongan suatu objek atau benda berdasarkan ciri-ciri atau karakteristik yang sama (Ansori, 2015).

Beberapa metode klasifikasi umum dalam machine learning seperti *K-Nearest Neighbors (KNN)*, *Decision Trees*, *Naive Bayes*, *Support Vector Machines (SVM)*, *Random Forest*, dan *Neural Networks* memiliki keunggulan, kelemahan, serta kondisi penggunaan yang berbeda-beda. Pemilihan metode klasifikasi sangat bergantung pada karakteristik data, skala masalah, kompleksitas data, serta tujuan akhir dari analisis yang diinginkan.

2.4 Deteksi Tingkat Keganasan Kanker Payudara

Dalam penilaian tingkat keganasan, terdapat dua kategori yaitu B atau M. Kategori B mengindikasikan kanker jinak, sementara kategori M mengindikasikan kanker ganas. Meskipun kategori B dan M sebenarnya memiliki tingkat keganasan yang berbeda-beda, namun karena hanya terdapat dua kategori, maka hasil penilaian tidak akan berbeda antara menggunakan regresi logistik binary atau regresi logistik ordinal. Oleh karena itu, dalam hal ini, sifat ordinal dari tingkat keganasan tidak terlalu berpengaruh (Vinarti & Anggraeni, 2015).

Objek penelitian dalam deteksi tingkat keganasan kanker payudara melibatkan analisis mendalam terhadap sejumlah variabel yang berperan dalam mengevaluasi tingkat keganasan kanker payudara pada individu. Studi ini meliputi:

Data Pasien:

- Gejala dan Tanda : Meliputi gejala atau tanda-tanda kanker payudara yang dirasakan oleh pasien.
- Riwayat Kesehatan : Termasuk riwayat keluarga, riwayat pribadi, dan faktor risiko yang terkait dengan kanker payudara.
- Hasil Tes Diagnostik : Data hasil tes pencitraan (seperti mammografi, USG, MRI), biopsi, dan tes lainnya yang membantu dalam diagnosis.
- Profil Genetik : Informasi tentang pola genetik atau ekspresi gen yang berhubungan dengan risiko atau tingkat keganasan kanker payudara.

Variabel Penelitian:

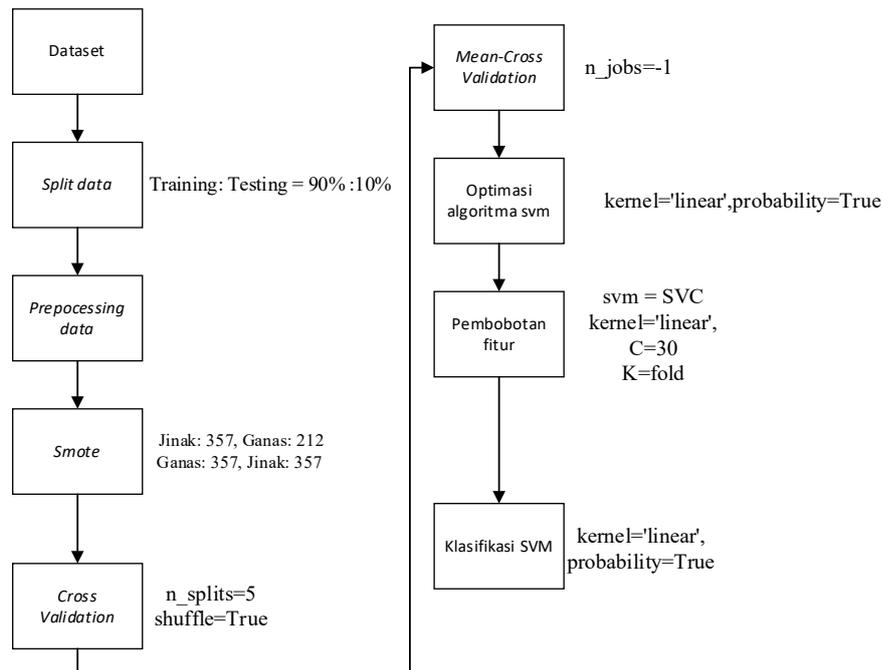
- Karakteristik Sel Kanker : Fokus pada ciri-ciri sel kanker yang memengaruhi tingkat keganasan, seperti ukuran, bentuk, tekstur, dan tingkat proliferasi.
- Faktor Biologis : Meliputi peran hormon, protein, atau aspek biologis lain yang berperan dalam perkembangan kanker payudara.
- Respon terhadap Pengobatan : Memeriksa bagaimana pasien merespons berbagai jenis pengobatan.

