

IMPLEMENTASI METODE *DECISION TREE* DALAM MEMPREDIKSI PENJUALAN PRODUK TERLARIS (STUDI KASUS : *KREATIF MART* ATAMBUA)

Sonya Oktavia Usnaat¹, Yoseph P.K Kelen², Krisantus J. Tey Seran³

^{1,2,3}Universitas Timor

(Program Studi Teknologi Informasi Fakultas Pertanian Sains dan Kesehatan Universitas
Timor)

(Jl. Km. 09 Kelurahan Sasi, Kecamatan Kota Kefamenanu, telp. 0812-3956-4700)

e-mail: ¹sonyaoktaviausnaat@gmail.com, ²yosepkelen@unimor.ac.id,

³krisantusteyseran@unimor.ac.id

Abstrak

Kebutuhan pokok atau dasar merupakan hal yang penting bagi manusia dalam menjalani kesehariannya. Pada masa era globalisasi ini, produk makanan dan minuman sangat mudah didapatkan karena saat ini banyak bidang usaha manusia yang bergerak dalam lingkup produsen dan jasa produk. Di negara Indonesia, persebaran usaha sudah sangat banyak, khususnya di Pulau Timor Kabupaten Belu yang beribukota di Atambua. Dengan letak geografinya yang menjadi pintu masuk ke negara tetangga, menjadikan kota ini sebagai tempat persinggahan bagi masyarakat yang ingin ke Timor Leste. Dijadikan tempat persinggahan tentunya memungkinkan terjadinya transaksi jual beli barang. Dengan banyaknya transaksi jual beli yang terjadi, perputaran uang di tempat ini menjadi cepat dan mengakibatkan stok barang yang dijual toko cepat habis khususnya barang produk. Satu dari sekian banyak pusat perbelanjaan yang paling laris di kota Atambua adalah toko Kreatif Mart yang berlokasi di Pasar Baru Atambua.

Tingginya transaksi berbelanja di toko ini, maka stok barang yang berada di gudang cepat mengalami penurunan. Diperlukan prediksi yang tepat dalam menangani masalah ini agar rantai pasok barang tidak terputus karena ada barang yang kosong saat dicari oleh pembeli. Penelitian ini bertujuan untuk menerapkan metode *Decision Tree* dengan algoritma ID3 untuk memprediksi produk terlaris di Kreatif Mart, Atambua. Data diambil selama satu bulan dan dianalisis berdasarkan variabel seperti stok, jumlah penjualan, dan harga. Hasil dari penelitian ini menunjukkan akurasi prediksi yang baik, membantu toko dalam manajemen persediaan untuk memenuhi kebutuhan konsumen dengan lebih efektif.

Kata kunci: *Decision Tree*, Data Mining, Python, ID3, Prediksi Penjualan.

Abstract

Basic or basic needs are important for humans in living their daily lives. In this era of globalization, food and beverage products are very easy to obtain because currently many fields of human business are engaged in the scope of product producers and services. In Indonesia, the distribution of businesses is very large, especially on the island of Timor, Belu Regency, whose capital is Atambua. With its geographical location being the entrance to neighboring countries, this city has become a stopover place for people who want to go to Timor Leste. Being a stopover place certainly allows for buying and selling goods transactions. With so many buying and selling transactions taking place, the circulation of money in this place becomes fast and causes the stock of goods sold by the shop to run out quickly, especially product goods. One of the best-selling shopping centers in Atambua city is the creative mart shop located at Pasar Baru Atambua. The high number of shopping transactions at this shop means that the stock of goods in the warehouse is rapidly decreasing. Correct predictions are needed in dealing with this problem so

that the supply chain of goods is not interrupted because an item is empty when the buyer is looking for it. This research aims to apply the Decision Tree method with the ID3 algorithm to predict the best-selling products at Creative Mart, Atambua. Data is taken for one month and analyzed based on variables such as stock, number of sales and price. The results of this research show good prediction accuracy, helping stores in inventory management to meet consumer needs more effectively.

Keywords: decision tree, data mining, python, ID3, Sales Prediction.

1. PENDAHULUAN

Kebutuhan pokok atau dasar merupakan hal yang penting bagi manusia dalam menjalani kesehariannya. Kebutuhan manusia sendiri terdiri dari tiga komponen kebutuhan pokok manusia yaitu Sandang, Papan dan Pangan [1]. Ketiga komponen ini sangat penting bagi kehidupan manusia, akan tetapi komponen utama dan mendasar dalam kehidupan manusia saat ini adalah komponen Pangan. Komponen pangan yang dimaksud adalah makanan dan minuman yang sangat dibutuhkan oleh manusia untuk menambahkan energi agar tetap beraktivitas sehari – hari.

Pada masa era globalisasi ini, produk makanan dan minuman sangat mudah didapatkan karena saat ini banyak bidang usaha manusia yang bergerak dalam lingkup produsen dan jasa produk. Toko dan kios merupakan contoh bidang usaha yang bergerak di pemasaran (penjualan) berbagai macam produk [2]. Kios merupakan tempat penjualan produk yang paling laris saat ini untuk masyarakat kalangan menengah ke bawah. Alasan ini menjadi salah satu ladang usaha bagi pendapatan masyarakat baik dari kota sampai di desa. Kios merupakan salah satu tempat usaha dagang yang penting bagi kebutuhan para pedagang [3].

Di Negara Indonesia, persebaran usaha kios sudah sangat banyak, khususnya di Pulau Timor Kabupaten Belu yang beribukota di Atambua. Atambua merupakan kota perbatasan antara Republik Indonesia dengan Timor Leste [4]. Dengan letak geografisnya yang menjadi pintu masuk ke negara tetangga, menjadikan kota ini sebagai tempat persinggahan (transit) bagi masyarakat yang ingin ke Timor Leste. Dijadikan tempat persinggahan tentunya memungkinkan terjadinya transaksi jual beli barang [5]. Sebagai contoh banyak warga negara Timor Leste yang datang dan berbelanja kebutuhan rumah tangga (Sandang, Pangan, Papan) di Kota Atambua. Dengan banyaknya transaksi jual beli yang terjadi, perputaran uang di tempat ini menjadi cepat dan mengakibatkan stok barang yang dijual di kios dan toko cepat habis khususnya barang produk. Satu dari sekian banyak pusat perbelanjaan yang paling laris di kota atambua adalah toko kreatif mart yang berlokasi di Pasar Baru Atambua.

Tingginya transaksi berbelanja di toko ini, maka stok barang yang berada di gudang cepat mengalami penurunan. Diperlukan prediksi yang tepat dalam menangani masalah ini agar rantai pasok barang tidak terputus karena ada barang yang kosong saat dicari oleh pembeli. Ada banyak cara dalam memprediksi stok barang di gudang, dengan prediksi disiplin ilmu yang berbeda. Mulai dari ilmu ekonomi sampai dengan ilmu Teknologi Informasi (*Data Science*). Dalam ilmu *Data Science* salah satu teknik yang terkenal dalam proses prediksi adalah teknik penambahan data atau yang sering disebut *Data Mining* [6].

