

Penerapan Algoritma C4.5 untuk Memprediksi Status Penjualan Laptop dan PC pada Toko Elektronik

Nurul Hikmah^{*1}, Jainal Abidin², Alwendi³

^{*1}Mahasiswa Ilmu Komputer, Fakultas Teknik, Universitas Graha Nusantara Padangsidempuan, Indonesia

^{*2,3}Dosen Ilmu Komputer, Fakultas Teknik, Universitas Graha Nusantara Padangsidempuan, Indonesia

Email : *1nurulhikmah@gmail.com , *2jainalabidin@gmail.com , *3alwendi60@gmail.com

Abstract

This study aims to apply the C4.5 algorithm to predict the sales status of laptops and PCs in electronics stores to support decision-making in stock management and web-based marketing strategies. The problem faced is the unavailability of a web-based prediction system capable of accurately classifying products into best-selling and non-best-selling categories based on certain attributes. The method used is a data mining technique with a classification approach using the C4.5 algorithm implemented in a web-based system. The research stages include collecting sales data, determining attributes consisting of product type, brand, and price, and forming a decision tree based on the calculation of gain values to generate classification rules. The results show that the product type attribute is the dominant factor in determining sales status. Based on the resulting decision tree, the rule obtained is that products with the PC type are categorized as best-selling, while products with certain types of laptops and brands are categorized as not best-selling, laptops with other specific brands are categorized as best-selling, and laptops with price categories B or C are categorized as not best-selling. The web-based system developed is capable of displaying prediction results quickly and easily understood, so it can be used as a basis for decision-making in electronics stores.

Keywords: Data Mining, C4.5 Algorithm, Sales Prediction, Web Based.

Abstrak

Penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam memprediksi status penjualan laptop dan PC pada toko elektronik guna mendukung pengambilan keputusan dalam pengelolaan stok dan strategi pemasaran berbasis web. Permasalahan yang dihadapi adalah belum tersedianya sistem prediksi berbasis web yang mampu mengklasifikasikan produk ke dalam kategori laris dan tidak laris secara akurat berdasarkan atribut tertentu. Metode yang digunakan adalah teknik data mining dengan pendekatan klasifikasi menggunakan algoritma C4.5 yang diimplementasikan dalam sebuah sistem berbasis web. Tahapan penelitian meliputi pengumpulan data penjualan, penentuan atribut yang terdiri dari jenis produk, merek, dan harga, serta pembentukan pohon keputusan berdasarkan perhitungan nilai gain untuk menghasilkan aturan klasifikasi. Hasil penelitian menunjukkan bahwa atribut jenis produk menjadi faktor dominan dalam menentukan status penjualan. Berdasarkan pohon keputusan yang dihasilkan, diperoleh aturan bahwa produk dengan jenis PC dikategorikan laris, sedangkan produk dengan jenis laptop dan merek tertentu dikategorikan tidak laris, laptop dengan merek tertentu lainnya dikategorikan laris, serta laptop dengan kategori harga B atau C dikategorikan tidak laris. Sistem berbasis web yang dibangun mampu menampilkan hasil prediksi secara cepat dan mudah dipahami, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan pada toko elektronik.

Kata kunci: Data Mining, Algoritma C4.5, Prediksi Penjualan, Berbasis Web.

PENDAHULUAN

Dalam kegiatan operasionalnya, toko elektronik kerap mengalami kesulitan dalam memahami pola penjualan produk, khususnya pada kategori laptop dan PC. Kondisi ini menyebabkan keputusan terkait pengelolaan stok dan strategi pemasaran sering kali kurang tepat. Data penjualan yang dimiliki sebenarnya cukup banyak, namun belum dimanfaatkan secara optimal untuk menghasilkan informasi yang dapat digunakan dalam memprediksi kecenderungan penjualan di masa mendatang. Laptop dan PC merupakan produk dengan variasi yang beragam, baik dari segi merek, tahun produksi, maupun harga. Perbedaan karakteristik tersebut memiliki pengaruh terhadap preferensi dan keputusan pembelian konsumen (Anugrah A, 2024). Tanpa adanya proses analisis yang terstruktur, pihak toko akan mengalami kesulitan dalam menentukan faktor-faktor utama yang memengaruhi tingkat penjualan suatu produk. Akibatnya, pengambilan keputusan masih cenderung bersifat subjektif dan kurang didukung oleh data yang akurat (Alam, F. N., dkk, 2024). Permasalahan yang sering muncul adalah ketidakmampuan dalam memperkirakan status penjualan produk secara tepat. Hal ini berpotensi menimbulkan ketidakseimbangan dalam pengelolaan persediaan, seperti terjadinya penumpukan barang yang kurang diminati atau kekurangan stok pada produk yang memiliki permintaan tinggi. Di sisi lain, sebagian besar toko elektronik telah memiliki data penjualan yang tersimpan secara digital, namun belum dimanfaatkan secara maksimal untuk menghasilkan