Metode Analisis:

- *Machine Learning* : Penggunaan algoritma dan teknik machine learning untuk mengenali pola dalam data yang berkaitan dengan tingkat keganasan.
- Analisis Genomik : Meneliti perubahan genetik dalam tumor untuk mengidentifikasi marker risiko atau keganasan.
- Statistik dan Model Matematika : Penggunaan model statistik dan matematis untuk menganalisis hubungan antar variabel.

Objek penelitian ini memungkinkan peneliti memahami faktor-faktor yang terkait dengan keganasan kanker payudara, mengidentifikasi pola yang bermanfaat untuk prediksi, serta mengembangkan strategi pendekatan yang lebih efektif dalam pencegahan, diagnosis, dan pengobatan kanker payudara.

3. HASIL DAN PEMBAHASAN



Gambar 3.1 Diagram Hasil Perbandingan

Proses pengolahan data dimulai dengan membagi dataset menjadi dua bagian, yakni data latih (90%) dan data uji (10%). Setelahnya, dilakukan langkah preprocessing untuk menyesuaikan skala data agar memiliki rata-rata nol dan deviasi standar satu. Tahap berikutnya, yaitu SMOTE, bertujuan untuk menyeimbangkan kelas data dengan menambah jumlah data manipulasi. Contohnya, pada awalnya kelas "Ganas" memiliki 212 data dan kelas "Jinak" memiliki 357 data, setelah SMOTE, keduanya memiliki 357 data, dengan penambahan 145 data pada kelas "Ganas". Setelah itu, proses cross-validation dilakukan lima kali untuk menguji model dengan skor AUC, memastikan pengujian data diacak untuk menghindari bias. Hasil dari kelima pengujian diambil rata-ratanya untuk menciptakan mean cross-validation.

Selanjutnya, fokus beralih ke tahap optimasi algoritma SVM, khususnya pada bagian kernel. Setelah proses optimasi, kernel linear dipilih. Pengaturan probabilitas prediksi diatur menjadi "true" agar dapat menghitung probabilitas prediksi untuk setiap kelas. Langkah berikutnya adalah pembobotan fitur untuk menentukan fitur yang paling berpengaruh terhadap tingkat keganasan kanker payudara. Akhirnya, implementasi semua optimasi dilakukan pada algoritma SVM sebagai tahap terakhir dari proses ini, dengan tujuan menghasilkan model yang efektif dalam mendeteksi kanker payudara.

3.1. Sub bab

Tabel 4.1 Hasil Model SVM

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
<i>malignant</i>	1.00	0.96	0.98	74
<i>benign</i>	0.96	1.00	0.98	69
<i>accuracy</i>			0.98	143
<i>macro avg</i>	0.98	0.98	0.98	143
<i>wighted avg</i>	0.98	0.98	0.98	143
<i>AUC Score</i>	0.97179			

Hasil dari model SVM yang digunakan menunjukkan performa yang sangat baik dalam mendeteksi kanker

payudara dengan akurasi sebesar 98%. Untuk kelas "Ganas" (*malignant*), model memiliki nilai *precision* sebesar 1.00 dan *recall* sebesar 0.96, yang berarti model cenderung memberikan prediksi yang benar untuk kasus kanker payudara yang sebenarnya (*malignant*), namun mungkin ada sedikit kasus yang terlewat. Sementara untuk kelas "Jinak" (*benign*), model memiliki *precision* sebesar 0.96 dan *recall* sebesar 1.00, yang menunjukkan bahwa model cenderung memberikan prediksi yang benar untuk semua kasus kanker yang sebenarnya tidak ganas (*benign*).

