

Data Mining Crisp-Dm C4.5 dalam Estimasi Penjualan Produk Pada PT. Sari Pati Abadi Berbasis android

Crisp-DM C4.5 Data Mining in Product Sales Estimation at PT. Sari Pati Abadi Based on Android

Rita Novita Sari¹, Mutiya Annisa²

^aProgram Studi Ilmu Komputer Universitas Islam Negeri Sumatera Utara Medan, Jl. Lap. Golf. No. 120 Kp. Tengah, Pancur Batu

^bProgram Studi Informatika Universitas Potensi Utama, Jl.K.L Yos Sudarso KM 6.5 Tj.Mulia, Medan, 20241, Indonesia

E-mail: ritanovitasari@uinsu.ac.id

E-mail : mutiyaannisa6@gmail.com

ABSTRAK

Perkembangan teknologi informasi mendorong perusahaan untuk memanfaatkan data penjualan sebagai dasar pengambilan keputusan yang lebih tepat. PT. Sari Pati Abadi menghadapi permasalahan dalam memprediksi tingkat penjualan produk karena pengolahan data yang masih terbatas dan belum dimanfaatkan secara optimal. Penelitian ini bertujuan untuk menerapkan metode CRISP-DM (Cross Industry Standard Process for Data Mining) dalam pengolahan data penjualan dengan menggunakan algoritma C4.5 sebagai model klasifikasi untuk estimasi penjualan. Data penjualan yang digunakan diolah melalui tahapan CRISP-DM, meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi. Hasil penelitian ini menghasilkan model klasifikasi yang mampu mengestimasi tingkat penjualan produk ke dalam kategori tinggi, sedang, dan rendah. Selanjutnya, model tersebut diimplementasikan ke dalam aplikasi berbasis Android agar memudahkan manajemen dalam memperoleh informasi estimasi penjualan secara cepat dan praktis. Dengan adanya sistem ini, PT. Sari Pati Abadi dapat memanfaatkan data historis penjualan secara optimal untuk mendukung perencanaan strategi bisnis, pengelolaan stok, serta distribusi produk yang lebih efektif.

Kata Kunci : Data Mining, CRISP-DM, C4.5, Estimasi Penjualan, Android

ABSTRACT

The rapid development of information technology encourages companies to utilize sales data as a basis for more accurate decision-making. PT. Sari Pati Abadi faces challenges in predicting product sales levels due to limited data processing and the lack of optimal utilization of sales records. This study aims to apply the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology in processing sales data, using the C4.5 algorithm as a classification model for sales estimation. The sales data were processed through the CRISP-DM stages, including business understanding, data understanding, data preparation, modeling, evaluation, and implementation. The results of this study produced a classification model capable of estimating product sales levels into high, medium, and low categories. Furthermore, the model was implemented into an Android-based application to provide management with fast and practical access to sales estimation information. With this system, PT. Sari Pati Abadi can optimally utilize historical sales data to support business strategy planning, stock management, and more effective product distribution.

Keywords : Data Mining, CRISP-DM, C4.5, Sales Estimation, Android

<i>Disubmit:</i>	<i>Info artikel :</i>	
	<i>Direview:</i>	<i>Diterima :</i>

Copyright © 2025 – senadimu. All rights reserved.

1. PENDAHULUAN

Banyak orang membicarakan tentang bisnis, keuntungan bisnis serta banyak nya jenis bisnis yang beraneka ragam yang tentu memiliki satu tujuan yaitu mendapatkan keuntungan secara finansial atau laba. Dengan begitu banyak jenis bisnis dan manfaat yang bisa di ambil dari bisnis tentu seringkali kita merasa ingin segera terjun ke dunia bisnis untuk mendapatkan keuntungan atau lebih spesifik yaitu penghasilan. Salah satu contoh bisnis adalah berjualan atau berdagang [1]

Banyaknya persaingan didunia usaha memaksa para pelaku usaha untuk menyusun trobosan baru yang dapat menjamin kelancaran usaha yang dijalani. Agar tidak kalah dengan pesaing baru, maka diperlukan prediksi penjualan dengan metode yang dapat digunakan untuk merancang sebuah strategi bisnis dalam meningkatkan penjualan. Diera dunia yang modern kebutuhan masyarakat merupakan hal yang sangat penting dan sangat dibutuhkan bagi setiap masyarakat [2]

PT. Sari Pati Abadi adalah salah satu perusahaan penjualan makanan ringan. Penjualan produk di PT. Sari Pati Abadi saat ini mengalami peningkatan yang sangat pesat, karena meningkatnya permintaan konsumen. Dengan meningkatnya permintaan produk, karyawan PT. Sari Pati Abadi mengalami kesulitan dalam mendata barang masuk, barang keluar, dan stok barang yang tersedia. Data transaksi penjualan disimpan dalam data warehouse swalayan. Namun, tumpukan data transaksi tersebut nyatanya belum dimanfaatkan secara efektif. Data yang begitu besar hanya dimanfaatkan sebagai bahan pembuatan laporan penjualan per bulan tanpa menggali informasi atau pengetahuan baru dari pola transaksi yang ada [3]

Belum adanya cara mengolah informasi yang efektif dari kumpulan data penjualan di tengah angka penjualan produk yang fluktuatif banyak kali menjadi sumber masalah baru bagi pihak perusahaan. Pihak perusahaan kerap kali mengalami kesulitan dalam memprediksi angka penjualan produk, terlebih lagi pihak perusahaan belum memiliki teknik khusus guna melakukan peramalan angka penjualan di masa depan. Hal ini mengakibatkan perusahaan kesulitan menentukan stok tiap barang secara efektif, sehingga yang terjadi adalah adanya penimbunan stok barang di gudang perusahaan ataupun habisnya stok barang yang paling diminati konsumen karena kebutuhan pasar tidak diimbangi persediaan barang.

