IMPLEMENTASI ALGORITMA DECISION TREE C4.5 UNTUK PREDIKSI EFEKTIVITAS DISKON TERHADAP PENDAPATAN HARIAN RUMAH BILLIARD

NAMA : ALDIAN ALIFEN

NIM : 20220801015

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dalam era persaingan bisnis yang semakin ketat, industri hiburan dan rekreasi seperti rumah billiard menghadapi tantangan dalam mengoptimalkan strategi pemasaran dan promosi. Pemberian diskon merupakan salah satu strategi yang umum digunakan untuk menarik pelanggan dan meningkatkan pendapatan. Namun, tanpa analisis yang mendalam, pemberian diskon dapat menjadi tidak efektif bahkan kontraproduktif, dimana diskon yang diberikan justru menurunkan margin keuntungan tanpa memberikan dampak signifikan terhadap peningkatan total pendapatan.

Klik Billiard & cafe saat ini memberikan diskon tanpa dasar analisis yang kuat tentang dampaknya terhadap total pendapatan harian dan komposisi pendapatannya. Permasalahan utama yang dihadapi adalah kurangnya pemahaman tentang korelasi antara total diskon yang diberikan dengan efektivitas pendapatan harian. Manajemen kesulitan menjawab pertanyaan kritis seperti: "Berapa budget diskon optimal untuk hari tertentu agar pendapatan total efektif?" dan "Bagaimana pola diskon mempengaruhi komposisi pendapatan antara sewa meja dan penjualan makanan/minuman?"

Data mining dengan algoritma Decision Tree C4.5 telah terbukti efektif dalam berbagai kasus klasifikasi dan prediksi bisnis. Algoritma ini memiliki keunggulan dalam menghasilkan aturan yang mudah diinterpretasi dan dapat diterapkan langsung dalam pengambilan keputusan bisnis. Berdasarkan systematic literature review yang dilakukan, algoritma C4.5 telah berhasil diimplementasikan dalam berbagai konteks bisnis dengan tingkat akurasi yang baik.

Penelitian ini bertujuan untuk mengembangkan model prediksi berbasis algoritma Decision Tree C4.5 yang dapat mengklasifikasikan efektivitas hari berdasarkan pola pemberian diskon. Model ini akan menganalisis hubungan antara total diskon harian, persentase diskon terhadap pendapatan, komposisi pendapatan (sewa meja vs F&B), dan hari dalam minggu untuk menghasilkan rekomendasi yang data-driven. Dengan sistem prediksi ini, diharapkan manajemen dapat mengambil keputusan strategis dalam menetapkan budget dan strategi diskon harian yang lebih optimal, sehingga dapat meningkatkan profitabilitas dan efisiensi promosi.

1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas, penulis mengidentifikasi beberapa masalah yang akan dijadikan bahan penelitian sebagai berikut:

1. Pemberian diskon di Klik Billiard & cafe dilakukan tanpa analisis mendalam tentang dampaknya terhadap total pendapatan harian dan komposisi pendapatan.
2. Belum ada pemahaman yang jelas tentang korelasi antara total diskon harian dengan efektivitas pendapatan, serta bagaimana pola diskon mempengaruhi komposisi pendapatan antara sewa meja dan penjualan F&B.
3. Diperlukan metode klasifikasi yang tepat untuk memprediksi kategori efektivitas hari berdasarkan pola pemberian diskon, sehingga dapat memberikan rekomendasi budget diskon yang optimal untuk setiap hari.

1.3 Tujuan Tugas Akhir

Adapun tujuan yang ingin dicapai dari penelitian yang dilakukan adalah sebagai berikut:

1. Membangun model klasifikasi menggunakan algoritma Decision Tree C4.5 yang dapat memprediksi kategori efektivitas hari (Efektif, Cukup Efektif, Tidak Efektif) berdasarkan Total Diskon, Persentase Diskon, Rasio Pendapatan Sewa Meja, dan Hari dalam minggu.
2. Mengidentifikasi kombinasi faktor yang paling berpengaruh terhadap pendapatan tinggi dan pola hubungan antara pemberian diskon dengan proporsi pendapatan dari sewa meja dan penjualan F&B.
3. Memberikan rekomendasi praktis berbasis data mengenai besaran budget diskon harian yang optimal untuk meningkatkan probabilitas suatu hari masuk dalam kategori "Efektif" serta meningkatkan pendapatan harian rata-rata pada periode sepi.