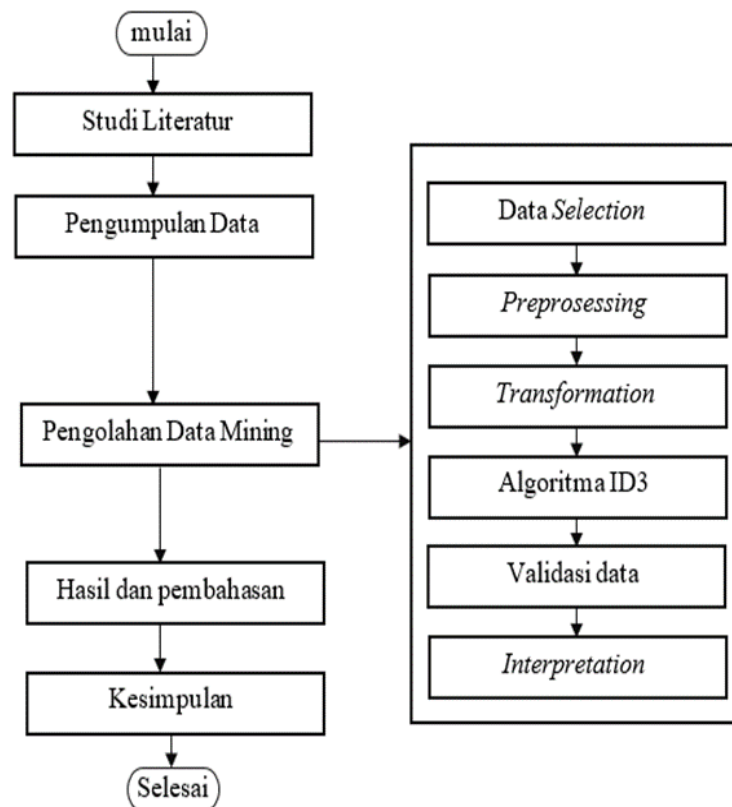
Data Mining terdiri dari dua kelompok besar yaitu klasifikasi dan klustering [7]. Klasifikasi merupakan proses mengelompokkan suatu data menjadi kelas-kelas tertentu dengan memanfaatkan struktur data. Klustering adalah teknik pengelompokan record data, pengamatan atau kasus dalam kelas yang memiliki kemiripan. Dalam klasifikasi ada berbagai macam algoritma yang bisa digunakan. Algoritma *Decision Tree* adalah salah satu metode yang cukup mudah untuk diinterpretasikan oleh manusia [8]. Algoritma *decision tree* mampu direpresentasikan pengguna dalam mempelajari alasan dibalik setiap keputusan yang diambil, proses pengambilan keputusan tidak terbatas dan mampu menangani dengan baik dalam masalah yang kompleks. Algoritma *decision tree* tidak memerlukan normalisasi dalam proses pemodelan. Data yang akan digunakan dalam penelitian ini diambil dari Toko Kreatif Mart selama Satu bulan yakni april 2024. Data yang dianalisis berupa transaksi penjualan produk sehingga dapat diprediksi barang apa saja yang paling dicari atau dibeli oleh konsumen. Atribut data analisis yang digunakan dalam penelitian ini adalah merk produk, stok produk, jumlah penjualan dan

harga. Diharapkan dengan implementasi metode *Decision Tree* dalam studi kasus ini, diharapkan dapat membantu memprediksi barang produk yang paling laris saat ini, metode decision tree cukup mudah untuk diinterpretasikan oleh manusia. Maka dalam penelitian ini diberi judul “**Penerapan Metode Decision Tree Dalam Memprediksi Penjualan Produk Terlaris (Studi Kasus: Pasar Baru Atambua)**”.

2. Metode Penelitian

Metode penelitian ini memberikan penjelasan tentang langkah-langkah, data, lokasi penelitian, metode evaluasi yang digunakan serta penjelasan terstruktur tentang algoritma atau metode dari penelitian yang dibahas.

Tahapan penelitian ini menggunakan tahapan *Knowledge Discovery in Database* (KDD) yaitu sebagai berikut :



Gambar 1. Tahapan Penelitian

Berikut ada 5 tahapan dalam menggunakan tahapan *knowledge Discovery In Database* (KDD) yaitu sebagai berikut :

1. Studi Literatur

Tahapan ini merupakan tahapan awal penelitian dimana melakukan penyusunan rencana penelitian terkait dengan objek yang diambil adalah barang produk di salah satu toko yang berada pasar baru Atambua yaitu Kreatif Mart.

2. Pengumpulan Data

Setelah melakukan persiapan penelitian, tahapan selanjutnya melakukan pengumpulan data. Pengumpulan data yang digunakan dalam penelitian ini yaitu observasi, wawancara dan studi pustaka. Pengumpulan data observasi dilakukan dengan cara mengamati secara langsung pada tempat penelitian serta mencatat hal – hal penting terkait dengan objek yang akan diambil. Sedangkan pengumpulan data wawancara dilakukan dengan cara mengajukan pertanyaan secara langsung pada pemilik toko terkait dengan objek yang ada ambil. Sementara

pengumpulan data studi pustaka di lakukan dengan cara mempelajari ilmu – ilmu dari penelitian sebelumnya yang berkaitan objek penelitian.

3. Pengolahan Data Mining

Pada pengolahan data mining ini merupakan tahapan selanjutnya apabila telah melakukan tahapan pengumpulan data. Pada tahap pengolahan data mining ini menggunakan tahapan *Knowledge Discovery in Database* (KDD). Berikut adalah pengolahan data mining menggunakan tahapan *Knowledge Discovery in Database* (KDD):

a. *Data selection*

Data selection merupakan proses pemilihan atau seleksi data yang nantinya akan digunakan pada proses data mining. Data yang digunakan dalam penelitian ini adalah data penjualan sembako selama 1 bulan yang berasal dari salah satu toko sembako yang berada di Pasar Baru Atambua yaitu toko Kreatif Mart. Semua data penjualan sembako yang di seleksi diantaranya yaitu merk produk, stok produk, jumlah penjualan dan harga, sehingga semua data penjualan yang akan digunakan untuk proses *knowledge Discovery in Database* (KDD) yaitu merk produk, stok produk, jumlah penjualan dan jenis kelamin pelanggan.

b. *Preprocessing*

Pada tahap ini dilakukan dengan cara membersihkan data tidak relevan terhadap *missing value* dalam data dan memastikan bahwa tidak ada kepalsuan nilai dalam data penjualan produk dan mendapatkan hasil yang benar.

c. *Transformation*

Tahap *transformation* yaitu tahapan mengubah atau mengalihkan data yang terpilih ke dalam bentuk format yang sesuai agar dapat di proses dalam data mining.

d. *Algoritma ID3*

Pada proses ini merupakan proses mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variable target. *Decision tree* sangat bagus sebagai langkah awal dalam proses pemodelan.

e. Validasi data

Pada proses ini, data yang telah diolah atau telah diekspolarsi dalam algoritma ID3, selanjutnya adalah melakukan validasi data pada hasil Algoritma ID3.

i. *Interpretation*

Tahap ini berfungsi untuk memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Pada tahap ini menggunakan metode *Decision Tree* untuk proses data mining dan hasilnya berupa *rules* yang didapat dari perhitungan *Decision Tree*.