informasi yang bernilai. Pemanfaatan teknik data mining dapat menjadi solusi dalam mengatasi permasalahan tersebut.

Melalui pendekatan ini, data penjualan dapat diolah untuk menemukan pola dan hubungan yang tersembunyi sehingga menghasilkan informasi yang berguna dalam pengambilan keputusan. Salah satu metode yang dapat digunakan adalah algoritma C4.5, yaitu metode klasifikasi yang mampu membentuk pohon keputusan berdasarkan atribut-atribut dalam data (Hofsah, T. P., & Anggoro, T., 2024). Penerapan algoritma C4.5 memungkinkan pengolahan data penjualan untuk menghasilkan model klasifikasi yang dapat memprediksi apakah suatu produk termasuk dalam kategori laris atau tidak laris. Model ini dibangun menggunakan beberapa atribut, seperti jenis produk, merek, tahun produksi, dan harga. Dengan adanya model tersebut, diharapkan pihak toko dapat menentukan strategi pengadaan dan pemasaran produk secara lebih efektif. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam mengklasifikasikan serta memprediksi status penjualan laptop dan PC pada toko elektronik, sehingga dapat mendukung pengambilan keputusan yang lebih tepat, terarah, dan berbasis data.

TINJAUAN PUSTAKA

Data mining merupakan proses pengolahan data untuk menemukan pola dan pengetahuan baru dari data berukuran besar dengan memanfaatkan teknik statistik, matematika, dan pembelajaran mesin. Secara umum, data mining

bertujuan mengekstraksi informasi bernilai berupa pola, hubungan, atau kecenderungan yang sebelumnya tidak diketahui dalam suatu kumpulan data (Syam S, dkk, 2024).

Berdasarkan fungsinya, metode dalam data mining dibagi menjadi dua kategori, yaitu predictive method dan descriptive method (Indriyawati, H., dan Winarti, T., 2021). Predictive method digunakan untuk memprediksi data di masa depan dengan teknik seperti klasifikasi dan regresi. Sedangkan descriptive method bertujuan mengidentifikasi pola atau karakteristik data tanpa melakukan prediksi, seperti clustering, association rule, dan sequential pattern mining (Jaya, I. K., Yohanna, dkk, 2023). Algoritma C4.5 merupakan salah satu metode klasifikasi dalam data mining yang menggunakan pendekatan pohon keputusan (decision tree). Algoritma ini dikembangkan sebagai penyempurnaan dari algoritma ID3 (Iterative Dichotomiser) dengan mekanisme pembentukan pohon secara top-down, bersifat rekursif, serta menggunakan strategi divide and conquer dalam proses pembagian data (Musa, Shantiony, dkk, 2024). Algoritma C4.5 memiliki kesamaan dengan ID3, namun dilengkapi dengan beberapa keunggulan, di antaranya mampu menangani atribut bertipe diskrit maupun numerik, mengelola data yang memiliki nilai hilang (missing value), serta menerapkan proses pemangkasan pohon (pruning) untuk meningkatkan akurasi model. Selain itu, pemilihan atribut terbaik pada C4.5 didasarkan pada nilai gain ratio, bukan hanya information gain seperti pada ID3.

Dalam proses pruning, algoritma C4.5 menggunakan metode pessimistic pruning, yaitu pemangkasan cabang pohon berdasarkan estimasi tingkat kesalahan pada data pelatihan tanpa memerlukan data khusus untuk pengujian (prune set). Penentuan kelas akhir (default class) dilakukan dengan memilih kelas yang memiliki jumlah data pelatihan paling dominan pada aturan tertentu. Tujuan utama algoritma C4.5 adalah membagi dataset ke dalam beberapa subset berdasarkan atribut yang paling informatif.