1.4 Manfaat Tugas Akhir

Adapun manfaat dari tugas akhir ini adalah sebagai berikut:

1. Bagi Perguruan Tinggi

1. Tugas akhir ini menjadi salah satu sarana promosi bagi Universitas Esa Unggul, memperkenalkan kompetensi mahasiswa Teknik Informatika kepada dunia industri, terutama bagi perusahaan yang membutuhkan lulusan dengan kemampuan analisis data dan data mining.
2. Dapat membuka peluang kemitraan strategis antara universitas dan sektor industri, sehingga tercipta kolaborasi yang berkelanjutan dalam bidang pengembangan tenaga kerja dan penelitian terapan.

3. Diharapkan dapat menjadi referensi bagi mahasiswa dan peneliti masa depan yang tertarik dengan implementasi algoritma klasifikasi dalam konteks bisnis dan pengambilan keputusan strategis.

2. Bagi Peneliti

1. Mengembangkan keterampilan teknis dalam pengolahan data, implementasi algoritma C4.5, dan evaluasi model machine learning, serta kemampuan analisis bisnis sebagai bagian dari pengembangan diri.
2. Memberikan pengalaman nyata dalam menerapkan ilmu yang telah diperoleh selama kuliah pada penelitian yang aplikatif dan dapat memberikan solusi terhadap permasalahan bisnis nyata.

3. Bagi Bisnis

1. Membantu manajemen Klik Billiard & cafe dalam membuat keputusan berbasis data terkait strategi pemberian diskon yang lebih efektif dan efisien.
2. Memberikan wawasan tentang pola hubungan antara diskon dengan komposisi pendapatan, yang dapat digunakan untuk optimasi strategi promosi dan peningkatan profitabilitas.
3. Menyediakan model prediksi yang dapat digunakan untuk perencanaan budget diskon harian dan evaluasi efektivitas strategi promosi yang sedang berjalan.

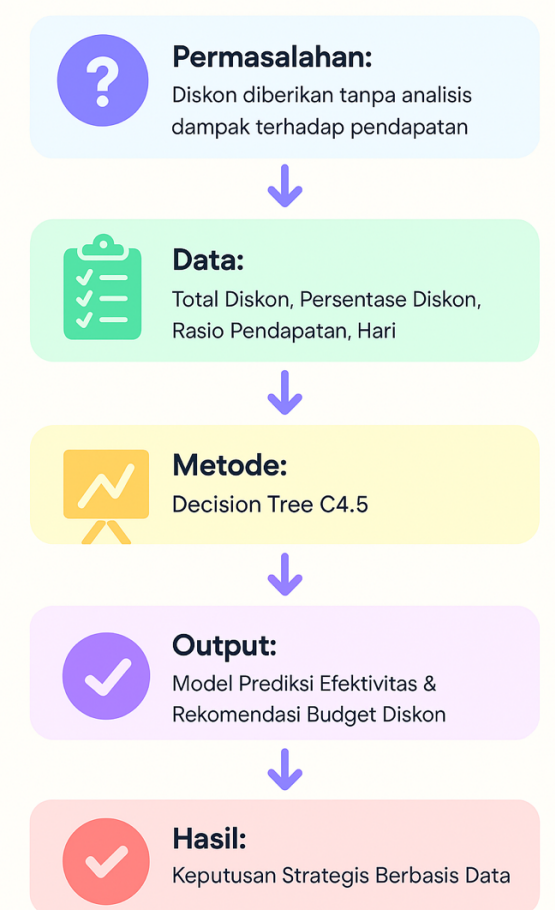
1.5 Lingkup Tugas Akhir

Ruang lingkup tugas akhir ini sebagaimana berikut:

1. Penelitian ini berfokus pada data tutup hari (closing) Klik Billiard & cafe periode Agustus 2024 – September 2025
2. Penelitian ini menggunakan algoritma Decision Tree C4.5 untuk mengklasifikasikan efektivitas pendapatan harian berdasarkan pola pemberian diskon.
3. Data mentah yang tersedia meliputi: Total Pendapatan Harian, Total Diskon Harian, Total Penjualan Sewa Meja Harian, Total Penjualan Makanan dan Minuman Harian, serta Total Pembayaran Tunai dan Non-Tunai.
4. Variabel input yang digunakan dalam model meliputi: Hari dalam minggu, Total Diskon Harian, Persentase Diskon Harian (variabel turunan dihitung dari Total Diskon/Total Pendapatan × 100%), dan Rasio Pendapatan Sewa Meja (variabel turunan dihitung dari Total Penjualan Sewa Meja/Total Pendapatan × 100%).
5. Variabel target adalah Kategori Efektivitas Pendapatan Harian yang dikategorikan menjadi: Tidak Efektif, Cukup Efektif, dan Sangat Efektif berdasarkan perhitungan kuartil dari Total Pendapatan Harian.

6. Penelitian ini tidak mencakup analisis pada level transaksi individu, detail jenis diskon spesifik, atau integrasi langsung dengan sistem Point of Sale (POS).

1.6 Kerangka Berpikir



1.7 Sistematika Penulisan Tugas Akhir

Sistematika penulisan Tugas Akhir ini disusun untuk memberikan pengaturan dan struktur yang jelas agar Tugas Akhir dapat disajikan dengan baik kepada pembaca. Umumnya, sistematika penulisan Tugas Akhir terdiri dari beberapa bagian yang saling terhubung, sebagai berikut:

BAB I PENDAHULUAN

Bab ini mencakup latar belakang penelitian, identifikasi masalah, tujuan yang ingin dicapai, manfaat yang diharapkan, lingkup penelitian, kerangka berpikir, serta sistematika penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini memaparkan penelitian-penelitian terdahulu yang relevan (systematic literature review) dan dasar teori yang mendukung penelitian pada tugas akhir ini, termasuk teori tentang data mining, algoritma Decision Tree C4.5, dan konsep-konsep terkait analisis bisnis.

BAB III METODE PENELITIAN

Bab ini menguraikan metodologi dan metode yang digunakan sebagai pedoman dalam menyelesaikan penulisan tugas akhir, meliputi pendekatan penelitian, tahapan penelitian, teknik pengumpulan data, preprocessing, implementasi algoritma C4.5, dan metode evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil yang diperoleh dari penelitian serta pembahasan mendalam terkait analisis dan interpretasi data yang telah dikumpulkan, visualisasi pohon keputusan, evaluasi model, dan rekomendasi bisnis.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan dan saran yang dapat bermanfaat untuk penelitian selanjutnya serta implementasi praktis di lapangan.

BAB 2

TINJAUAN PUSTAKA

2.1 Systematic Literature Review (SLR)

2.1.1 Pendahuluan SLR

Systematic Literature Review (SLR) telah dilakukan dengan tujuan menganalisis hasil-hasil penelitian sebelumnya yang relevan dengan penerapan algoritma Decision Tree C4.5 dalam konteks klasifikasi dan prediksi bisnis, khususnya terkait analisis penjualan dan strategi promosi. Tujuan utama dari SLR ini adalah untuk mengidentifikasi kontribusi signifikan dari studi-studi terdahulu yang berkaitan dengan implementasi algoritma C4.5, data mining untuk bisnis, serta manajemen strategi diskon dan promosi.