4. Hasil dan Pembahasan

Setelah melakukan pengolahan data mining tahap selanjutnya adalah tahap hasil dan pembahasan. Tahap ini memaparkan pembahasan dan hasil dari proses pengolahan data mining yang menerapkan metode *Decision Tree*.

5. Kesimpulan

Pada tahap ini merupakan tahapan terakhir yang dimana melakukan Kesimpulan dari hasil penelitian.

2.1 Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya [9]

2.2 Bahasa Pemrograman Python

Python adalah bahasa pemrograman yang diketik secara dinamis, sering disebut pengetikan bebek dalam rekayasa perangkat lunak [10]

2.3 Penghitungan Menggunakan Algoritma ID3

Pada tahapan kali ini proses yang cukup penting dalam pengimplementasian algoritma ID3 yang akan digunakan yakni penghitungan nilai entropy dan Gain Seperti diketahui perhitungan tersebut dilakukan untuk menentukan nilai dari sebuah atribut yang mana akan menentukan proses pembuatan Decision Tree yang akan dibuat [11]

2.4 Decision Tree

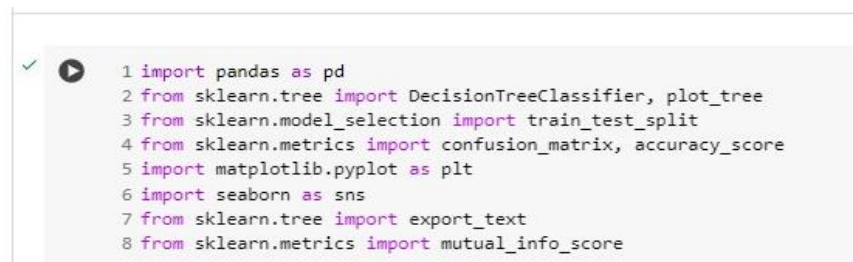
Decision tree merupakan teknik model prediksi yang dapat digunakan untuk klasifikasi dan prediksi tugas. Decision tree menggunakan teknik “membagi dan menaklukkan” untuk membagi ruang pencarian masalah menjadi himpunan masalah [12]

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Import Labrary Phyton

Dalam hasil pengujian data menggunakan *phyton*, langkah awal yang dilakukan adalah mengimport *library* dimana *library* yang digunakan adalah *pandas*, *sklearn*, *seaborn* dan *matplotlib*. Dalam *library sklearn* memiliki beberapa modul yakni *decisiontreeclassifier*, *metrics*, *train_test_split*, *confusion_matrix*, *export_text*, dan *mutual_info_score*. Codingan yang digunakan dalam mengimport *library* dan kemudian di *run* atau dijalankan jika berhasil maka akan menampilkan centang berwarna hijau di sebelah kiri *codingan* seperti pada gambar dibawah ini:



```

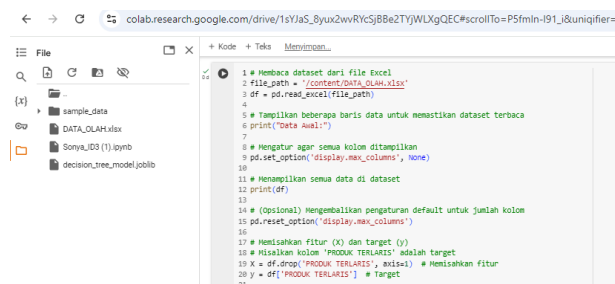
1 import pandas as pd
2 from sklearn.tree import DecisionTreeClassifier, plot_tree
3 from sklearn.model_selection import train_test_split
4 from sklearn.metrics import confusion_matrix, accuracy_score
5 import matplotlib.pyplot as plt
6 import seaborn as sns
7 from sklearn.tree import export_text
8 from sklearn.metrics import mutual_info_score

```

Gambar 2. Coding Import Labrary Python

3.1.2 Menginput Data dan Membaca Data

Setelah mengimport *labrary*, langkah berikutnya adalah menginput data dan membaca data. Data yang dimaksud adalah data yang akan dianalisis yakni data produk penjualan. Data *train* yang di *input* harus berupa *format excel* dan *csv* yang mana proses menginput dilakukan dengan cara unggah file data ke penyimpanan sesi *python*. Codingan untuk mengunggah file ke *phyton* seperti gambar di bawah ini:



```

1 # Membaca dataset dari file Excel
2 file_path = "/content/DATA_OLAH.xlsx"
3 df = pd.read_excel(file_path)
4
5 # Tampilkan beberapa baris data untuk memastikan dataset terbaca
6 print("Data Awal:")
7
8 # Mengatur agar semua kolom ditampilkan
9 pd.set_option('display.max_columns', None)
10
11 # Menampilkan semua data di dataset
12 print(df)
13
14 # (Optional) Mengembalikan pengaturan default untuk jumlah kolom
15 pd.reset_option('display.max_columns')
16
17 # Memisahkan fitur (X) dan target (y)
18 # Hapus kolom 'PRODUK TERLARIS' sebagai target
19 X = df.drop('PRODUK TERLARIS', axis=1) # Memisahkan fitur
20 y = df['PRODUK TERLARIS'] # Target

```

Gambar 3. Codingan Unggah File

Pada gambar diatas merupakan kode yang berfungsi untuk membaca file data dan yang menampilkan data awal.


```

21
22 # Membagi dataset menjadi data training dan testing
23 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=52)
24
25 # Menampilkan hasil pembagian data
26 print("Data Training (X_train):")
27 print(X_train)
28
29 print("Data Target Training (y_train):")
30 print(y_train)
31
32 print("Data Testing (X_test):")
33 print(X_test)
34
35 print("Data Target Testing (y_test):")
36 print(y_test)
37

```

Gambar 4. Coding Python untuk menampilkan jumlah data yang dibagi

Adanya pemisahan fitur dan target, yang membagi data menjadi set pelatihan (*trainingset*) dengan nilai 80% dan set pengujian (*testingset*) dengan nilai 20%, untuk melatih dan mengevaluasi model data secara terpisah.

```

1 # Memisahkan fitur (X) dan target (y)
2 X = df.drop('PRODUK TERLARIS', axis=1)
3 y = df['PRODUK TERLARIS']
4
5 # Membagi data menjadi training dan testing set (80% training, 20% testing)
6 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=52)
7

```

Gambar 5. coding python untuk menampilkan pemisahan fitur dan target.

Setelah itu, kode pada gambar diatas menyatakan bahwa kode berhasil di *run* atau dijalankan maka menampilkan data yang telah diunggah.

