JURNAL INSY PRO Information System and Processing
Volume 10, Nomor 1, Mei 2025 P–ISSN 2597-419X, E-ISSN 2579-468X
Ojs : http://journal.uinalauddin.ac.id/index.php/insypro
Email : insypro@uin-alauddin.ac.id

ANALISIS DETEKSI DINI PENYAKIT JANTUNG DENGAN METODE *ENSEMBLE LEARNING* PADA DATA PASIEN

Early Detection Analysis of Heart Disease Using Ensemble Learning Methods on Patient Data

Fahrim Irhamna Rachman[1)], Rizki Yusliana Bakti[2)] , Titin Wahyuni[3)], Rizka Adrianingsih[4)]

[1,2,3,4]Program Studi Informatika Universitas Muhammadiyah Makassar
[1,2,3,4]Jl. Sultan Alauddin No 259, Kota Makassar, Sulawesi Selatan 90123

E-mail: fachrim141020@unismuh.ac.id[1)], Rizkiyusliana@unismuh.ac.id[2)], Titinwahyuni@unismuh.ac.id[3)], 105841108520@student.unismuh.ac.id[4)]

Abstrak – Penyakit jantung merupakan salah satu penyebab utama kematian yang memerlukan deteksi dini agar dapat dilakukan penanganan yang cepat dan tepat. Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit jantung menggunakan metode *ensemble learning*, khususnya teknik *Adaptive Boosting* (AdaBoost). Metode ini menggabungkan beberapa model lemah untuk meningkatkan akurasi klasifikasi penyakit jantung pada data pasien. Hasil penelitian menunjukkan bahwa penerapan teknik *ensemble learning* dengan metode AdaBoost menghasilkan model dengan akurasi yang sangat tinggi, terutama setelah menambahkan fitur demografis seperti jenis kelamin dan usia. Akurasi model meningkat dari 93,75% menjadi 100% dengan *precision*, *recall*, dan *f1-score* mencapai nilai 1,00 untuk kedua kelas. Dengan hasil yang sangat baik ini, metode AdaBoost terbukti efektif dalam mendeteksi penyakit jantung pada tahap awal, memberikan peluang untuk tindakan medis yang lebih tepat waktu dan efektif. Penelitian ini diharapkan memberikan kontribusi signifikan dalam pengembangan teknologi deteksi dini penyakit jantung serta meningkatkan kualitas hidup pasien melalui diagnosis yang lebih akurat.

Kata Kunci: Penyakit jantung, Deteksi dini, *Ensemble learning*, AdaBoost, *Machine learning*

Abstract *– Heart disease is one of the leading causes of death, requiring early detection for prompt and accurate treatment. This study aims to develop a heart disease prediction model using ensemble learning methods, specifically the Adaptive Boosting (AdaBoost) technique. This method combines several weak models to improve the accuracy of heart disease classification based on patient data. The results show that applying the ensemble learning technique with the AdaBoost method produces a highly accurate model, especially after adding demographic features such as gender and age. The model's accuracy increased from 93.75% to 100%, with precision, recall, and F1-score reaching a perfect score of 1.00 for both classes. With these excellent results, the AdaBoost method has proven to be effective in detecting heart disease at an early stage, providing opportunities for more timely and effective medical interventions. This research is expected to make a significant contribution to the development of early heart disease detection technology and improve patient quality of life through more accurate diagnoses.*

Keywords: *Heart disease, Early detection, Ensemble learning, AdaBoost, Machine learning*

PENDAHULUAN

Penyakit Jantung telah menjadi masalah kesehatan yang mendesak di berbagai negara, dengan dampak yang meluas dan mempengaruhi kehidupan jutaan orang setiap tahunnya. Penyakit Jantung merupakan salah satu kondisi yang paling umum terjadi di masyarakat yang dapat terjadi pada siapa saja tanpa memandang usia, jenis kelamin atau gaya hidup (Utomo & Mesran, 2020). Karena sifatnya yang kronis dan seringkali berkembang, penyakit jantung membutuhkan perawatan jangka panjang. Hal ini deteksi dini menjadi faktor penting dalam menangani penyakit jantung, diagnosis yang cepat dan tepat juga dapat memungkinkan untuk penanganan yang lebih efektif. Oleh Karena itu, penting untuk mengembangkan metode deteksi dini yang dapat mengenali gejala awal penyakit jantung, dengan memperhatikan hubungan antara hipertensi dan peningkatan kadar glukosa, kolesterol, kreatinin, ureum, serta aktivitas enzim SGOT dalam darah (Karwiti *et al*., 2023).