Dengan *Fi-score* sebesar 0.98 untuk kedua kelas, model ini memiliki keseimbangan yang baik antara *precision* dan *recall* untuk kedua jenis kanker payudara. *AUC score* (*Area Under Curve*) yang cukup tinggi, yaitu 0.9717897, menunjukkan bahwa model ini memiliki kinerja yang baik dalam memisahkan antara kelas positif dan negatif dalam pengujian. Dengan demikian, dapat disimpulkan bahwa model SVM ini sangat baik dalam deteksi kanker payudara dengan kinerja yang tinggi dan dapat diandalkan. Dan pembahasan dapat disajikan ke beberapa sub bab dengan format 10 pt, Times bold. Format tabel seperti pada Tabel 1. Format gambar seperti Gambar 1. Setiap tabel dan gambar harus dikutip di paragraf, dengan cara menulis "Gambar 1" "Tabel 1". Gambar harus memiliki resolusi yang tinggi sehingga citra yang dihasilkan tajam dan jelas.

Tabel 1. Pengujian Sistem Minimum Mikrokontroler AVR ATmega 16

No	Bagian Pengujian	Hasil Pengujian	
		Berhasil	Gagal
1	Port A, B, C dan D	√	-
3	USART	√	-
4	LCD	√	-

4. KESIMPULAN

Kesimpulan yang didapatkan dari penelitian ini adalah algoritma *SVM* mendapatkan akurasi tertinggi dibanding 6 algoritma lainnya pada kasus klasifikasi tingkat keganasan pada kanker payudara menggunakan dataset *Wisconsin Breast Cancer*. Pada proses pengolahan data, dataset telah melalui proses *SMOTE* dan *preprocessing* agar tidak terjadi *missing value* serta dapat diolah dalam program Machine Learning dengan menggunakan bahasa pemrograman *Python*. Algoritma *SVM* menunjukkan kinerja yang sangat baik dengan akurasi sebesar 0.98, presisi 1.00, *recall* 0.98, *FI-score* 0.98, *AUC* 0.98, dan rata-rata *AUC* dari validasi silang sebesar 0.97. Hal ini dicapai dengan menggunakan kernel linier yang menghasilkan nilai *AUC* hampir sempurna sebesar 0.971789.

DAFTAR PUSTAKA

- A.Vincent, J. P. J. F. (2022). Komparasi Tingkat Akurasi Random Forest Dan Knn Untuk Mendiagnosis Penyakit Kanker Payudara. *Universitas Pelita Harapan PSDKU Medan Jurusan Sistem Informasi*, 7(1), 49–61.
- Alfarizi, M. R. sirfatullah, Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman Untuk Machine Learning dan Deep Learning. *Karimah Tauhid*, 2(1), 1–6.
- Anis Zubair, M. M. (2018). Penerapan Metode Naive Bayes Untuk Klasifikasi Status Gizi (Studi Kasus Di Klinik Bromo Malang). *Jurnal Gaussian*.
- Ansori. (2015). Pembahasan Klasifikasi. *Paper Knowledge . Toward a Media History of Documents*, 3(April), 49–58.
- Batta, M. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 18(8), 381–386. <https://doi.org/10.21275/ART20203995>
- Chazar, C., & Erawan, B. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 12(1), 67–80. <https://doi.org/10.37424/informasi.v12i1.48>
- Cynthia, E. P. (2018). *MENGLASIFIKASI DATA PENJUALAN BISNIS GERAJ*. July. <https://doi.org/10.30645/jurasik.v3i0.60>
- Derisma, D., & Febrian, F. (2020). Perbandingan Teknik Klasifikasi Neural Network, Support Vector Machine, dan Naive Bayes dalam Mendeteksi Kanker Payudara. *Bina Insani Ict Journal*, 7(1), 53. <https://doi.org/10.51211/biict.v7i1.1343>
- Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 427. <https://doi.org/10.25126/jtiik.201854773>
- Hanifah, S., & Gunawan, I. (2024, July). IMPLEMENTASI METODE NAIVE BAYES PADA EVALUASI KEPUASAN MAHASISWA DALAM LAYANAN AKADEMIK DAN NON AKADEMIK. In *Prosiding SENDIKO (Seminar Nasional Hasil Penelitian & Pengabdian Masyarakat Bidang Ilmu Komputer)* (Vol. 3).