	KATEGORI	JUMLAH PENJUALAN	STOK	HARGA	PRODUK TERLARIS
0	Bahan Pokok	Banyak	Banyak	Mahal	Ya
1	Minuman Olahan	Banyak	Banyak	Murah	Ya
2	Minuman Olahan	Banyak	Sedikit	Murah	Ya
3	Makanan Olahan	Banyak	Sedang	Murah	Ya
4	Minuman Olahan	Banyak	Sedang	Sedang	Ya
5	Makanan Olahan	Banyak	Sedang	Murah	Ya
6	Minuman Olahan	Banyak	Sedang	Murah	Ya
7	Makanan Olahan	Banyak	Sedikit	Murah	Tidak
8	Minuman Olahan	Banyak	Sedikit	Murah	Ya
9	Makanan Olahan	Banyak	Sedikit	Murah	Ya
10	Makanan Olahan	Banyak	Sedang	Murah	Ya
11	Makanan Olahan	Banyak	Sedang	Mahal	Ya
12	Minuman Olahan	Banyak	Sedikit	Sedang	Ya
13	Makanan Olahan	Banyak	Banyak	Murah	Tidak
14	Makanan Olahan	Banyak	Sedikit	Murah	Ya
15	Bahan Pokok	Banyak	Sedikit	Murah	Ya
16	Bahan Pokok	Banyak	Sedikit	Murah	Ya
17	Makanan Olahan	Banyak	Banyak	Murah	Tidak
18	Bahan Pokok	Banyak	Sedikit	Mahal	Ya
19	Minuman Olahan	Banyak	Sedang	Murah	Tidak
20	Minuman Olahan	Banyak	Sedang	Murah	Tidak
21	Makanan Olahan	Banyak	Sedikit	Murah	Tidak
22	Minuman Olahan	Banyak	Sedang	Murah	Tidak
23	Minuman Olahan	Banyak	Sedang	Murah	Tidak
24	Makanan Olahan	Banyak	Sedikit	Murah	Ya
25	Minuman Olahan	Banyak	Sedang	Mahal	Ya
26	Minuman Olahan	Banyak	Sedikit	Mahal	Ya
27	Makanan Olahan	Banyak	Sedikit	Murah	Tidak
28	Makanan Olahan	Banyak	Sedikit	Mahal	Ya
29	Makanan Olahan	Banyak	Sedikit	Mahal	Ya
30	Minuman Olahan	Banyak	Sedikit	Mahal	Ya
31	Makanan Olahan	Banyak	Sedikit	Mahal	Ya
32	Minuman Olahan	Sedikit	Sedikit	Murah	Tidak
33	Makanan Olahan	Sedikit	Sedang	Mahal	Tidak
34	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
35	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
36	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
37	Makanan Olahan	Sedikit	Sedikit	Murah	Tidak
38	Bahan Pokok	Sedikit	Sedikit	Mahal	Tidak
39	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
40	Makanan Olahan	Sedikit	Sedang	Mahal	Tidak
41	Minuman Olahan	Sedikit	Sedikit	Murah	Tidak

42	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
43	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
44	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
45	Makanan Olahan	Sedikit	Sedikit	Murah	Tidak
46	Makanan Olahan	Sedikit	Banyak	Murah	Tidak
47	Makanan Olahan	Sedikit	Sedikit	Murah	Tidak
48	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak
49	Bahan Pokok	Sedikit	Sedikit	Mahal	Tidak
50	Makanan Olahan	Sedikit	Sedikit	Murah	Tidak
51	Makanan Olahan	Sedikit	Sedikit	Mahal	Tidak

Gambar 6 Tampilan Data Awal (lanjutan)

File

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

<

Gambar 7. Tampilan Data Train

File

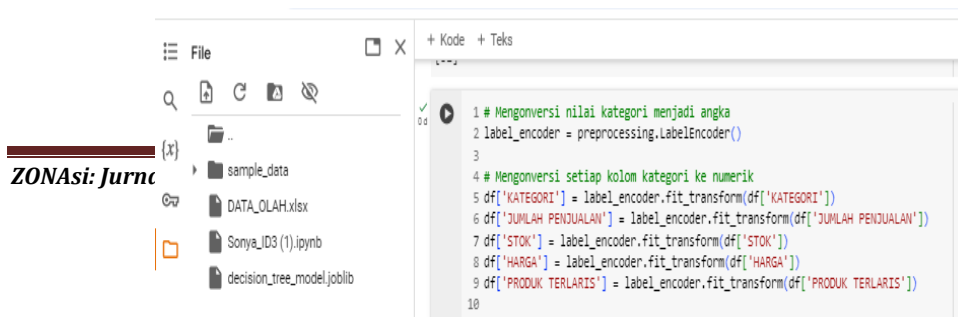
<

Gambar 8. Tampilan Data Setting

3.1.3 Mengkonversi data menjadi nilai numerik

Sebelum melakukan penerapan ID3 data yang telah diupload dilakukan *konversi* beberapa kolom dataFrame (kategori, jumlah penjualan, stok, harga, dan produk terlaris), kemudian mengonversinya ke angka menggunakan *LabelEncoder*.

Ini dilakukan agar data dapat digunakan sebagai input dalam algoritma *machine learning*, yang membutuhkan data numerik. Data dapat ditampilkan dengan *coding* seperti gambar berikut :



The screenshot shows a Jupyter Notebook interface with a single cell containing the following code:

```

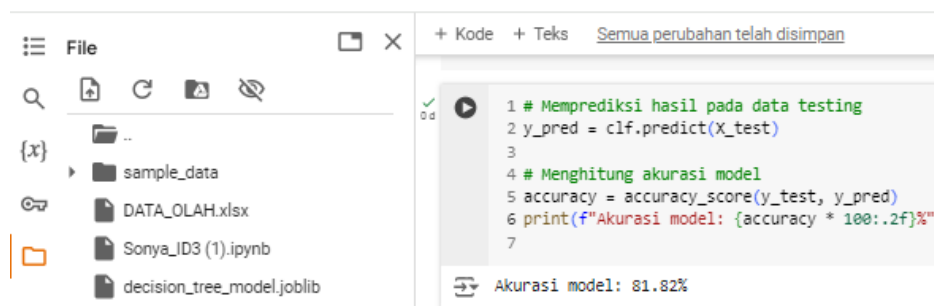
1 # Mengonversi nilai kategori menjadi angka
2 label_encoder = preprocessing.LabelEncoder()
3
4 # Mengonversi setiap kolom kategori ke numerik
5 df['KATEGORI'] = label_encoder.fit_transform(df['KATEGORI'])
6 df['JUMLAH PENJUALAN'] = label_encoder.fit_transform(df['JUMLAH PENJUALAN'])
7 df['STOK'] = label_encoder.fit_transform(df['STOK'])
8 df['HARGA'] = label_encoder.fit_transform(df['HARGA'])
9 df['PRODUK TERLARIS'] = label_encoder.fit_transform(df['PRODUK TERLARIS'])
10