Pada penelitian ini diambil permasalahan terhadap data penjualan produk yang mengalami kesulitan mendapatkan informasi akurat seperti penjualan produk Terlaris. Dengan ketersediaan data yang banyak membuat perusahaan mandapatkan kesulitan untuk menentukan penjualan produk terlaris jika dilakukan secara manual. Selain itu juga membutuhkan waktu yang lama serta tingkat ketelitian yang tinggi karena besar kemungkinan terjadinya kesalahan dalam pengerjaannya. PT. Sari Pati Abadi mengalami kesulitan mendapatkan informasi strategis seperti tingkat penjualan per periode atau produk terlaris. Ketersedian data yang banyak dan kebutuhan akan informasi atau pengetahuan sebagai pendukung pengambilan keputusan untuk membentuk solusi bisnis.

Penelitian ini menggunakan tahap tahap Crisp-DM (*cross industry standar process for data mining*) yang banyak digunakan dalam proses data mining. Sebelum memasuki bagian tahapan dari crisp-dm, terlebih dahulu dilakukan tahap awal yaitu pengumpulan data. Enam langkah dari teknik ini adalah pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran. Crisp-DM ini juga membatasi penggunaan metode ilmiah dalam *knowledge mining*. Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya agar dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu [4]

Penerapan yang sesuai untuk mengatasi masalah tersebut adalah dengan dikembangkannya konsep Algoritma C4.5 dapat dengan mudah menentukan barang-barang yang mempunyai kaitan yang lebih dekat hubungannya, sehingga mengetahui bagaimana cara menentukan klasifikasi penjualan berdasarkan data barang [5]. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field – field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data [6].

Berdasarkan penelitian yang dilakukan menghasilkan kesimpulan bahwa klasifikasi dengan akurasi sebesar 99% yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur perkiraan penjualan sepeda motor Honda terlaris [7]

Pada penelitian lain yang telah dilakukan menghasilkan kesimpulan bahwa proses klasifikasi konsumen menggunakan data mining yang telah dilakukan mendapatkan hasil yang membantu perusahaan Telekomunikasi dalam mengambil keputusan kedepannya agar customer dapat bertahan lebih lama dalam menggunakan layanan mereka [8]

Adapun penelitian lain dengan kesimpulan bahwa dengan menggunakan metode algoritma C.45, data mining ini dapat menghasilkan output yang sempurna dalam mengklasifikasi kepribadian siswa dalam studi kasus di SMP N 5 Jepara [9]

Penelitian lain yang dilakukan menghasilkan kesimpulan bahwa berdasarkan hasil pengujian dengan menggunakan 437 data diperoleh nilai *accuracy* sebesar 91,53%, nilai *precision* sebesar 27,77%, dan nilai *recall* sebesar 17,24% [10]. Hal ini menunjukkan bahwa sistem memiliki kinerja yang baik dan algoritma C4.5 dapat diterapkan pada kasus klasifikasi penentuan produk terlaris di toko XYZ Palembang [11]

Berdasarkan Penelitian Pemetaan produk yang tidak seimbang mengakibatkan penjualan yang disusun tidak merata sehingga menyebabkan adanya penumpukan stok barang sehingga melebihi kapasitasnya. Selain itu ketidakrataan penjualan bisa juga mengakibatkan tidak adanya barang yang terjual sehingga menurunkan profit Perusahaan [12]. Pada penelitian ini dengan metode bisnis CRISP-DM menggunakan model algoritma K-Means sebagai salah satu algoritma yang digunakan untuk mengklaster produk untuk mendapatkan model pengelompokan produk. *K-Means* banyak dipakai karena kemudahan dan kesederhanaan algoritmanya [13].

Berdasarkan Penelitian Dalam penelitian ini akan dibandingkan kinerja antara algoritma - algoritma klasifikasi yaitu k-NN, NB, dan CART untuk prediksi cacat *software* dengan pendekatan CRISP-DM. CRISP-DM merupakan model proses data mining dengan 6 tahapan yaitu: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* dalam menentukan perbandingan algoritma klasifikasi dalam memprediksi cacat *software*. *Software Matrix* yang digunakan pada penelitian ini adalah tujuh dataset dari NASA MDP [14].

Berdasarkan Penelitian Kelulusan mahasiswa dalam suatu mata kuliah dapat diprediksikan berdasarkan data masukkan yang menjadi parameternya. Banyak metode klasifikasi yang dapat digunakan dengan keunggulan dan kekurangannya masing-masing yang dapat digunakan untuk melakukan prediksi tersebut [15]. Dengan menggunakan metodologi data mining CRISP-DM, yaitu mengkomparasi beberapa klasifikasi dalam *supervised learning*, dapat diperoleh nilai terbaik yang berkaitan dengan akurasi dan *error*[16]. Klasifikasi yang diperbandingkan dengan menggunakan *software Orange Data Mining* tersebut adalah *Naive Bayes* (NB), *Neural Network* (NN), *Logistic Regression* (LR) dan *Support Vector Machine* (SVM). Objek penelitian dilakukan terhadap mahasiswa yang mengikuti mata kuliah PTI (Pengantar Teknologi Informasi) [17].

Hasil ini kemudian diuji menggunakan *confusion matrix*. Pengujian mendapatkan hasil 1949 True Positive, 102 False Negative, 527 False Positive, 2477 True Negative. Tingkat akurasi hasil yaitu sebesar 87.56%, Presisi 78.72%, Recal 95.03% dan Error rate sebesar 12.44%. Penelitian kedepannya bisa dibandingkan dengan menggunakan metode yang berbeda [18].