Proses ini dilakukan dengan menghitung nilai entropy dan information gain. Entropy digunakan untuk mengukur tingkat ketidakpastian atau keragaman data dalam suatu dataset, sedangkan information gain menunjukkan besarnya pengurangan nilai entropy setelah data dibagi berdasarkan atribut tertentu. Atribut dengan nilai gain tertinggi akan dipilih sebagai node pada pohon keputusan. Tahapan dalam pembentukan pohon keputusan menggunakan algoritma C4.5 adalah sebagai berikut (Hakim, Putrawansyah, 2024): Menentukan atribut akar berdasarkan nilai gain tertinggi. Membentuk cabang dari setiap nilai pada atribut yang terpilih. Menentukan atribut berikutnya sebagai cabang berdasarkan nilai gain pada subset data. Mengulangi proses tersebut hingga seluruh data terklasifikasi atau memenuhi kondisi tertentu. Dalam menentukan atribut terbaik sebagai akar, diperlukan perhitungan nilai gain yang diawali dengan menghitung nilai entropy. Perhitungan tersebut dirumuskan dalam persamaan berikut.

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

Keterangan :

S : himpunan kasus

A : atribut

n : jumlah partisi S

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

Sementara itu, penghitungan nilai *gain* dapat dilihat pada persamaan 2 berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke-i

$|S|$: jumlah kasus dalam S

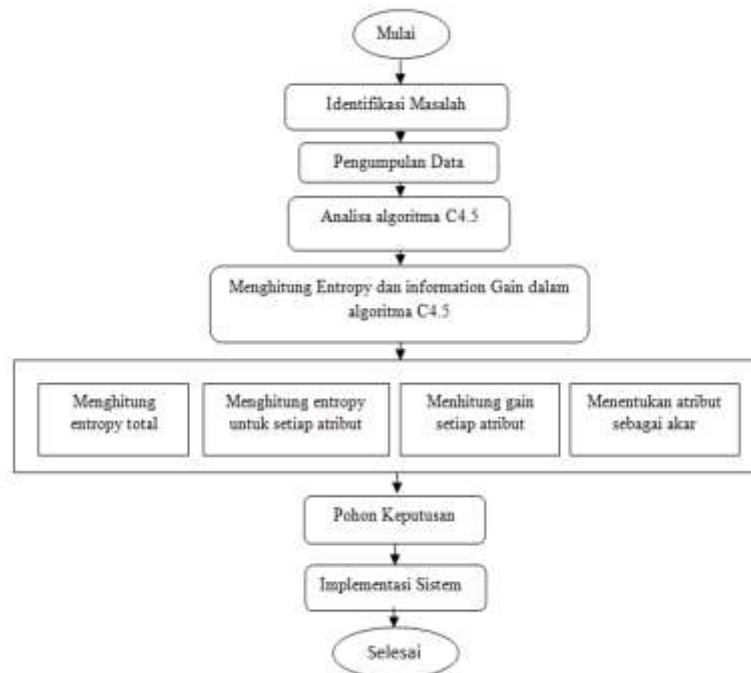
Pohon keputusan (decision tree) merupakan model berbentuk struktur hierarki yang digunakan dalam proses klasifikasi pada data mining. Struktur ini terdiri dari node, cabang, dan daun, di mana setiap node internal merepresentasikan atribut yang diuji, cabang menunjukkan hasil pengujian, dan daun (leaf) menyatakan kelas atau

keputusan akhir (Zacksavira, S., dkk., 2025). Pada algoritma C4.5, pembentukan pohon keputusan didasarkan pada perhitungan information gain dari setiap atribut. Atribut dengan nilai gain tertinggi akan dipilih sebagai akar karena memiliki kemampuan terbaik dalam memisahkan data ke dalam kelas yang berbeda. Proses pembentukan pohon dilakukan secara

berulang (rekursif) hingga seluruh data terklasifikasi atau tidak terdapat atribut lain

yang dapat digunakan (Agustin, R., & Defit, S., 2024).