```

Gambar 9. Coding dan Hasil Running Konversi Data Menjadi Nilai Numerik

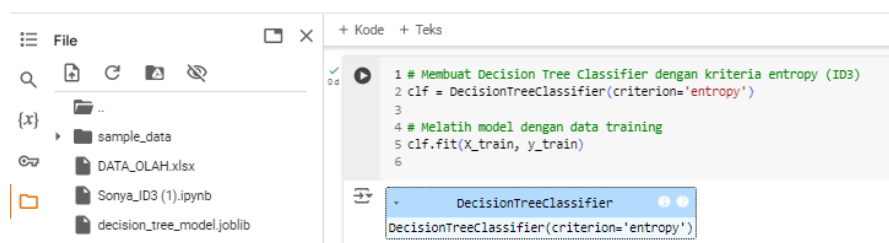
3.1.4 Menghitung Akurasi Model

Tahap ini memperoleh prediksi untuk data uji (*testingset*) menggunakan model yang telah dilatih. Menghitung *akurasi model* seberapa baik model memprediksi nilai target berdasarkan data uji dan menampilkan akurasi model dalam bentuk presentase, sehingga dapat dinilai seberapa baik model bekerja dalam memprediksi hasil. Berikut ini adalah *source code akurasi model* sebagai berikut:



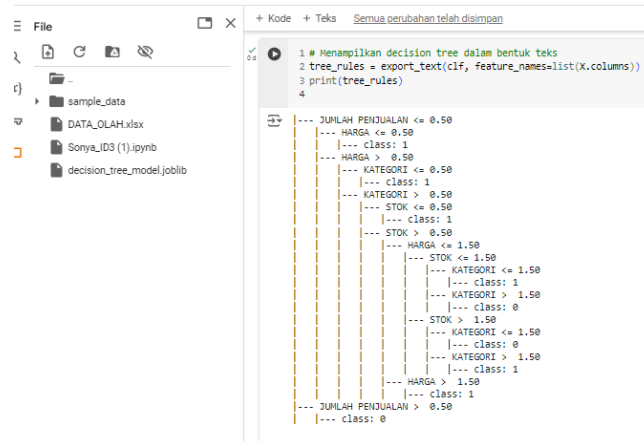
Gambar 10. Akurasi Model Dan Hasil Runningnya

Gambar di atas merupakan *akurasi model* selanjutnya kita membangun model *decisiontree* dengan menggunakan algoritma *id3* (berdasarkan kriteria entropi). Kita juga melatih model tersebut dengan data pelatihan (*trainingdata*) untuk memahami pola antar fitur-fitur dan target, sehingga model dapat memprediksi target berdasarkan input (*fitur*). Berikut ini *coding* membangun dan melatih model *decisiontree* sebagai berikut:

Gambar 11. Codingan membangun dan melatih model *decisiontree*

3.1.5 Mengekspor Decisiontree dalam bentuk aturan-aturan yang berbasis teks

Mengekspor *decisiontree* yang telah dilatih dalam bentuk aturan-aturan berbasis teks menggunakan fitur-fitur asli sebagai referensi, dan juga menampilkan *decisiontree* tersebut dalam format yang mudah dibaca, dimana setiap aturan menunjukkan kondisi pemisahan pada fitur dan hasil prediksinya.



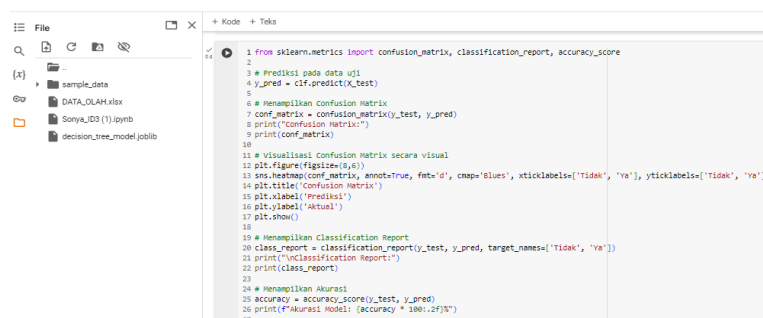
```

1 # Menampilkan decision tree dalam bentuk teks
2 tree_rules = export_text(cif, feature_names=list(X.columns))
3 print(tree_rules)
4
--- JUMLAH PENJUALAN <= 0.50
|--- HARGA <= 0.50
|   |--- class: 1
|   |--- HARGA > 0.50
|       |--- KATEGORI <= 0.50
|           |--- class: 1
|           |--- KATEGORI > 0.50
|               |--- STOK <= 0.50
|                   |--- class: 1
|                   |--- STOK > 0.50
|                       |--- HARGA <= 1.50
|                           |--- STOK <= 1.50
|                               |--- KATEGORI <= 1.50
|                                   |--- class: 1
|                                   |--- KATEGORI > 1.50
|                                       |--- class: 0
|                                       |--- STOK > 1.50
|                                           |--- KATEGORI <= 1.50
|                                               |--- class: 0
|                                               |--- KATEGORI > 1.50
|                                                   |--- class: 1
|--- HARGA > 1.50
|--- class: 1
--- JUMLAH PENJUALAN > 0.50
|--- class: 0
  
```

Gambar 12. Coding dan Hasil Running Pengeksporan Decision Tree

3.1.6 Confusion Matrix

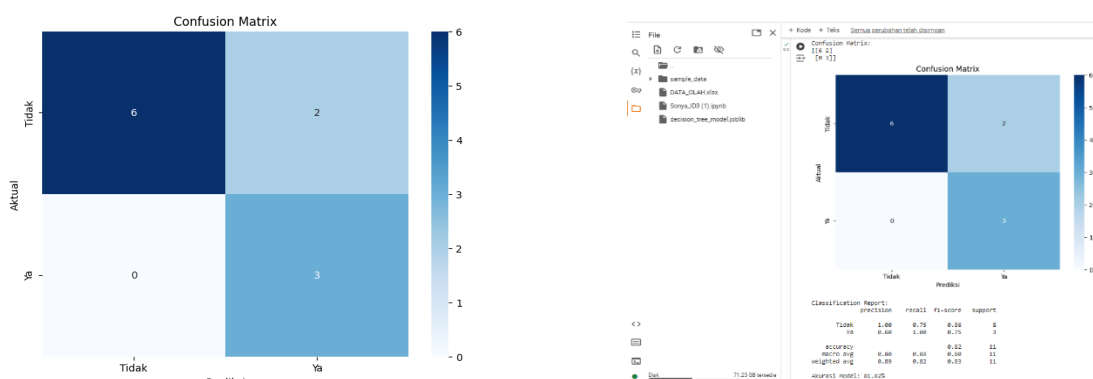
Mengevaluasi kinerja model *decisiontree* yang telah dilatih dengan menggunakan *confusion matrix*, *classification report*, dan *akurasi*. Kinerja dievaluasi berdasarkan prediksi pada *testingset*. Berikut codingnya :



```

1 from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
2
3 # Prediksi pada data uji
4 y_pred = cif.predict(X_test)
5
6 # Menampilkan Confusion Matrix
7 conf_matrix = confusion_matrix(y_test, y_pred)
8 print("Confusion Matrix")
9 print(conf_matrix)
10
11 # Visualisasi Confusion Matrix secara visual
12 plt.figure(figsize=(8,6))
13 sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=['Tidak', 'Ya'], yticklabels=['Tidak', 'Ya'])
14 plt.title('Confusion Matrix')
15 plt.xlabel('Prediksi')
16 plt.ylabel('Aktual')
17 plt.show()
18
19 # Menampilkan Classification Report
20 class_report = classification_report(y_test, y_pred, target_names=['Tidak', 'Ya'])
21 print("Classification Report")
22 print(class_report)
23
24 # Menampilkan Akurasi
25 accuracy = accuracy_score(y_test, y_pred)
26 print(f"Akurasi Model: {accuracy * 100:.2f}%")
27
  
```

Gambar 13 Coding Confusion Matrix



Gambar 14. Hasil Running Confusion Matrix

Interpretasi Hasil

a) **True Positive (TP): 3** (jumlah produk yang benar-benar terlaris dan diprediksi terlaris).