Diagnosis yang cepat dan akurat diperlukan untuk tindakan pencegahan dan pengobatan dini guna mengurangi risiko komplikasi serius. Pengembangan metode deteksi dini penyakit jantung sangat penting, meski analisis data medis umum telah membantu. Tantangan dalam meningkatkan akurasi prediksi penyakit jantung tetap ada, sehingga penggunaan metode *Ensemble Learning* menjadi relevan. Metode ini menggabungkan beberapa model dalam *machine learning* untuk meningkatkan kinerja. Salah satu teknik yang sering digunakan adalah *Adaptive Boost* yang merupakan teknik *boosting* dengan cara menggunakan metode secara besama-sama dalam pembelajaran mesin untuk meningkatkan akurasi (Andhika *et al*., 2023).

Melalui penelitian ini, dikembangkan model prediksi penyakit jantung berbasis AdaBoost pada data pasien. Dengan menggabungkan beberapa fitur klinis seperti usia, jenis kelamin, serta hasil laboratorium, dan pemeriksaan kesehatan model ini diharapkan mampu memberikan akurasi yang lebih tinggi dalam mendeteksi penyakit jantung pada tahap awal, sehingga dapat meningkatkan kualitas keputusan medis dan intervensi yang tepat waktu.

Dalam beberapa tahun terakhir, pendekatan *machine learning*, khususnya *ensemble learning*, telah terbukti mampu meningkatkan performa model prediksi di berbagai bidang, termasuk kesehatan. Salah satu metode *ensemble* yang sering digunakan adalah *Adaptive Boosting* (AdaBoost), yang dikenal efektif dalam menggabungkan beberapa model prediksi lemah menjadi model yang lebih kuat. Dalam upaya meningkatkan kinerja, setiap iterasi latihan dengan memberikan bobot pada data dan bobot tersebut digunakan untuk melatih klasifikasi yang berbeda (Novianti *et al*., 2022).

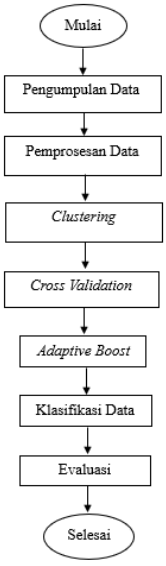
Penelitian sebelumnya oleh Al-Jamimi (2024) menunjukkan bahwa penggunaan *ensemble learning*, khususnya kombinasi XGBoost dan Bayesian, mampu meningkatkan akurasi prediksi penyakit kronis, termasuk penyakit jantung. Selain itu, Aziz *et al.* (2023) menemukan bahwa penerapan metode AdaBoost secara signifikan meningkatkan kinerja klasifikasi penyakit jantung dibandingkan dengan model dasar.

Dengan demikian seperti yang telah dijelaskan, Penelitian ini bertujuan untuk mengembangkan model prediksi deteksi dini penyakit jantung menggunakan metode *Ensemble Learning*, khususnya dengan algoritma *Adaptive Boost* (AdaBoost), yang menggabungkan beberapa *classifier* untuk meningkatkan performa prediksi. Algoritma ini mampu mengatasi tantangan akurasi dengan menggabungkan model-model yang lebih sederhana (*weak learners*) menjadi model yang lebih kuat dan handal dalam mendeteksi dini penyakit jantung.

METODOLOGI PENELITIAN

Model prediksi penyakit jantung menggunakan algoritma *Adaptive Boosting* (AdaBoost), salah satu teknik *ensemble learning*. Tujuannya adalah untuk meningkatkan akurasi dalam mendeteksi risiko penyakit jantung pada data pasien rumah sakit. Yang dimana *ensemble Learning* merupakan teknik yang menggabungkan beberapa model dasar *Machine Learning* (baik homogen maupun heterogen) untuk meningkatkan kualitas prediksi dengan mengurangi *noise* atau kesalahan antara data yang diamati dan data yang diprediksi (Dachi, 2023). Adaboost merupakan tipikal algortima pembelajaran *ensemble*, hasil yang didapatkan memiliki tingkat akurasi yang kuat.