- Indrayana, T., Warijan, W., Sutarmi, S., Purnomo, D., & Gunawan, I. (2022). Developing systems application based on android as tool for determinant stunting factors in the COVID-19 pandemic era. *International Journal of Health Sciences*, 6(S1), 257-268.
- Nurul, B., & Gunawan, I. (2024, July). Penerapan Machine Learning Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree. In *Prosiding SENDIKO (Seminar Nasional Hasil Penelitian & Pengabdian Masyarakat Bidang Ilmu Komputer)* (Vol. 3).
- Oktavianto, H., & Handri, R. P. (2020). Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes. *INFORMAL: Informatics Journal*, 4(3), 117. <https://doi.org/10.19184/isj.v4i3.14170>
- Pratiwi, D. (2021). *IMPLEMENTASIMETODE SIMPLE ADDITIVE WEIGHTING DAN MACHINE LEARNING PADAAPLIKASI REKOMENDASI PRODUK SKINCAREBERDASARKANMASALAH KULITWAJAHBERBASIS ANDROID*.
- Purwa, T. (2019). *Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced*. 16(1), 58–73. <https://doi.org/10.20956/jmsk.v16i1.6494>
- Rahutomo, F., Saputra, P. Y., & Fidyawan, M. A. (2018). Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Polinema*, 4(2), 93. <https://doi.org/10.33795/jip.v4i2.152>
- Ridwan, R., Lubis, H., & Kustanto, P. (2020). Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Media Informatika Budidarma*, 4(2), 286. <https://doi.org/10.30865/mib.v4i2.2035>
- Rohman, A., & Rocham, M. (2019). *KOMPARASI METODE KLASIFIKASI DATA MINING UNTUK PREDIKSI Abstrak*. 5(1), 23–29.
- Sandag, G. A. (2020). *Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest Application Rating Prediction on App Store using Random Forest Algorithm*. 6(2), 167–178.
- Saragih, R. R. (2016). Pemrograman dan bahasa Pemrograman. *STMIK-STIE Mikroskil, December*, 1–91.
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *Journal of Intelligent System and Computation*, 1(1), 43–49. <https://doi.org/10.52985/insyst.v1i1.36>
- Sesrianty, V., & Amalia, E. (2023). Edukasi Pencegahan Kanker Payudara Melalui Deteksi Dini Sadari Di Jorong Sungai Sariak. *Community Development Journal*, 4(1), 423–427.
- Sutarmi, S., Warijan, W., Indrayana, T., & Gunawan, I. (2023). Machine Learning Model For Stunting Prediction. *Jurnal Health Sains*, 4(9), 10-23.
- Sutoyo, I. (2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Jurnal Pilar Nusa Mandiri*, 14(2), 217. <https://doi.org/10.33480/pilar.v14i2.926>
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). *Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)*. 4, 650–658. <https://doi.org/10.30865/mib.v4i3.2181>
- Via, Y. V., Nugroho, B., & Syafrizal, A. (2015). Sistem Pendukung Keputusan Klasifikasi Tingkat Keganasan Kanker Payudara Dengan Metode Naive Bayes Classifier. *SCAN-Jurnal Teknologi Informasi Dan Komunikasi*, 10(2), 64–65. <http://ejournal.upnjatim.ac.id/index.php/scan/article/view/609>
- Vinarti, R. A., & Anggraeni, W. (2015). Identifikasi Faktor Prediksi Diagnosis Tingkat Keganasan Kanker Payudara Metode Stepwise Binary Logistic Regression. *Jurnal Informatika*, 12(2), 70–76. <https://doi.org/10.9744/informatika.12.2.70-76>