METODOLOGI PENELITIAN

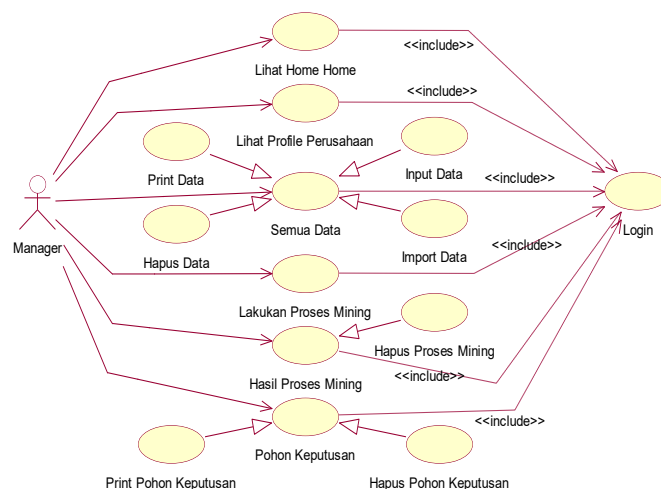


Gambar 1. Diagram Alir Penelitian

ANALISA DAN PERANCANGAN

Analisis sistem merupakan proses penguraian suatu sistem informasi ke dalam komponen-komponen yang lebih kecil dengan tujuan untuk mengidentifikasi serta mengevaluasi permasalahan yang ada, sekaligus menentukan kebutuhan yang diperlukan sebagai dasar perbaikan sistem. Tahap ini menjadi langkah awal dalam perancangan dan pengembangan sistem, karena pada tahap inilah kinerja sistem dianalisis dan dievaluasi untuk mengetahui kekurangan yang perlu diperbaiki. Melalui analisis sistem, dilakukan identifikasi terhadap berbagai permasalahan serta penentuan kebutuhan yang harus dipenuhi dalam proses perancangan.

Oleh karena itu, pemahaman yang menyeluruh terhadap sistem sangat diperlukan sebelum melakukan analisis. Selain itu, proses analisis juga membutuhkan data yang relevan sebagai dasar dalam melakukan evaluasi dan perumusan kebutuhan sistem. Analisis data merupakan tahapan untuk mengkaji dan mengolah data yang dibutuhkan dalam perancangan sistem. Pada penelitian ini, data diperoleh melalui studi literatur yang berkaitan dengan topik penelitian. Informasi yang dikumpulkan digunakan sebagai dasar dalam penyusunan landasan teori, sehingga menghasilkan keterkaitan yang sistematis antara konsep-konsep yang digunakan dalam penelitian.



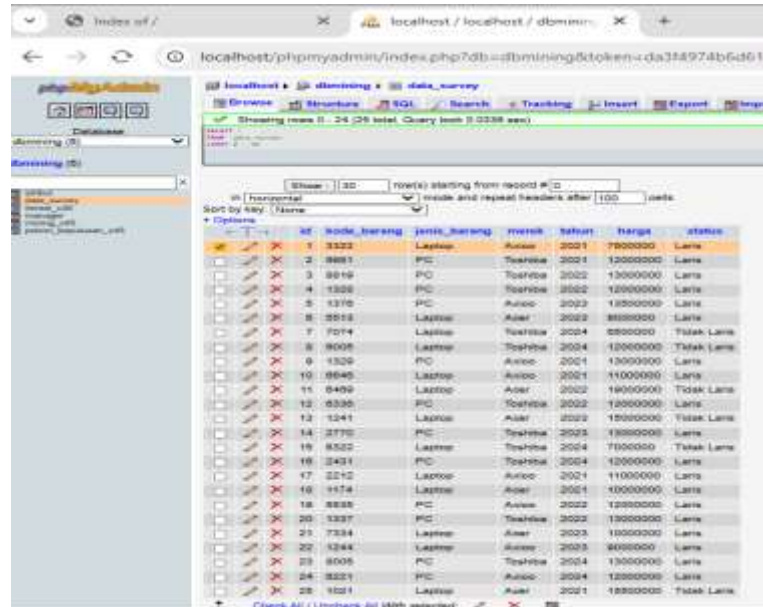
Gambar 2. Use Case Diagram

Gambar tersebut menunjukkan diagram use case yang menggambarkan hubungan interaksi antara aktor dan sistem yang dikembangkan. Aktor dalam diagram merepresentasikan pengguna yang berperan dalam menjalankan berbagai fungsi yang tersedia pada sistem. Setiap elemen berbentuk oval mencerminkan fitur atau proses yang dapat dilakukan, seperti pengolahan data, analisis, hingga menghasilkan informasi sebagai output. Sementara itu, garis penghubung antara aktor dan use case menunjukkan adanya interaksi, di mana pengguna dapat