2.1.2 Metodologi SLR

Proses SLR mengikuti langkah-langkah sistematis yang konsisten dengan ruang lingkup penelitian ini:

1. **Identifikasi:** Penelusuran artikel yang relevan menggunakan kata kunci seperti "Decision Tree C4.5", "klasifikasi penjualan", "prediksi bisnis", "data mining", dan "strategi diskon".
2. **Penyaringan:** Evaluasi artikel berdasarkan relevansi topik, fokus pada penelitian yang menerapkan C4.5 untuk klasifikasi atau prediksi dalam konteks bisnis.
3. **Kelayakan:** Kriteria mencakup relevansi terhadap tema C4.5, data mining bisnis, publikasi dalam rentang waktu relevan, dan metodologi yang jelas.
4. **Inklusi Akhir:** Lima artikel terpilih untuk analisis sistematis lebih lanjut.

2.1.3 Hasil SLR

Tabel 2.1 Hasil Systematic Literature Review

(Ringkasan dari 5 studi yang dianalisis)

1. Penerapan Data Mining untuk Klasifikasi Penjualan Barang Terlaris Menggunakan Metode Decision Tree C4.5

Hasil: Menunjukkan penerapan langsung C4.5 untuk identifikasi produk terlaris, namun kurang detail dalam metrik kinerja.

2. Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 untuk Prediksi Kelayakan Kredit

Hasil: Mencapai akurasi 83,67% dengan pendekatan metodis melibatkan seleksi fitur.

3. Implementasi Algoritma C4.5 untuk Klasifikasi Penjualan Barang di Swalayan

Hasil: Melaporkan akurasi 100% (mengindikasikan potensi overfitting).

4. Penerapan Data Mining untuk Klasifikasi Data Penjualan Sembako Terlaris dengan Algoritma C4.5

Hasil: Mencapai akurasi 85% dengan aplikasi langsung untuk optimasi stok dan pemasaran.

5. Penerapan Algoritma C4.5 untuk Memprediksi Penjualan Barang pada PT Prima Niaga Indomas

Hasil: Fokus pada pengembangan model prediktif dengan variabel eksternal.

2.1.4 Analisis Hasil SLR

Analisis terhadap kelima literatur tersebut menghasilkan beberapa temuan krusial:

1. Algoritma C4.5 secara konsisten dipilih untuk klasifikasi dan prediksi dalam konteks bisnis, mengonfirmasi efektivitasnya.
2. Tingkat akurasi bervariasi (83,67% - 100%), dengan akurasi sangat tinggi mengindikasikan potensi overfitting.
3. Sebagian besar penelitian belum mengintegrasikan variabel diskon atau strategi promosi dalam analisis penjualan.
4. Fokus dominan pada klasifikasi produk terlaris, bukan pada efektivitas strategi bisnis seperti pemberian diskon.
5. Kurangnya penelitian yang menghubungkan pola diskon dengan komposisi pendapatan (misalnya: sewa vs F&B).

2.1.5 Analisis Kesenjangan Penelitian (Research Gap)

Tabel 2.2 Analisis Gap Penelitian

Aspek	Penelitian Sebelumnya	Kontribusi Penelitian Ini
Variabel Diskon	Belum menjadi fokus utama dalam klasifikasi	Menggunakan Total Diskon dan Persentase Diskon sebagai fitur utama
Komposisi Pendapatan	Tidak dianalisis hubungannya dengan strategi promosi	Menganalisis Rasio Pendapatan Sewa Meja vs F&B terhadap efektivitas diskon
Konteks Bisnis	Fokus pada retail umum, supermarket	Spesifik untuk bisnis rumah billiard dengan karakteristik unik
Output Model	Klasifikasi produk terlaris	Prediksi efektivitas hari dan rekomendasi budget diskon optimal

2.1.6 Kesimpulan Hasil SLR

Berdasarkan hasil tinjauan sistematis, dapat disimpulkan bahwa algoritma C4.5 telah terbukti efektif untuk klasifikasi bisnis dengan tingkat akurasi yang baik. Namun, terdapat kesenjangan signifikan dalam penelitian yang mengintegrasikan analisis diskon dengan prediksi efektivitas pendapatan. Penelitian ini menawarkan kontribusi baru dengan mengembangkan model yang secara spesifik memprediksi efektivitas strategi diskon terhadap pendapatan harian, dengan mempertimbangkan komposisi pendapatan dan pola hari dalam minggu.