Adapun tahapan dalam perancangan sistem proses pendeteksian dini penyakit jantung sebagai berikut:



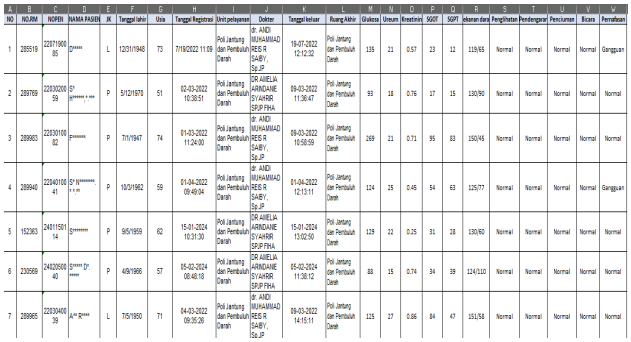
Gambar 1 Perancangan Sistem

Pada gambar 1. Proses dimulai dengan pengumpulan data pasien penyakit jantung, diikuti dengan pemrosesan data untuk pembersihan. *Clustering* dilakukan untuk membagi data ke dalam kelompok serupa. Setelah itu, *Cross Validation* membagi dataset untuk pengujian dan pelatihan model. Metode *Adaptive Boost* digunakan untuk

JURNAL INSY PRO Information System and Processing
Volume 10, Nomor 1, Mei 2025 P–ISSN 2597-419X, E-ISSN 2579-468X
Ojs : http://journal.uinalauddin.ac.id/index.php/insypro
Email : insypro@uin-alauddin.ac.id

meningkatkan akurasi prediksi dengan menggabungkan model lemah secara adaptif. Setelah model terlatih, data diklasifikasikan untuk mendeteksi kemungkinan penyakit jantung. Setelah evaluasi, sistem siap digunakan untuk deteksi dini penyakit jantung.

Data yang diperoleh dari Rumah Sakit Umum Daerah (RSUD) Haji Makassar yang berisi informasi rekam medis pasien penyakit jantung. Data yang diambil dari ruang rekam medis mencakup 640 pasien dengan diagnosa penyakit jantung dalam periode pemeriksaan dari 1 Januari 2021 hingga 15 Juli 2024. Informasi yang diperoleh meliputi berbagai kategori, yaitu data demografis, data registrasi dan pelayanan, serta data laboratorium. Proses pengumpulan data laboratorium, data fisiologis serta hasil pemeriksaan kesehatan seperti Tekanan Darah diambil dari berkas rekam medis pasien selama proses pemeriksaan.



Gambar 2 Dataset

Preprocessing data meliputi pembersihan dan transformasi. Pada pembersihan, data duplikat pasien dihapus dan kesalahan penulisan, seperti nomor rekam medis atau nama, diperbaiki untuk menghindari inkonsistensi. Dalam transformasi, data Jenis Kelamin diubah menjadi angka biner ("0" untuk laki-laki dan "1" untuk perempuan), dan kolom Tekanan Darah dipisah menjadi dua kolom: *Systolic* dan *Diastolic*, untuk memudahkan pemrosesan oleh model.

Tabel 1 Transformasi Data

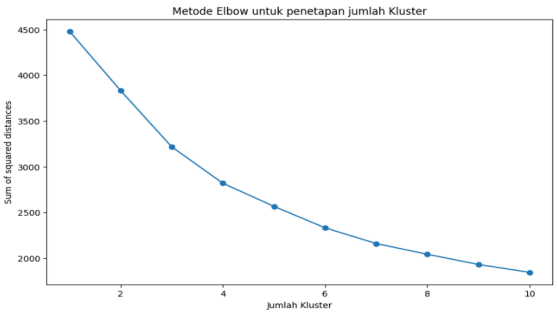
Nama	JK	Tekanan Darah	*Systolic*	*Diastolic*
D*****	0	119/65	119	65
S* H******	1	130/90	130	90
F*******	1	74	269	21
S* N*******	1	59	124	25
S********	1	62	129	22
…	…	…	…	…

Proses *clustering* dengan mengelompokkan kumpulan data ke dalam kelompok-kelompok yang memiliki kesamaan ke dalam *cluster* yang sama dan yang tidak sama ke *cluster* yang sama (Sekar Setyaningtyas *et al*., 2022). *Clustering* dilakukan dengan K-Means menggunakan metode *elbow* untuk menentukan jumlah kluster optimal berdasarkan grafik *Sum of Square Error* (SSE). Algoritma K-Means menentukan jumlah klaster, inisialisasi, dan jarak antar objek, baik secara manual maupun dengan bantuan *software* (Lobo *et al*., 2022).