2. METODE

2.1. Metode Pengumpulan Data

Di dalam melakukan penelitian diperlukan beberapa cara untuk mengumpulkan data yang diperlukan dalam kegiatan penelitian ini. Adapun teknik dalam pengumpulan data adalah :

a) Pengamatan (*Observation*)

Pengamatan adalah metode pengumpulan data yang dilakukan dengan memperhatikan dan mencatat peristiwa atau perilaku secara langsung tanpa campur tangan atau pertanyaan kepada subjek. Peneliti melakukan observasi dengan mengamati secara sistematis situasi atau objek penelitian.

b) Wawancara (*Interview*)

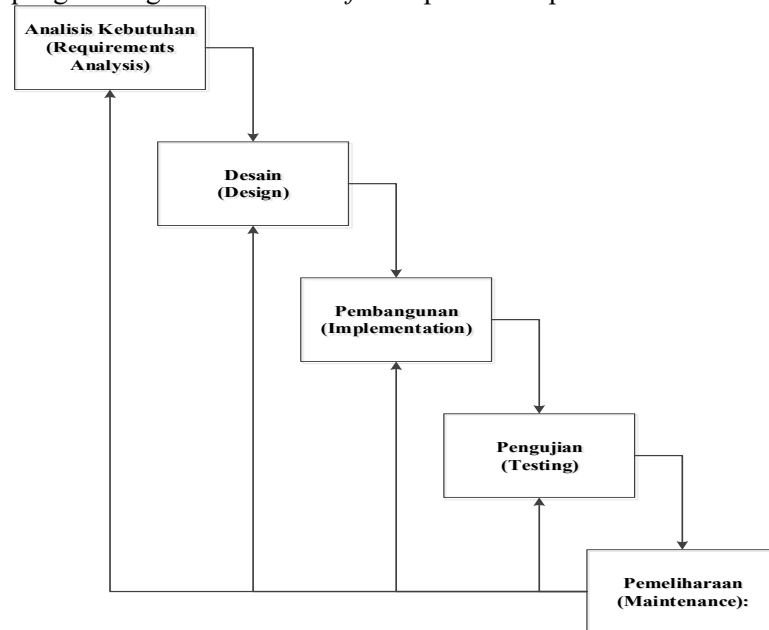
Wawancara adalah metode pengumpulan data yang melibatkan interaksi langsung antara peneliti dan responden. Peneliti mengajukan pertanyaan tertentu untuk mendapatkan informasi yang diinginkan. Wawancara dapat dilakukan dalam bentuk formal atau informal, peneliti melakukan wawancara langsung kepada bagian penjualan pada PT. Sari Pati Abadi.

c) Sampel (*Sampling*)

Sampling adalah proses pemilihan sebagian kecil dari populasi secara representatif, yang dijadikan objek penelitian, karena keterbatasan waktu dan sumber daya. Dalam penelitian, seringkali tidak memungkinkan atau tidak efisien untuk mengumpulkan data dari seluruh populasi. Oleh karena itu, peneliti memilih sampel yang merupakan subset dari populasi yang dianggap mewakili karakteristik umum dari keseluruhan populasi.

2.2. Metode Pengembangan Sistem

Pengembangan sistem dapat berupa menyusun suatu sistem yang baru dan menggantikan sistem yang lama secara keseluruhan atau memperbaiki sistem yang telah ada. Setiap tahap harus di selesaikan terlebih dahulu kemudian diteruskan ke tahap berikutnya untuk menghindari terjadinya pengulangan tahap. Metodologi pengembangan sistem *Waterfall* dapat dilihat pada Gambar 1 berikut:



Gambar 1. Diagram *Waterfall*

1. Analisis Kebutuhan (*Requirements Analysis*)

Tahapan ini merupakan langkah awal dimulai dengan memahami kebutuhan dan tujuan dari sudut pandang bisnis, mengubah pengetahuan menjadi definisi masalah dalam konteks data mining, kemudian merencanakan serta menyusun strategi untuk mencapai tujuan dalam proses data mining. Analisa masalah dalam penelitian ini adalah proses estimasi penjualan. Maka dengan demikian dirancang sebuah aplikasi dalam pengelompokan estimasi penjualan.

2. Desain (*Design*)

Desain dalam model *waterfall* dapat dihubungkan dengan pemilihan algoritma *C4.5*, konfigurasi parameter, dan persiapan data. Ini melibatkan pemilihan model dan pengaturan eksperimen. Tahapan ini untuk menentukan teknik, alat, dan algoritma data mining yang sesuai. Dalam penelitian ini menggunakan model *C4.5*.

3. Pembangunan (*Implementation*)

Implementasi pada model *waterfall* sebagian diwakili untuk melakukan pelatihan model menggunakan data yang telah dipersiapkan sebelumnya. Pada tahap ini, model machine learning dihasilkan berdasarkan desain yang telah dibuat. Tahapan ini dilakukan untuk proses design sistem

dengan menggunakan UML yang terdiri dari *usecase diagram*, *class diagram*, *sequence diagram* dan *activity diagram*, kemudian proses desain database dan desain *interface*.

4. Pengujian (*Testing*)

Digunakan untuk menguji model terhadap dataset uji yang terpisah untuk mengevaluasi kinerjanya. Tahapan ini menerapkan data mining untuk menentukan teknik, alat, dan algoritma data mining yang sesuai. Dalam penelitian ini menggunakan model C4.5.

5. Pemeliharaan (*Maintenance*)

Pemeliharaan dalam model eksperimen dapat mencakup peningkatan model berdasarkan performa, perubahan dalam dataset, atau perubahan kondisi lingkungan yang mempengaruhi kinerja model.