mengakses dan menjalankan fungsi-fungsi tersebut sesuai kebutuhan. Diagram juga memperlihatkan alur proses yang berujung pada keluaran sistem, seperti laporan atau hasil prediksi. Hal ini menandakan bahwa sistem tidak hanya menerima input, tetapi juga mengolahnya menjadi informasi yang bermanfaat. Secara keseluruhan, diagram tersebut menggambarkan sistem yang interaktif, yang memungkinkan pengguna melakukan proses mulai dari input data hingga memperoleh hasil analisis sebagai dasar pengambilan keputusan.

NODE			Jml Kasus (S)	Tidak laris (S ₁)	Laris (S ₂)	Entropy	Gain
1.1.2	Jenis Laptop dan Merek Acer		6	3	3	1.000000	
	Harga						1
		A	3	0	3	0.000000	
		B	1	1	0	0.000000	
		C	2	2	0	0.000000	
	Tahun						1
		A	3	0	3	0.000000	
		B	3	3	0	0.000000	

Gambar 3. Perhitungan Algoritma C4.5



id	kode_barang	jenis_barang	merek	tahun	harga	status
1	3323	Laptop	Acer	2021	7800000	Laris
2	9881	PC	Toshiba	2021	12000000	Laris
3	9819	PC	Toshiba	2022	13000000	Laris
4	1328	PC	Toshiba	2022	12000000	Laris
5	1310	PC	Acer	2023	18000000	Laris
6	5515	Laptop	Acer	2022	8000000	Laris
7	7014	Laptop	Toshiba	2024	5500000	Tidak Laris
8	9005	Laptop	Toshiba	2024	12000000	Tidak Laris
9	1329	PC	Acer	2021	13000000	Laris
10	9840	Laptop	Acer	2021	11000000	Laris
11	5409	Laptop	Acer	2022	16000000	Tidak Laris
12	5335	PC	Toshiba	2022	12000000	Laris
13	1241	Laptop	Acer	2022	15000000	Tidak Laris
14	2770	PC	Toshiba	2023	13000000	Laris
15	8522	Laptop	Toshiba	2024	7000000	Tidak Laris
16	2431	PC	Toshiba	2024	12000000	Laris
17	2212	Laptop	Acer	2021	11000000	Laris
18	1174	Laptop	Acer	2021	10000000	Laris
19	8835	PC	Acer	2022	12000000	Laris
20	1337	PC	Toshiba	2022	13000000	Laris
21	7334	Laptop	Acer	2023	10000000	Laris
22	1244	Laptop	Acer	2023	6000000	Laris
23	8005	PC	Toshiba	2024	13000000	Laris
24	8321	PC	Acer	2024	12000000	Laris
25	1021	Laptop	Acer	2021	18000000	Tidak Laris

Gambar 4. Penyimpanan data dalam Database

PENUTUP

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan mengenai penerapan algoritma C4.5 untuk memprediksi status penjualan laptop dan PC pada toko elektronik, dapat disimpulkan bahwa sistem berbasis web yang dikembangkan mampu mengklasifikasikan data penjualan secara otomatis menjadi kategori laris dan tidak laris berdasarkan data historis dengan tingkat akurasi yang cukup baik. Keberadaan sistem ini memberikan kemudahan bagi pihak toko dalam melakukan analisis tanpa perlu perhitungan manual, sehingga hasil prediksi yang dihasilkan dapat dimanfaatkan dalam pengambilan keputusan terkait pengelolaan stok, penetapan harga, serta strategi promosi. Selain itu, hasil pengujian

menunjukkan bahwa algoritma C4.5 efektif digunakan dalam menganalisis data penjualan karena mampu mengidentifikasi pola-pola yang relevan, sehingga dapat meningkatkan efisiensi operasional dan meminimalkan risiko penumpukan barang.