Pengujian dilakukan dengan *Cross-Validation*, di mana dataset dibagi menjadi subset data latih dan uji untuk memastikan model diuji pada data yang beragam. Setelah *clustering*, performa model diukur menggunakan metrik akurasi, presisi, recall, dan F1-score. Kemudian, teknik *Adaptive Boost* diterapkan untuk meningkatkan performa. Akhirnya, model dievaluasi dengan dataset uji baru untuk memastikan prediksi yang akurat.

HASIL DAN PEMBAHASAN

Pada proses ini dengan membuang kolom yang tidak relevan, seperti No. Rekam Medis, No. Pendaftaran, Tanggal Lahir, Tanggal Registrasi, Unit Pelayanan, Dokter, Tanggal Keluar, Ruang Akhir, serta data pemeriksaan fisiologi seperti Penglihatan, Pendengaran, Penciuman, Bicara, dan Pernafasan, lalu menggunakan kolom Jenis Kelamin, Usia, Glukosa, Ureum, Kreatinin, SGOT, SGPT, *Systolic*, dan *Diastolic* untuk *clustering*. Penentuan jumlah *cluster* yang optimal diidentifikasi dengan mempertimbangkan perbandingan perhitungan SSE pada setiap nilai *cluster*, peningkatan jumlah *cluster* akan membentuk siku, sehingga semakin besar nilai k, nilai SSE akan semakin kecil (Syahfitri *et al*., 2023).



Gambar 3 Grafik Metode *Elbow*

Pada gambar 3. sumbu X menunjukkan jumlah kluster 1 hingga 10, sementara sumbu Y

menampilkan nilai SSE (*sum of squared distances*). Penurunan tajam SSE terjadi saat kluster meningkat dari 1 ke 2, menunjukkan pengurangan signifikan dalam variasi data. Namun, laju penurunan melambat setelahnya. Titik di mana penurunan SSE mulai melambat secara signifikan disebut "*elbow*," mengindikasikan jumlah kluster optimal kemungkinan di sekitar 3, karena setelah itu, penurunan SSE menjadi kurang signifikan.

Tabel 2 Hasil Cluster

No	Nama Pasien	…	*Systolic*	*Diastolic*	Kluster
1	D*****	…	119	65	0
2	S* H******	…	130	90	1
3	F*******	…	269	21	1
4	S* N*******	…	124	25	1
5	S********	…	129	22	1
…	…	…	…	…	…

Berdasarkan tabel 2. Data dari beberapa pasien yang mencakup hasil pengukuran berbagai parameter kesehatan seperti glukosa darah, ureum, kreatinin, enzim hati (SGOT dan SGPT), serta tekanan darah (*Systolic* dan *Diastolic*). Selain itu, tabel ini juga mengelompokkan pasien ke dalam kluster berdasarkan karakteristik kesehatan mereka. Informasi ini memberikan pandangan awal tentang kondisi metabolik, fungsi ginjal, dan risiko kardiovaskular setiap pasien, yang sangat penting dalam mendukung diagnosis dini dan perencanaan pengobatan yang tepat.

Cluster	NO	NO.RM	NOPEN	JK	Usia	Glukosa
0	313.209524	286328.971429	2.296849e+09	0.0	57.752381	149.911111
1	328.340557	280909.095975	2.301406e+09	1.0	57.173375	151.811146
2	202.500000	269474.500000	2.208515e+09	0.5	59.500000	102.000000

Cluster	Ureum	Kreatinin	SGOT	SGPT	Systolic	Diastolic
0	33.269841	1.086317	30.92381	30.615873	148.669841	84.726984
1	30.600619	1.041053	30.22291	29.631579	145.244582	82.071207
2	83.000000	2.050000	654.00000	163.000000	118.000000	73.000000