3. HASIL DAN PEMBAHASAN

3.1. Pembahasan

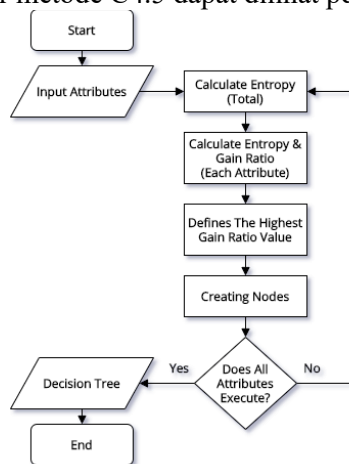
Studi kasus :

Pada penelitian ini digunakan algoritma C4.5 yang berfungsi untuk mendapatkan estimasi penjualan. Tujuan pada tahapan ini adalah menerapkan algoritma data mining (dalam hal ini C4.5 Decision Tree) untuk membangun model prediksi. Implementasi pada penelitian:

- Menerapkan algoritma C4.5 untuk membuat pohon keputusan.
- Proses melibatkan pemilihan atribut terbaik berdasarkan Gain Ratio.
- Pohon keputusan akan menghasilkan aturan (rules) untuk mengestimasi penjualan produk (tinggi, sedang, rendah).
- Membagi data menjadi training set dan testing set untuk mengukur performa model.

Adapun flowchart dari metode C4.5 adalah sebagai berikut :

Berikut ini merupakan flowchar metode C4.5 dapat dilihat pda gambar berikut ini:



Gambar 2 Flowchart Metode C4.5

Berikut ini ialah langkah-langkah perhitungan manual dalam Data Mining Crisp-Dm C4.5 dalam Estimasi Penjualan Produk Pada PT. Sari Pati Abadi Berbasis android. Dalam proses implementasi algoritma Algoritma C4.5 terdapat beberapa tahap yang harus diLariskan.

Tabel 1. Penjelasan Atribut

No	Atribut		Penjelasan
1	Jenis makanan ringan	1.	Makanan Ringan Kemasan
		2.	Makanan ringan segar
2	Satuan	1.	Kemasan/ Kotak
		2.	Kantong
		3.	Sachet
		4.	Gram

No	Atribut		Penjelasan
3	Stok Awal	1. 2. 3.	Jika stok awal < 200 = Sedikit Jika stok awal 200 – 600 = Sedang Jika stok awal > 600 = Banyak
4	Stok Akhir	1. 2. 3.	Jika stok Akhir < 200 = Sedikit Jika stok Akhir 200 – 600 = Sedang Jika stok Akhir > 600 = Banyak
5	Bulan	1. 2. 3.	April Mei Juni
5	Keterangan	1. 2.	Tidak Laris Laris

Dataset dibagi menjadi data training dan data testing yang merupakan data penjualan. Proses pembagian data set dilakukan yaitu 20% untuk testing dan 80% untuk data training. Maka hasil penentuan split data sebagai berikut :

Data Testing = 20% * 500 = 100

Data Training = 80% * 500 = 400

Langkah 1: Menghitung jumlah kasus, jumlah kasus untuk keputusan Tidak Laris, jumlah kasus untuk keputusan Laris, dan Entropy dari semua kasus yang dibagi berdasarkan atribut Jenis makanan, Satuan, Stok awal, stok akhir, bulan dan keterangan. Setelah itu dilakukan penghitungan Gain untuk masing-masing atribut.

Langkah 2: Menghitung Entropy pada baris total dengan persamaan (1). Selanjutnya dilakukan penghitungan Gain untuk masing-masing atribut dengan persamaan (2). Berikut ini adalah perhitungan nilai entropy dan gain.

Berikut ini merupakan Total Keterangan dari tabel

Tabel 2. Tabel Total

No	Keterangan	Jumlah
1	Laris	43
2	Tidak Laris	57

Menghitung Entropy total :

$$\begin{aligned}
 \text{Entropy [Total]} &= \left(-\frac{57}{100} \times \log_2 \frac{57}{100}\right) + \left(-\frac{43}{100} \times \log_2 \frac{43}{100}\right) \\
 &= \left(-\frac{57}{100} \times 0,30103 \times \frac{57}{100}\right) + \left(-\frac{43}{100} \times 0,30103 \times \frac{43}{100}\right) \\
 &= \left(-\frac{57}{100} \times 0,30103 \times 0,57\right) + \left(-\frac{43}{100} \times 0,30103 \times 0,43\right) \\
 &= \left(-\frac{57}{100} \times 0,17158\right) + \left(-\frac{43}{100} \times 0,12944\right) \\
 &= (-0,57 \times 0,17158) + (-0,43 \times 0,12944) \\
 &= -0,0978 + (-0,0556) \\
 &= -0,1534
 \end{aligned}$$

Menghitung entropy dan gain Jenis makanan :

Tabel 3 Jenis makanan

No	Keterangan	Laris	Tidak Laris	Total
1	Makanan ringan segar	20	42	62
2	Makanan Ringan Kemasan	23	15	38

Makanan Ringan Kemasan = $[23,15] = 38$
 $= -(23/38) * 0.30103 * (23/38) - ((15/38) * 0.30103 * (15/38))$
 Entropy Makanan Ringan Kemasan = -0.15719
 Makanan Ringan Segar = $[20,42] = 62$
 $= -(20/62) * 0.30103 * (20/62) - ((42/62) * 0.30103 * (42/62))$
 Entropy Makanan Ringan Segar = -0.16947
 $IG = -0.15347 - ((38/100) * -0.15719) + ((62/100) * -0.16947) = 0.0113340$
 Menghitung entropy dan gain Satuan :

Tabel 4.Satuan

No	Keterangan	Laris	Tidak Laris	Total
1	Kemasan/ Kotak	1	3	4
2	Kantong	6	8	14
3	Sachet	17	27	44
4	Gram	19	19	38

Kemasan/Kotak = $[1,3] = 4$
 $Kemasan/Kotak = -(1/4) * 0.30103 * (1/4) - ((3/4) * 0.30103 * (3/4))$
 Entropy Kemasan/Kotak = -0.18814
 Kantong = $[6,8] = 14$
 $Kantong = -(6/14) * 0.30103 * (6/14) - ((8/14) * 0.30103 * (8/14))$
 Entropy Kantong = -0.15359
 Sachet = $[17,27] = 44$
 $Sachet = -(17/44) * 0.30103 * (17/44) - ((27/44) * 0.30103 * (27/44))$
 Entropy Sachet = -0.15829
 Gram = $[19,19] = 38$
 $Gram = -(19/38) * 0.30103 * (19/38) - ((19/38) * 0.30103 * (19/38))$
 Entropy Gram = -0.15052
 $IG = -0.15347 - ((4/100) * -0.18814) + ((14/100) * -0.15359) + ((44/100) * -0.15829) + ((38/100) * -0.15052) = 0.0024030$
 Menghitung entropy dan gain Stok Awal :