Code ini menghasilkan visualisasi dari *decisiontree* yang memudahkan pemahaman tentang cara model membuat keputusan berdasarkan fitur-fitur input. Dengan menampilkan nama fitur dan kelas serta menggunakan pengisian warna dan bentuk node yang bulat, visualisasi ini menjadi alat yang berguna untuk :

- Menganalisis dan memahami model : memberikan wawasan tentang bagaimana model bekerja dan aturan apa yang dihasilkan.
- Komunikasi hasil : membantu dalam menjelaskan keputusan model kepada audiens non-teknis dengan cara yang lebih intuitif.

```
[x] 1 from sklearn import tree
2 import matplotlib.pyplot as plt
3
4 # Visualisasi Decision Tree dengan lebih detail
5 plt.figure(figsize=(20,10)) # Ukuran gambar diperbesar agar lebih jelas
6
7 # Visualisasi tree dengan menampilkan nama fitur dan nama kelas
8 tree.plot_tree(clf,
9               feature_names=X.columns, # Menampilkan nama fitur
10              class_names=['Tidak', 'Ya'], # Menampilkan nama kelas
11              filled=True, # Mengisi warna berdasarkan probabilitas
12              rounded=True, # Membuat node berbentuk bulat
13              fontsize=10) # Ukuran font
14
15 plt.title("Visualisasi Decision Tree ID3 untuk Prediksi Produk Terlaris")
16 plt.show()
17
```

```

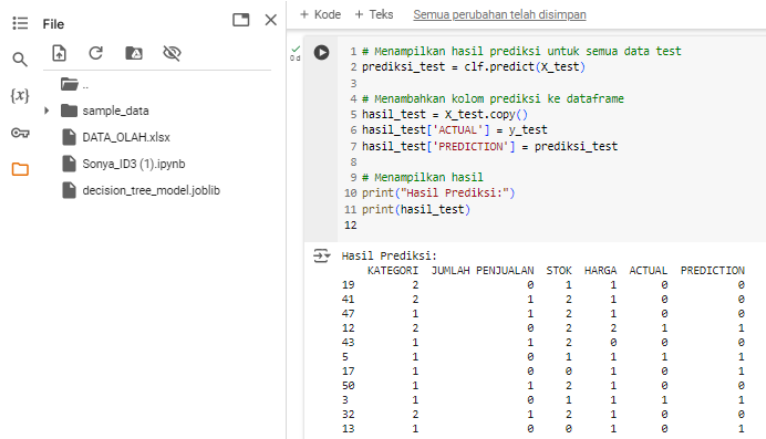
graph TD
    Root["Jumlah Penjualan <= 0.5  
entropy = 1.0  
samples = 41  
value = [20, 21]  
class = Ya"]
    Root -- True --> H1["Harga <= 0.5  
entropy = 0.706  
samples = 26  
value = [5, 21]  
class = Ya"]
    Root -- False --> H1R["entropy = 0.0  
samples = 15  
value = [15, 0]  
class = Tidak"]
    H1 --> K1["Kategori <= 0.5  
entropy = 0.896  
samples = 16  
value = [5, 11]  
class = Ya"]
    H1 --> K1L["entropy = 0.0  
samples = 10  
value = [0, 10]  
class = Ya"]
    K1 --> S1["Stok <= 0.5  
entropy = 0.94  
samples = 14  
value = [5, 9]  
class = Ya"]
    K1 --> S1L["entropy = 0.0  
samples = 2  
value = [0, 2]  
class = Ya"]
    S1 --> H2["Harga <= 1.5  
entropy = 0.961  
samples = 13  
value = [5, 8]  
class = Ya"]
    S1 --> S1R["entropy = 0.0  
samples = 1  
value = [0, 1]  
class = Ya"]
    H2 --> S2["Stok <= 1.5  
entropy = 0.98  
samples = 12  
value = [5, 7]  
class = Ya"]
    H2 --> K2["Kategori <= 1.5  
entropy = 0.954  
samples = 8  
value = [3, 5]  
class = Ya"]
    S2 --> K2L["entropy = 0.0  
samples = 1  
value = [0, 1]  
class = Ya"]
    S2 --> K2R["entropy = 1.0  
samples = 6  
value = [3, 3]  
class = Tidak"]
    K2 --> K2L2["entropy = 0.0  
samples = 2  
value = [0, 2]  
class = Ya"]
    K2 --> K2R2["entropy = 0.0  
samples = 1  
value = [0, 1]  
class = Ya"]
    K2 --> K2R3["entropy = 0.918  
samples = 3  
value = [2, 1]  
class = Tidak"]
    K2 --> K2R4["entropy = 0.0  
samples = 1  
value = [0, 1]  
class = Ya"]

```

3.1.8 Menampilkan hasil prediksi

Kode ini secara keseluruhan menghasilkan tabel yang memungkinkan pengguna untuk membandingkan hasil prediksi model dengan nilai aktual untuk setiap contoh dalam data pengujian. Ini sangat berguna untuk:

- Analisis Hasil: Memungkinkan analisis kesalahan dan evaluasi prediksi model.
 - Pemahaman Model: Memberikan wawasan tentang seberapa baik model dalam membuat prediksi berdasarkan data yang tersedia.
 - Pengujian Model: Memfasilitasi proses pengujian dan validasi model untuk memastikan bahwa model berfungsi dengan baik dalam konteks yang diinginkan.
- Berikut codingnya :



```

1 # Menampilkan hasil prediksi untuk semua data test
2 prediksi_test = clf.predict(X_test)
3
4 # Menambahkan kolom prediksi ke dataframe
5 hasil_test = X_test.copy()
6 hasil_test['ACTUAL'] = y_test
7 hasil_test['PREDICTION'] = prediksi_test
8
9 # Menampilkan hasil
10 print("Hasil Prediksi:")
11 print(hasil_test)
12

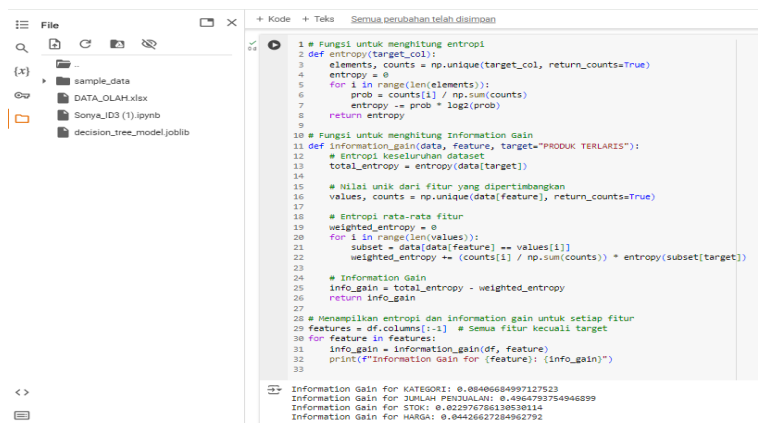
```