Cluster	Kluster	Cluster
0	0.0	0.0
1	1.0	1.0
2	2.0	2.0

Gambar 4 Centroids

Pada gambar 4. Setiap *cluster* dalam analisis ini menunjukkan karakteristik unik berdasarkan nilai rata-rata dari variabel-variabel yang dianalisis. *Cluster* 0 mewakili pasien dengan kondisi kesehatan yang lebih stabil, ditandai oleh kadar ureum dan kreatinin yang rendah, yang menunjukkan fungsi ginjal yang baik dan metabolisme yang seimbang. *Cluster* 1 menunjukkan pasien dengan tekanan darah sistolik dan kadar glukosa yang tinggi, mengindikasikan potensi risiko hipertensi dan diabetes yang memerlukan perhatian medis lebih lanjut. Sementara itu, *cluster* 2 memiliki nilai SGOT dan SGPT yang sangat tinggi, yang kemungkinan menunjukkan adanya kerusakan hati atau gangguan metabolisme. Kondisi ini juga dapat berdampak pada kesehatan jantung, karena gangguan hati sering kali berhubungan dengan peningkatan risiko kardiovaskular dan memerlukan evaluasi serta penanganan yang mendalam untuk memastikan kesehatan jantung yang optimal.

Kategori risiko ini kemungkinan didasarkan pada kombinasi berbagai parameter kesehatan seperti glukosa darah, tekanan darah, dan fungsi organ penting lainnya. Data ini menunjukkan bahwa sebagian besar pasien memiliki risiko penyakit yang relatif rendah, namun ada sekelompok pasien yang membutuhkan perhatian lebih intensif karena memiliki risiko yang tinggi hingga sangat tinggi terhadap komplikasi kesehatan. Informasi ini penting dalam perencanaan intervensi medis dan manajemen penyakit untuk mengurangi risiko komplikasi lebih lanjut.

Proses evaluasi model dimulai dengan membagi dataset, di mana 80% digunakan untuk pelatihan dan 20% untuk pengujian. Model AdaBoostClassifier, bagian dari *scikit-learn*, diterapkan sebagai teknik *boosting* yang menggabungkan 100 pohon keputusan lemah untuk membentuk model yang kuat. Algoritma kuat ini menggunakan *Decision Tree* sebagai model dasar dan menerapkan parameter SAMME untuk menangani prediksi multikelas. Pelatihan dilakukan dengan data pelatihan, di mana model belajar mengenali pola dan melakukan prediksi kelas target berdasarkan kombinasi dari model-model pohon keputusan yang diperkuat.

Setelah melatih model AdaBoostClassifier, langkah berikutnya adalah membuat prediksi pada set pengujian menggunakan metode *predict*. Hasil prediksi disimpan, kemudian, untuk mengevaluasi kinerja model, kita menghitung akurasi dengan membandingkan label prediksi dengan label sebenarnya. Akurasi ini menunjukkan seberapa

sering model memberikan prediksi yang benar. Akurasi dan laporan klasifikasi kemudian dicetak untuk menilai efektivitas model dalam mengklasifikasikan data pengujian, baik dengan mempertimbangkan fitur jenis kelamin dan usia maupun tanpa mempertimbangkan fitur tersebut.

Akurasi: 0.9375

Laporan Klasifikasi:

Tabel 3 Hasil Evaluasi Tanpa JK dan Usia

	precision	recall	f1-score	support
0	0.94	0.96	0.95	81
1	0.93	0.89	0.91	47
accuracy			0.94	128
macro avg	0.94	0.93	0.93	128
weighted avg	0.94	0.94	0.94	128

Akurasi: 1.0

Laporan Klasifikasi:

Tabel 4 Hasil Evaluasi dengan JK dan Usia

	precision	recall	f1-score	support
0	1.00	1.00	1.00	72
1	1.00	1.00	1.00	56
accuracy			1.00	128
macro avg	1.00	1.00	1.00	128
weighted avg	1.00	1.00	1.00	128

Penambahan fitur jenis kelamin dan usia meningkatkan kinerja model. Pada tabel pertama, model hanya mencapai akurasi 93.75% dengan beberapa kesalahan. Setelah menambahkan jenis kelamin dan usia, akurasi meningkat menjadi 100%, dengan semua metrik evaluasi mencapai nilai sempurna, menunjukkan peningkatan signifikan dalam prediksi. Penambahan fitur ini memungkinkan model untuk lebih baik mengenali pola-pola yang sebelumnya tidak terdeteksi, mengarah pada perbaikan yang signifikan dalam kemampuan klasifikasi.