Tabel 5.Stok Awal

No	Keterangan	Laris	Tidak Laris	Total
1	Banyak	23	16	39
2	Sedang	17	34	51
3	Sedikit	3	7	10

Banyak = $[23,16] = 39$
 $Banyak = -(23/39) * 0.30103 * (23/39) - ((16/39) * 0.30103 * (16/39))$
 Entropy Banyak = -0.15536
 Sedang = $[17,34] = 51$
 $Sedang = -(17/51) * 0.30103 * (17/51) - ((34/51) * 0.30103 * (34/51))$
 Entropy Sedang = -0.16724
 Sedikit = $[3,7] = 10$
 $Sedikit = -(3/10) * 0.30103 * (3/10) - ((7/10) * 0.30103 * (7/10))$
 Entropy Sedikit = -0.1746
 $IG = -0.15347 - ((39/100) * -0.15536) + ((51/100) * -0.16724) + ((10/100) * -0.1746) = 0.0098730$
 Menghitung entropy dan gain Stok Akhir :

Tabel 6. Stok Akhir

No	Keterangan	Laris	Tidak Laris	Total
1	Banyak	2	3	5
2	Sedang	13	16	29
3	Sedikit	28	38	66

Banyak = $[2,3] = 5$

Banyak = $(-(2/5) * 0.30103 * (2/5)) - ((3/5) * 0.30103 * (3/5))$

Entropy Banyak = -0.15654

Sedang = $[13,16] = 29$

Sedang = $(-(13/29) * 0.30103 * (13/29)) - ((16/29) * 0.30103 * (16/29))$

Entropy Sedang = -0.15213

Sedikit = $[28,38] = 66$

Sedikit = $(-(28/66) * 0.30103 * (28/66)) - ((38/66) * 0.30103 * (38/66))$

Entropy Sedikit = -0.15397

IG = $-0.15347 - ((5/100) * -0.15654) + ((29/100) * -0.15213) + ((66/100) * -0.15397) = 0.0000950$

Menghitung entropy dan gain Bulan :

Tabel 7. Bulan

No	Keterangan	Laris	Tidak Laris	Total
1	April	8	16	24
2	Mei	23	26	49
3	Juni	12	15	27

April = $[8,16] = 24$

April = $(-(8/24) * 0.30103 * (8/24)) - ((16/24) * 0.30103 * (16/24))$

Entropy April = -0.16724

Mei = $[23,26] = 49$

Mei = $(-(23/49) * 0.30103 * (23/49)) - ((26/49) * 0.30103 * (26/49))$

Entropy Mei = -0.15108

Juni = $[12,15] = 27$

Juni = $(-(12/27) * 0.30103 * (12/27)) - ((15/27) * 0.30103 * (15/27))$

Entropy Juni = -0.15237

IG = $-0.15347 - ((24/100) * -0.16724) + ((49/100) * -0.15108) + ((27/100) * -0.15237) = 0.0018370$

Hasil perhitungan ditunjukkan pada tabel IV.7. Perhitungan Node 1

Tabel 8.. Hasil Perhitungan Entropy

Node akar 1		Jumlah	Tidak Laris	Laris	Entropy	Gain
Total		100	57	43	-0.15347	
Jenis makanan	Makanan Ringan Kemasan	38	12	15	-0.15719	0.0113340
	Makanan Ringan Segar	62	20	42	-0.16947	
Satuan	Kemasan/ Kotak	4	3	1	-0.18814	0.0024030
	Kantong	14	8	6	-0.15359	
	Sachet	44	27	7	-0.15829	
	Gram	38	19	19	-0.15052	
Stok Awal	Banyak	39	16	13	-0.15536	0.0098730
	Sedang	51	34	17	-0.16724	
	Sedikit	10	7	3	-0.1746	
Stok Akhir	Banyak	5	3	2	-0.15654	0.0000950
	Sedang	29	16	13	-0.15213	
	Sedikit	66	38	28	-0.15397	

Bulan	April	24	16	8	-0.16724	0.0018370
	Mei	49	26	23	-0.15108	
	Juni	27	15	12	-0.15237	

Maka dengan demikian dapat diketahui hasil Kesimpulan dari table diatas Adalah sebagai berikut :

Tabel 9. Urutan Hasil Perhitungan Entropy

No	Atribut	Gain
1	Jenis	0.0113340
2	Stok Awal	0.0098730
3	Satuan	0.0024030
4	Bulan	0.0018370
5	Stok Akhir	0.0000950

Tahapan ini melibatkan penilaian kinerja pola yang dihasilkan oleh algoritma. Evaluasi dilakukan dengan menggunakan Confusion Matrix untuk mengukur nilai akurasi, presisi dan recall. Digunakan untuk menguji model terhadap dataset uji yang terpisah untuk mengevaluasi kinerjanya. Maka dengan demikian dapat diketahui rule yang terbentuk sebagai berikut:

Tabel 10. Rule Yang Terbentuk

No.	Aturan
1	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Sachet AND stok awal = Banyak THEN keterangan = Laris
2	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Sachet AND stok awal = Sedang THEN keterangan = Laris
3	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Kantong THEN keterangan = Laris
4	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Gram THEN keterangan = Tidak Laris
5	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = Juni THEN keterangan = Laris
6	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = Mei AND stok awal = Sedang AND satuan = Gram THEN keterangan = Tidak Laris
7	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = Mei AND stok awal = Banyak AND satuan = Gram THEN keterangan = Laris
8	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Banyak AND bulan = Mei AND satuan = Gram THEN keterangan = Laris
9	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Banyak AND bulan = April THEN keterangan = Laris
10	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Banyak AND bulan = Juni THEN keterangan = Laris
11	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Sedang AND bulan = Mei AND satuan = Gram THEN keterangan = Laris
12	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Sedang AND bulan = April THEN keterangan = Laris
13	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Sedang AND bulan = Juni THEN keterangan = Tidak Laris
14	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Sedikit AND bulan = Mei THEN keterangan = Tidak Laris

No.	Aturan
15	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Sedikit AND bulan = Juni THEN keterangan = Laris
16	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedikit AND stok awal = Sedikit AND bulan = April THEN keterangan = Tidak Laris
17	IF jenis = Makanan Ringan Kemasan AND stok akhir = Banyak THEN keterangan = Tidak Laris
18	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Sachet AND stok akhir = Sedang AND bulan = Juni THEN keterangan = Laris
19	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Sachet AND stok akhir = Sedang AND bulan = April THEN keterangan = Tidak Laris
20	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Sachet AND stok akhir = Sedang AND bulan = Mei THEN keterangan = Laris
21	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Sachet AND stok akhir = Sedikit AND bulan = Juni THEN keterangan = Laris
22	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Sachet AND stok akhir = Sedikit AND bulan = Mei THEN keterangan = Laris
23	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Sachet AND stok akhir = Banyak AND bulan = April THEN keterangan = Laris
24	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Kantong AND bulan = Mei AND stok akhir = Sedikit THEN keterangan = Laris
25	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Kantong AND bulan = Mei AND stok akhir = Sedang THEN keterangan = Laris
26	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Kantong AND bulan = Juni THEN keterangan = Tidak Laris
27	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Gram THEN keterangan = Tidak Laris
28	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Kemasan/Kotak AND stok akhir = Banyak THEN keterangan = Tidak Laris
29	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Kemasan/Kotak AND stok akhir = Sedikit THEN keterangan = Tidak Laris
30	IF jenis = Makanan Ringan Segar AND stok awal = Banyak AND satuan = Kemasan/Kotak AND stok akhir = Sedang THEN keterangan = Laris
31	IF jenis = Makanan Ringan Segar AND stok awal = Sedikit AND satuan = Kantong AND stok akhir = Sedikit AND bulan = Juni THEN keterangan = Laris
32	IF jenis = Makanan Ringan Segar AND stok awal = Sedikit AND satuan = Sachet THEN keterangan = Tidak Laris
33	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = Juni AND stok akhir = Sedikit THEN keterangan = Tidak Laris
34	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = Juni AND stok akhir = Sedang AND satuan = Kantong THEN keterangan = Tidak Laris
35	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = Juni AND stok akhir = Sedang AND satuan = Sachet THEN keterangan = Tidak Laris
36	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = Mei AND satuan = Sachet AND stok akhir = Sedikit THEN keterangan = Tidak Laris
37	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = Mei AND satuan = Gram AND stok akhir = Sedikit THEN keterangan = Tidak Laris

No.	Aturan
38	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = Mei AND satuan = Gram AND stok akhir = Sedang THEN keterangan = Tidak Laris
39	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = Mei AND satuan = Kantong THEN keterangan = Laris
40	IF jenis = Makanan Ringan Segar AND stok awal = Sedang AND bulan = April THEN keterangan = Tidak Laris

Maka dengan demikian deiperolah hasil perhitungan confusion matriks sebagai berikut :

Tabel 11. confusion matrix

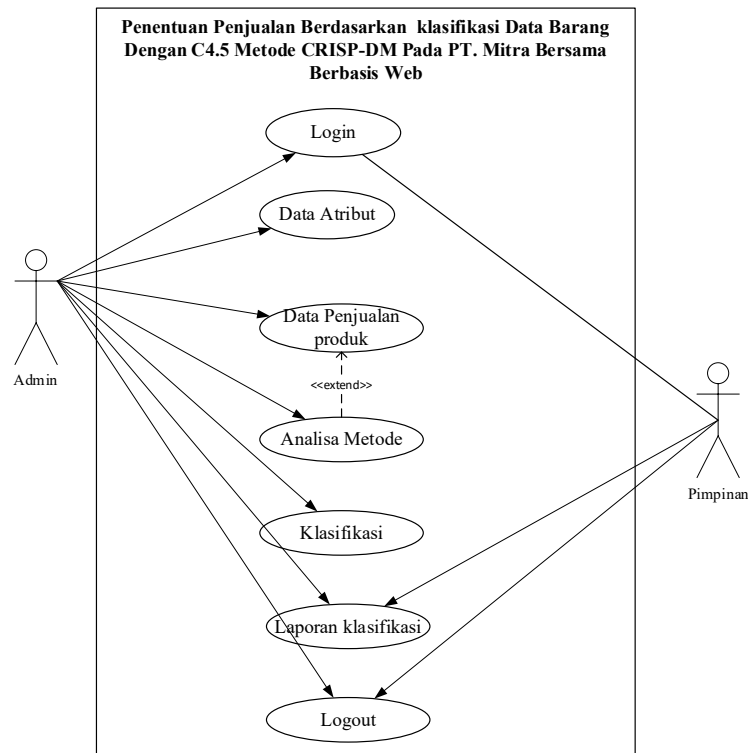
Confusion Matrix		Nilai aktual	
		Positif	Negatif
Nilai Prediksi	Positif	170	30
	Negatif	40	160
Jumlah		210	190
Total		400	

Confusion matrix bertujuan untuk memvisualisasikan prediksi dan keadaan aktual dari data yang dihasilkan oleh algoritma machine learning. Hal ini dilakukan dengan menghitung akurasi, presisi, recall, dan F1-score. Metrik-metrik ini sangat penting dalam mengevaluasi kinerja klasifikasi atau algoritma machine learning yang digunakan untuk melakukan prediksi. Rumus untuk keempat metrik ini ditunjukkan dalam tabel berikut (Stacyana Jesika et al., 2023):