	KATEGORI	JUMLAH PENJUALAN	STOK	HARGA	ACTUAL	PREDICTION
19	2	0	1	1	0	0
41	2	1	2	1	0	0
47	1	1	2	1	0	0
12	2	0	2	2	1	1
43	1	1	2	0	0	0
5	1	0	1	1	1	1
17	1	0	0	1	0	1
50	1	1	2	1	0	0
3	1	0	1	1	1	1
32	2	1	2	1	0	0
13	1	0	0	1	0	1

Gambar 17. Code dan Running Hasil Prediksi Data Test

Menghitung *entropy* dan *informasi gain*

Kode ini dirancang untuk menghitung *entropy* dan *information gain* dari setiap fitur dalam dataset dengan tujuan menentukan seberapa baik fitur tersebut dalam memprediksi kelas target, yang dalam konteks ini adalah "PRODUK TERLARIS". Berikut codenya :



```

1 # Fungsi untuk menghitung entropi
2 def entropi(target_col):
3     elements, counts = np.unique(target_col, return_counts=True)
4     entropy = 0
5     for i in range(len(elements)):
6         prob = counts[i] / np.sum(counts)
7         entropy += prob * log2(prob)
8     return entropy
9
10 # Fungsi untuk menghitung Information Gain
11 def information_gain(data, feature, target="PRODUK TERLARIS"):
12     # Entropi keseluruhan dataset
13     total_entropy = entropi(data[target])
14
15     # Nilai unik dari fitur yang dipertimbangkan
16     values, counts = np.unique(data[feature], return_counts=True)
17
18     # Entropi rata-rata fitur
19     weighted_entropy = 0
20     for i in range(len(values)):
21         subset = data[data[feature] == values[i]]
22         weighted_entropy += (counts[i] / np.sum(counts)) * entropi(subset[target])
23
24     # Information Gain
25     info_gain = total_entropy - weighted_entropy
26     return info_gain
27
28 # Menampilkan entropi dan information gain untuk setiap fitur
29 features = df.columns[:-1] # Semua fitur kecuali target
30 for feature in features:
31     info_gain = information_gain(df, feature)
32     print(f"Information gain for {feature}: {info_gain}")
33

```

Information gain for KATEGORI: 0.88486684997127523
Information gain for JUMLAH PENJUALAN: 0.4964793754946899
Information gain for STOK: 0.822976786138538114
Information gain for HARGA: 0.8442682734962792

Gambar 18. Code dan Running Perhitungan Entropy dan Informasi Gain

Kode ini secara keseluruhan berfungsi untuk:

- Menghitung dan mengevaluasi *entropi* dan *information gain* dari setiap fitur dalam dataset, yang merupakan langkah penting dalam pembelajaran mesin, khususnya dalam pemilihan fitur untuk model *Decision Tree*.
- Dengan menggunakan nilai Information Gain, kita dapat mengidentifikasi fitur mana yang paling informatif dan relevan untuk klasifikasi, membantu dalam meningkatkan akurasi model.

3.2 Pembahasan

Proses memprediksi penjualan produk terlaris sangat membutuhkan *decision tree* agar hasil prediksi penjualan produk terlaris yang didapatkan akurat. Dalam melakukan prediksi penjualan produk memerlukan variabel - variabel untuk digunakan dalam memprediksi penjualan produk terlaris di Kreatif Mart Atambua yakni merk produk, stok produk, jumlah penjualan dan harga produk. Berikut ini adalah data produk penjualan di Kreatif Mart Atambua.

Tabel 1. Data Produk Penjualan

Nama Barang	Jumlah Penjualan	Harga	Stok
Telur Ayam Ras	15279	22.000	850
Aqua Botol 600 Ml	7387	3.500	550
Aqua Botol 330 Ml	6058	2.500	70
So Nice Sosis Sapi Pcs 21g	3235	1.000	330
Aqua Botol 1500 Ml	3190	6.500	400
Pasta Coklat Pcs 8g	3077	1.000	165
Le Minerale Btl 600ml	2914	3.500	137
Pasta Keju 8 G	2559	1.000	165
Teh Pucuk Harum Btl 350ml	2185	4.000	70
Chocolatos Wafer Stick 8.5 G	1872	500	70
Le Minerale Btl 330ml	1603	2.500	40
So Nice Sosis Ayam Pcs 21g	1531	1.000	200
Choki Choki Cashew 10 G	1373	13.000	100
Le Minerale Btl 1,5 L	1360	6.500	50
Beng-Beng Regular 25g	1346	2.500	65
Roti Boni 1.000	1148	1.000	2.500
Eko Mie Bumbu Kaldu Ayam Pcs 60 G	1035	1.500	45
Sedaap Goreng Pcs 90g	1035	3.500	55
Sedaap Mie Soto 75 G	1007	3.500	55
Aice Miki-Miki Stick 30 G	1007	2.500	1.200
Gula Pasir Timbang 1 Kg	1003	18.500	30
Floridina Orange Btl 350 Ml	1000	Rp3.500	125
Floridina Cocobit Btl 350ml	963	Rp3.500	100
Nabati Time Break Extra 48g	950	3.000	15
Bear Brand Rtd Milk 189ml	938	11.000	280
Sprite Lemon Lime Btl 250ml	930	4.000	300
Masker Duckbil 1.000	892	1.000	70
Permen Lunak Aneka Rasa Buah	891	1.500	50
Bonafide Gelas 24*220ml	890	20.000	170
Dancow Fortigro Putih 5*27g	861	17.500	80
Ahh Nabati Richeese 5.5 G	850	1.000	43
Siip Nabati Chocolate 5 G	840	10.000	43
Siip Nabati Roasted Corn Pcs 5g	817	11.200	43
Club Gelas 220ml	805	23.000	45
Tobelo Golden Balls 5g	780	25.000	50
Ultra Milk Coklat 125 Ml	712	3.500	85
Superstar Triple Choco 18 G	703	15.000	100
Nabati Time Break Richoco 16g	694	20.600	78
Nabati Rolls Richeese 8 G	673	10.000	60
Ahh Nabati Richoco 5.5 G	672	10.000	70
Nabati Rolls Richoco 8 G	670	500	70
Tugu Buaya Kipas 12g	670	25.000	10
Biskuat Coklat Pcs 15.2 G	655	10.000	52
Masako Sapi Rtg 12*9g	647	10.000	200
Surya Gudang Garam 12	646	28.000	82