Saran

Berdasarkan hasil penelitian, sistem yang dikembangkan masih memiliki beberapa keterbatasan sehingga memerlukan pengembangan lebih lanjut di masa mendatang. Oleh karena itu, disarankan agar sistem dilengkapi dengan fitur pencadangan data (backup) untuk menjaga keamanan data apabila terjadi kerusakan pada server. Selain itu, aplikasi dapat dikembangkan ke platform Android atau sistem berbasis mobile agar lebih

fleksibel dan mudah diakses oleh pengguna di berbagai perangkat. Dengan pengembangan tersebut, diharapkan sistem dapat digunakan secara lebih efektif dan efisien dalam mendukung aktivitas operasional toko.

DAFTAR PUSTAKA

- Anugrah A, A. (2024). Pengaruh Digital Marketing Dan Viral Marketing Terhadap Keputusan Pembelian (Studi Pada Chocolicious Indonesia Kota Parepare) (Doctoral dissertation, IAIN Parepare).
- Alam, F. N., Setyawan, P. A., Dwianto, S., Danuarta, A., Syah, R. A. O., & Kusumastuti, R. (2024, December). Prediksi Ketersediaan Stok Bulanan Kategori Susu di Toko Koperasi menggunakan Metode Decision Tree. In Prosiding Seminar Nasional Amikom Surakarta (Vol. 2, pp. 25-34).
- Hofsah, T. P., & Anggoro, T. (2024). Implementasi Data Mining Penjualan menggunakan Algoritma Apriori (Studi Kasus: Toko SRC Aska Musa). *Innovative: Journal Of Social Science Research*, 4(6), 1999-2014.
- Syam, S., Tokoro, Y., Judijanto, L., Garonga, M., Sinaga, F. M., Umar, N., ... & Sitanggang, A. T. (2024). *Data Mining: Teori dan Penerapannya dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia.
- Indriyawati, H., & Winarti, T. (2021). Pemodelan data mining pola kelayakan kemampuan lulusan dengan kebutuhan stakeholder menggunakan algoritma apriori. *JITSI: Jurnal Ilmiah Teknologi Sistem Informasi*, 2(3), 78-84.
- Jaya, I. K., Yohanna, M., Rumapea, H., Rumapea, Y. Y. P., & Juli, N. (2023). *Buku Monograf ALGORITMA C. 45 KELANCARAN PEMBAYARAN KREDIT*. MEGA PRESS NUSANTARA.
- Erdisna, E., Murniasih, I., Diantoro, K., Hardini, I. R., Rinaldo, R., Asruddin, A., ... & Adiya, M. H. (2025). *Data Warehouse*. CV. Gita Lentera.
- PRAMUDYA, M. I. A. (2022). Penerapan Market Basket Analysis Menggunakan Proses Knowledge Discovery in Database (Kdd) Sebagai Strategi Penjualan.... Repository. Radenintan. Ac. Id, 1-2.
- Wijaya, Y. (2024). Implementasi data mining dengan metode support vector machine pada aplikasi fundraising UKM LDK Syahid (Bachelor's thesis, Fakultas Sains Teknologi UIN Syarif Hidayatullah Jakarta).
- Musa, D. M., Sakti, D., Shantiony, K. A., Zega, S. K. P., Hamzah, S., Zega, Y. J., & Lubis, B. O. (2024). Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4. 5. *Jurnal Teknologi Informatika Dan Komputer*, 10(1), 168-182.

Hakim, R. X., Putrawansyah, F., & Syahri, R. (2024). Penerapan Algoritma C4. 5 Untuk Prediksi Anak Stunting Di Kota Pagar Alam. JATI (Jurnal Mahasiswa Teknik Informatika), 8(2), 2469-2478.

Zacksavira, S., Sebayang, C. A. B., Imanuel, J., Sibarani, G. B., & Pratama, A. (2025). Implementasi Algoritma Decision Tree untuk Deteksi Depresi Pada Pelajar. Innovative: Journal Of Social Science

Research, 5(3), 5915-5934.

Agustin, R., & Defit, S. (2024). Perbandingan Algoritma CART dan C. 4 5 Pada Citra Tandan Buah Sawit Untuk Mengetahui Tingkat Kematangan Dalam Penentuan Harga. Jurnal KomtekInfo, 263-273.

Zahrani, S. I. (2024). Sistem Informasi Toko Roti Berbasis Website (Doctoral dissertation, Politeknik Harapan Bersama).