Tabel 12.rumus confusion matrix

No	Pengukuran	Rumus	Hasil	Hasil (%)
1	Akurasi	$Accuracy = \frac{170+160}{170+160+40+30} * 100\%$	0.825	82.5%
2	Presisi	$Precision = \frac{170}{170+40} * 100\%$	0.8095	80.95%
3	Recall	$Recall = \frac{170}{170+30} * 100\%$	0.85	85%
		$F1-Score = \frac{2 \times 0.8095 \times 0.85}{0.8095 + 0.85} * 100\%$	0.8290	82.9%

Sebuah *use case* digambarkan sebagai elips horizontal dalam suatu diagram UML *use case*, dapat dilihat pada gambar 3:



Gambar 3. Use Case Diagram

3.2. Hasil

1. Tampilan Form Penjualan

Tampilan ini merupakan tampilan *form* penjualan yang berfungsi untuk mengisi Input penjualan. Gambar tampilan *form* penjualan pada gambar 4

Gambar 4 Tampilan Form Penjualan

2. Tampilan Form Analisa Metode

Tampilan ini merupakan tampilan *form* analisa metode yang berfungsi untuk mengisi Input analisa metode. Gambar tampilan *form* analisa metode pada gambar 5.

No.	Aturan
1	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Sachet AND stok awal = Banyak THEN keterangan = Tidak Laris
2	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Sachet AND stok awal = Sedang THEN keterangan = Tidak Laris
3	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Kantong THEN keterangan = Laris
4	IF jenis = Makanan Ringan Kemasan AND stok akhir = Sedang AND bulan = April AND satuan = Gram THEN keterangan = Tidak Laris
5	IF jenis = Makanan Ringan Kemasan AND stok akhir = Banyak AND bulan = April AND satuan = Sachet THEN keterangan = Tidak Laris

Gambar 5. Tampilan Form Analisa Metode

3. Tampilan *Form* perhitungan C4.5

Tampilan ini merupakan tampilan *form* input perhitungan C4.5 yang berfungsi untuk mengisi form perhitungan C4.5. Berikut gambar *form* input data perhitungan C4.5 ditunjukkan pada Gambar 6:

Entropy Keterangan [43,57]
 $= -((43/100) \times \log_2(43/100)) - ((57/100) \times \log_2(57/100)) = -0.15347$

$= [23,15] = 38$
 $= -((23/38) \times 0.30103 \times (23/38)) - ((15/38) \times 0.30103 \times (15/38))$
 Entropy = -0.15719

$= [20,42] = 62$
 $= -((20/62) \times 0.30103 \times (20/62)) - ((42/62) \times 0.30103 \times (42/62))$
 Entropy = -0.16947

IG = -0.15347 - ((38/100) * -0.15719) + ((62/100) * -0.16947) = 0.0113340

$= [1,3] = 4$
 $= -((1/4) \times 0.30103 \times (1/4)) - ((3/4) \times 0.30103 \times (3/4))$
 Entropy = -0.18814

$= [6,8] = 14$
 $= -((6/14) \times 0.30103 \times (6/14)) - ((8/14) \times 0.30103 \times (8/14))$
 Entropy = -0.15359

$= [17,27] = 44$
 $= -((17/44) \times 0.30103 \times (17/44)) - ((27/44) \times 0.30103 \times (27/44))$
 Entropy = -0.15829

$= [19,19] = 38$
 $= -((19/38) \times 0.30103 \times (19/38)) - ((19/38) \times 0.30103 \times (19/38))$
 Entropy = -0.15052

IG = -0.15347 - ((44/100) * -0.18814) + ((14/100) * -0.15359) + ((62/100) * -0.15829) + ((38/100) * -0.15052) = 0.0024030

Gambar 6. Tampilan *Form* perhitungan C4.5

4. Tampilan *Form* klasifikasi

Tampilan ini merupakan tampilan *form* input klasifikasi yang berfungsi untuk mengisi form klasifikasi. Berikut gambar *form* input data klasifikasi ditunjukkan pada Gambar 7 :

Gambar 7. Tampilan *Form* Klasifikasi

5. Tampilan *Form* Atribut

Tampilan ini merupakan tampilan *form* atribut yang berfungsi untuk mengisi data-data atribut. Berikut gambar *form* laporan atribut ditunjukkan pada Gambar 8 :

Gambar 8. Tampilan *Form* Atribut

6. Tampilan *Form* Laporan klasifikasi

Tampilan ini merupakan tampilan *form* laporan klasifikasi yang berfungsi untuk mengisi data-data klasifikasi. Berikut gambar *form* laporan klasifikasi ditunjukkan pada Gambar 9 :



The screenshot shows the Sari Pati Abadi application interface. At the top, there is a header bar with the company name and logo. Below the header, there is a section titled 'Laporan Klasifikasi'. This section contains a table with the following data:

No.	Nama Produk	Periode	Status
1	Astor Coklat	Agustus 2025	Tidak Laris

Gambar 9. Tampilan *Form* Laporan Klasifikasi

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan selama membuat Data Mining Crisp-Dm C4.5 dalam Estimasi Penjualan Produk Pada PT. Sari Pati Abadi Berbasis android ini, maka dapat ditarik beberapa kesimpulan sebagai berikut :

1. Dengan membangun aplikasi ini dapat lebih mudah dalam menentukan klasifikasi Penjualan produk pada PT. Sari Pati Abadi dapat membantu PT. Sari Pati Abadi dalam menyediakan produk yang akan di jual.
2. Dengan adanya aplikasi ini dapat membantu PT. Sari Pati Abadi dalam penentuan stok barang sesuai kebutuhan dengan menggunakan metode C4.5.
3. Aplikasi yang dirancang dengan menggunakan Android Studio dan database Mysql, proses perancangan menggunakan UML yang terdiri dari usecase diagram, class diagram, activity diagram dan sequence diagram.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Potensi Utama yang telah banyak memberikan masukan dan saran dalam penyelesaian penelitian ini.