Perbedaan kedua tabel menunjukkan peningkatan kinerja model setelah menambahkan fitur jenis kelamin dan usia. Pada tabel pertama, model hanya menghasilkan akurasi 93.75% dengan beberapa kesalahan klasifikasi. Setelah penambahan fitur jenis kelamin dan usia, akurasi meningkat menjadi 100% dengan semua metrik evaluasi mencapai nilai sempurna (1.00). Ini menunjukkan bahwa variabel demografis ini sangat meningkatkan akurasi model, membantu dalam membuat prediksi yang lebih tepat.

KESIMPULAN

Hasil penelitian menggunakan metode *ensemble learning*, terutama AdaBoost, menunjukkan kinerja sangat baik dalam deteksi dini penyakit jantung. Model awal dengan akurasi 93.75% meningkat menjadi 100% setelah menambahkan fitur demografis, dengan *precision*, *recall*, dan *f1-score* mencapai 1.00 untuk kedua kelas. Ini menunjukkan potensi besar AdaBoost dalam meningkatkan deteksi penyakit jantung dan kualitas hidup pasien. Untuk penelitian lebih lanjut, disarankan memperluas studi ke berbagai jenis penyakit jantung dan penyakit kronis lainnya dengan dataset yang lebih beragam selain dari parameter Jenis Kelamin, Usia, Glukosa, Ureum, Kreatinin, SGOT, SGPT dan Tekanan Darah. Pengujian metode *Machine Learning* lain serta implementasi klinis dan uji lapangan di berbagai fasilitas kesehatan juga penting untuk menguji penerapan nyata dan sistem pendukung keputusan medis.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada rekan-rekan yang telah banyak membantu dalam menyelesaikan penyusunan penelitian ini, serta kepada semua pihak yang telah memberikan dukungan dan kontribusi berharga selama proses penelitian.

DAFTAR PUSTAKA

Al-Jamimi, H. A. (2024). S*ynergistic Feature Engineering and Ensemble Learning for Early Chronic Disease Prediction. IEEE Access*, *12*(March), 62215–62233. https://doi.org/10.1109/ACCESS.2024.3395512

Andhika, L. D., Cahyani, D. R., Saputra, D., & Herawati, T. (2023). *Analisis Sentimen Kosumen KFC Berdasarkan Pendekatan Naive Bayes dan Ada Boost Berbasis Data Twitter*. *3*(1), 55–61.

Aziz, M. I., Fanani, A. Z., & Affandy, A. (2023). *Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis Decision Tree. Jurnal Media Informatika Budidarma*, *7*(1), 1–12. https://doi.org/10.30865/mib.v7i1.5169

Jan Melvin Ayu Soraya Dachi. (2023). *Analisis Perbandingan Algoritma XGBoost dan Algoritma*

Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam, *2*(2), 87–103. https://doi.org/10.55606/jurrimipa.v2i2.1470

Karwiti, W., Rezekiyah, S., & Lestari, W. S. (2023). *Blood Chemistry Profile as Early Detection of Degenerative Diseases in the Productive Age Group*. *9*(November 2022), 494–503.

Lobo, M. A. A., Prasetyo, S. Y. J., & Hartomo, K. D. (2022). *Pemetaan Karakteristik Sekolah Sasaran Promosi pada UNKRISWINA SUMBA menggunakan K-Means. Jurnal Media Informatika Budidarma*, *6*(4), 1842. https://doi.org/10.30865/mib.v6i4.4464

Novianti, N., Zarlis, M., & Sihombing, P. (2022). *Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes. Jurnal Media Informatika Budidarma*, *6*(2), 1200. https://doi.org/10.30865/mib.v6i2.4017

Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). *Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, *10*(2), 52–61. https://doi.org/10.21063/jtif.2022.v10.2.52-61

Syahfitri, N., Budianita, E., Nazir, A., & Afrianty, I. (2023). *Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid. KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(3), 1668–1675. https://doi.org/10.30865/klik.v4i3.1525

Utomo, D. P., & Mesran, M. (2020). *Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. Jurnal Media Informatika Budidarma*, *4*(2), 437. https://doi.org/10.30865/mib.v4i2.2080