Nama Barang	Jumlah Penjualan	Harga	Stok
Ale-Ale Cocopandan Cup 180ml	622	1.000	90
Nabati Siip Richeese 5 G	620	10.000	15
Merries Pants Pcs L1	607	20.000	55
Golden Canon 500	573	20.000	40
Chocolatos Dark 18 G	572	25.000	72
Pusan Permen Kacamata 5.4g	570	1.000	25
Aice Chocolate Stick 40 G	565	3.000	950
Nabati Richoco 50 G	564	2.500	17
Beng-Beng Maxx 32g	562	35.000	80
Gula Piramid Kristal 1 Kg	545	21.000	40
Coklat Bola 500	545	1.500	30
Fullo Coklat Pcs 8 G	536	13.000	32

4. Kesimpulan

Berdasarkan hasil dan pembahasan diatas maka dapat disimpulkan bahwa:

1. Implementasi algoritma Decision Tree menggunakan metode ID3 pada data penjualan produk di Kreatif Mart telah berhasil dilakukan. Algoritma ini memungkinkan pembuatan model klasifikasi yang dapat memprediksi produk-produk terlaris berdasarkan variabel-variabel yang dianalisis, seperti jumlah penjualan, stok, dan harga produk. Pohon keputusan yang dihasilkan memberikan informasi yang jelas tentang faktor-faktor yang mempengaruhi laris tidaknya suatu produk. Dari hasil analisis dengan Decision Tree, ditemukan bahwa produk yang memiliki stok cukup, harga terjangkau, dan jumlah penjualan yang tinggi cenderung diklasifikasikan sebagai produk yang laris. Dengan kata lain, variabel-variabel seperti stok yang memadai dan harga murah atau sedang sangat mempengaruhi keputusan konsumen dalam membeli suatu produk di Kreatif Mart.
2. Penerapan metode Decision Tree juga memberikan kontribusi signifikan dalam meningkatkan efisiensi pengelolaan stok di toko. Dengan memprediksi produk apa saja yang berpotensi terjual habis dalam waktu cepat, toko dapat mengelola persediaan barang secara lebih efektif, sehingga dapat menghindari risiko kekosongan stok dan meminimalisir kelebihan stok pada produk yang tidak laris.

Model pohon keputusan yang terbentuk berdasarkan hasil perhitungan dengan algoritma id3 adalah berdasarkan evaluasi confusion matrix melalui pembagian dataset yang telah dilakukan, diperoleh hasil rata-rata akurasi sebesar 81.82%. Model yang dihasilkan memiliki tingkat akurasi yang cukup baik, dengan menunjukkan prediksi yang sesuai dengan kondisi nyata di toko. Akurasi yang tinggi ini membuktikan bahwa algoritma ID3 cukup efektif untuk diterapkan dalam konteks prediksi penjualan produk di lingkungan ritel.

Mengambil keputusan strategis terkait pengelolaan stok dan penyediaan barang. Dengan demikian, manajemen dapat lebih mudah menentukan produk yang perlu diutamakan untuk pengadaan ulang, serta mengurangi barang-barang yang kurang diminati konsumen. Dengan demikian, algoritma ID3 terbukti efektif dalam membantu memprediksi produk terlaris di Kreatif Mart dan dapat diandalkan sebagai alat bantu dalam pengambilan keputusan manajemen stok.

Daftar Pustaka

- [1] E. Hermawan and R. Sulastri, "Pemberdayaan Masyarakat: Pentingnya Pemenuhan Kebutuhan Dasar Masyarakat," *Distingsi J. Digit. Soc.*, vol. 1, no. 3, pp. 1–6, 2023.
- [2] Hesniati and S. T. Putri, "Jurnal Pengabdian kepada Masyarakat Nusantara (JPkMN) Vol.3 No.2.2 Februari 2023 1830-1840 e-ISSN : 2745 4053 Perancangan," *J. Pengabd. Kpd. Masy. Nusantara*, vol. 3, no. 2, pp. 1830–1840, 2023.
- [3] F. Faturrahman, "Pertambahan jumlah umkm sebagai potensi pad kabupaten sumbawa," *J. accounting, Financ. Audit.*, vol. 3, no. 2, pp. 48–56, 2021.
- [4] C. L. Tjandra and D. S. MintoTorogo, "Fasilitas Wisata Pembuatan Kain Tenun di

- Atambua , NTT,” *J. eDimensi Arsit.*, vol. XI, no. 1, pp. 329–336, 2023.
- [5] S. H. Syahid, P. P. Amtiran, R. F. Makatita, and C. C. Foenay, “FAKTOR-FAKTOR YANG MEMPENGARUHI PERDAGANGAN LINTAS BATAS NEGARA INDONESIA-TIMOR LESTE Factors Affecting Cross-Border Trade Between Indonesia And Timor,” pp. 1239–1250, 2019.
- [6] P. S. Ramadhan and N. Safitri, “Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang,” *J. saintikom*, vol. 18, no. 1, pp. 55–61, 2019.
- [7] P. Mai, S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, “IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA APRIORI DALAM MENENTUKAN PERSEDIAAN BARANG (STUDI KASUS : TOKO SINAR HARAHAP),” *J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 12, no. 2, pp. 51–61, 2022.
- [8] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi produk terlaris,” *J. Ilm. ilmu Komput.*, vol. 7, no. 2, p. 217, 2021.
- [9] F. Zahra, M. A. Ridla, and N. Azise, “Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap),” *JUSTIFY J. Sist. Inf. Ibrahimy*, vol. 3, no. 1, pp. 55–65, 2024, doi: 10.35316/justify.v3i1.5335.
- [10] A. C. Darmawan, “Pengembang Aplikasi Berbasis Web dengan Python Flask untuk Klasifikasi Data Menggunakan Metode Decision Tree C4.5,” *J. Pendiidikan Konseling*, vol. 4, no. 5, pp. 5351–5362, 2022.
- [11] G. Gunawan, A. C. Fauzan, and H. Harliana, “Implementasi Algoritma Decision Tree Iterative Dichotomiser 3 (ID3) untuk Prediksi Keberhasilan Pengobatan Penyakit Kutil Menggunakan Cryotherapy,” *J. Bumigora Inf. Technol.*, vol. 4, no. 1, pp. 73–82, 2022, doi: 10.30812/bite.v4i1.1949.
- [12] S. Bahri and A. Lubis, “Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League,” *J. Sintaksis*, vol. 2, no. 1, pp. 63–70, 2020.
- [13] Fairuzabadi, M., Permana, A. A., Istiono, W., Pomalingo, S., Prasetyo, A. B., Kelen, Y. P. K., ... & Fernando, E. (2024). SISTEM PENDUKUNG KEPUTUSAN: KONSEP, METODE DAN IMPLEMENTASI. Get Press Indonesia.
- [14] Kelen, Yoseph PK, et al. "Decision support system for the selection of new prospective students using the simple additive weighted (SAW) method." *AIP Conference Proceedings*. Vol. 2798. No. 1. AIP Publishing, 2023.



ZONAsi: Jurnal Sistem Informasi

is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)