REFERENSI

- [1] Audrilia, M., & Budiman, A. (2020). Perancangan Sistem Informasi Manajemen Bengkel Berbasis Web (Studi Kasus : Bengkel Anugrah). *Jurnal Madani : Ilmu Pengetahuan, Teknologi, Dan Humaniora*, 3(1), 1–12. <https://doi.org/10.33753/madani.v3i1.78>
- [2] Fatawa Imam Al Muftin, & Fendi Hidayat. (2024). Sistem Informasi Penjualan. *Zona Komputer: Program Studi Sistem Informasi Universitas Batam*, 13(3), 232–237. <https://doi.org/10.37776/zkomp.v13i3.1461>
- [3] Gustami, E., & Astuti, K. P. (2022). Klasifikasi Barang Menggunakan Algoritma C4. 5 Dalam Penentuan Prediksi Stok Barang Pada Pt Aerofood Indonesia. *Reputasi: Jurnal Rekayasa Perangkat ...*, 3(2), 12–18.
- [4] Hidayati, N., Suntoro, J., & Setiaji, G. G. (2021). Perbandingan Algoritma Klasifikasi untuk Prediksi Cacat Software dengan Pendekatan CRISP-DM. *Jurnal Sains Dan Informatika*, 7(2), 117–126. <https://doi.org/10.34128/jsi.v7i2.313>
- [5] Indah Sari, A., & Sewaka. (2022). Prediksi Penjualan Sepatu Dengan Algoritma C4.5 Di T&T Collection Di Tangerang. *OKTAL : Jurnal Ilmu Komputer Dan Science*, 1(9), 1435–1442.

- [5] Janis, J. W., Mamahit, D. J., Sugiarto, B. A., Rumagit, A. M., Elektro, T., Sam, U., & Manado, R. (2020). Rancang Bangun Aplikasi Online Sistem Pemesanan Jasa Tukang Bangunan Berbasis Lokasi. *Jurnal Teknik Informatika*, 15(1), 1–12. <https://doi.org/10.35793/jti.15.1.2020.29023>
- [7] Leasiwal, A. A., Andrian, A. M., & Masala, I. V. (2021). Implementasi Algoritma C4. 5 Untuk Klasifikasi Harga Emas. *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, 06, 1–12.
- [8] Narayana Sakti Aji, Fauzan Natsir, & Siti Istianah. (2023). Penentuan Penjualan Barang Berdasarkan Pengelompokan Produk dengan K-Means Clustering Metode CRISP-DM Pada CV Sembako Dina. *Journal Zetroom*, 5(2), 119–126. <https://doi.org/10.36526/ztr.v5i2.3041>
- [9] Navisa, S., Luqman Hakim, & Aulia Nabilah. (2021). Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM. *Jurnal Sistem Cerdas*, 4(2), 114–125. <https://doi.org/10.37396/jsc.v4i2.162>
- [10] Pala'langan, E. I., Saerang, D. P. E., & Gamaliel, H. (2020). Analysis of Accounting Information System Sales At Pt. Wahana Wirawan Manado-Nissan Datsun Martadinata. *Jurnal EMBA*, 8(4), 824–834.
- [11] Pratiwi, Y. A., Ginting, R. U., Situmoran, H., & Sitanggang, R. (2020). Perancangan Sistem Informasi Akademik Berbasis Web Di Smp Rahmat Islamiyah. *Jurnal Teknologi, Kesehatan Dan Ilmu Sosial*, 2(1), 27–32.
- [12] Priyatna, A., & Sanwani. (2024). Evaluasi Pemahaman Siswa Dalam Proses Belajar Secara Online dengan Menggunakan Algoritma C5.0. *Resolusi : Rekayasa Teknik Informatika Dan Informasi*, 4(3), 300–309.
- [13] Ramayu, I. M. S., Susanto, F., & Mahendra, G. S. (2022). Penerapan Data Mining Dengan Algoritma C4.5 Dalam Pemesanan Obat Guna Meningkatkan Keuntungan Apotek. *Prosiding Seminar Nasional Manajemen, Desain & Aplikasi Bisnis Teknologi (SENADA)*, 5, 237–245.
- [14] Salsabilla, T., & Sulastri, S. (2022). Implementasi Algoritma C4.5 Untuk Klasifikasi Produk Laris Sepeda Motor Honda Pada Cv Cendana Motor Cepiring. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 7(2), 164–171. <https://doi.org/10.36341/rabit.v7i2.2489>
- [15] Saputra, A. D., & Qoiriah, A. (2022). Penerapan Algoritma C4.5 Untuk Mengatur Persediaan Stok Barang Berbasis Website. *Journal of Informatics and Computer Science (JINACS)*, 3(04), 481–493. <https://doi.org/10.26740/jinacs.v3n04.p481-493>
- [16] Shedriko, S., & Firdaus, M. (2022). Penentuan Klasifikasi Dengan Crisp-Dm Dalam Memprediksi Kelulusan Mahasiswa Pada Suatu Mata Kuliah. *Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi)*, 6(1), 826–831. <https://doi.org/10.30998/semnasristek.v6i1.5814>
- [17] Suhartini, Sadali, M., & Putra, Y. K. (2020). Sistem Informasi Berbasis Web Sma Al- Mukhtariyah Mamben Lauk Berbasis Php Dan Mysql . *Jurnal Informatika Dan Teknologi*, 3(1), 79–83.
- [18] Wahyudi, M. D. (2023). Penerapan Data Mining Dengan Algoritma C4. 5 Dalam Prediksi Penjualan Buku. *Jurnal Teknorama (Informatika Dan ...)*, 1(1), 1–6.