2.2 Data Mining

Data mining adalah proses ekstraksi pola yang bermakna dan pengetahuan yang berguna dari kumpulan data yang besar. Dalam konteks bisnis, data mining digunakan untuk mengidentifikasi tren, pola, dan hubungan tersembunyi dalam data yang dapat mendukung pengambilan keputusan strategis. Teknik data mining mencakup klasifikasi, clustering, asosiasi, dan prediksi. Penelitian ini fokus pada teknik klasifikasi menggunakan algoritma Decision Tree C4.5 untuk memprediksi kategori efektivitas pendapatan harian berdasarkan pola pemberian diskon.

2.3 Algoritma Decision Tree C4.5

Decision Tree C4.5 adalah algoritma klasifikasi yang dikembangkan oleh Ross Quinlan sebagai perbaikan dari algoritma ID3. C4.5 menggunakan konsep information gain untuk memilih atribut yang paling baik dalam memisahkan data. Keunggulan utama C4.5 meliputi:

1. Mampu menangani data numerik dan kategorikal
2. Dapat menangani missing values
3. Menghasilkan aturan yang mudah diinterpretasi
4. Memiliki mekanisme pruning untuk menghindari overfitting

Algoritma C4.5 membangun pohon keputusan dengan cara rekursif, memilih atribut dengan information gain tertinggi sebagai node pemisah. Proses ini berlanjut hingga semua data terklasifikasi atau memenuhi kriteria stopping.

2.4 Konsep Information Gain dan Entropy

Entropy adalah ukuran ketidakpastian atau keacakan dalam data. Dalam konteks klasifikasi, entropy mengukur seberapa campur data dalam suatu set. Information gain mengukur pengurangan entropy setelah dataset dipecah berdasarkan suatu atribut. Atribut dengan information gain tertinggi dipilih sebagai node pemisah dalam pohon keputusan.

Formula Entropy:

$$Entropy(S) = -\Sigma\ pi \times \log_2(pi)$$

dimana pi adalah proporsi sampel kelas i dalam set S

Formula Information Gain:

$$Gain(S,A) = Entropy(S) - \Sigma\ (|Sv|/|S|) \times Entropy(Sv)$$

dimana A adalah atribut, Sv adalah subset S dengan nilai v untuk atribut A

2.5 Evaluasi Model Machine Learning

Evaluasi model klasifikasi menggunakan beberapa metrik standar:

1. **Confusion Matrix:** Matriks yang menampilkan prediksi benar dan salah untuk setiap kelas.
2. **Akurasi:** Persentase prediksi yang benar dari total prediksi. Formula: (TP+TN)/(TP+TN+FP+FN)
3. **Presisi:** Proporsi prediksi positif yang benar. Formula: TP/(TP+FP)
4. **Recall:** Proporsi data positif aktual yang teridentifikasi. Formula: TP/(TP+FN)

5. **F1-Score:** Harmonic mean dari presisi dan recall. Formula: 2×(Precision×Recall)/(Precision+Recall)

2.6 Preprocessing Data

Preprocessing adalah tahap penting dalam data mining yang meliputi:

1. **Data Cleaning:** Menangani missing values, duplikasi, dan outlier
2. **Feature Engineering:** Membuat variabel turunan dari data mentah
3. **Data Transformation:** Normalisasi atau standarisasi data numerik
4. **Encoding:** Mengubah data kategorikal menjadi numerik
5. **Data Splitting:** Membagi data menjadi training dan testing set

2.7 Teknologi Implementasi Sistem

2.7.1 Python

Python adalah bahasa pemrograman tingkat tinggi yang populer untuk data science dan machine learning. Python menyediakan ekosistem library yang kaya untuk analisis data, visualisasi, dan implementasi algoritma machine learning.

2.7.2 Scikit-learn

Scikit-learn adalah library Python untuk machine learning yang menyediakan implementasi berbagai algoritma klasifikasi, termasuk Decision Tree. Library ini juga menyediakan tools untuk preprocessing data, evaluasi model, dan cross-validation.

2.7.3 Pandas

Pandas adalah library Python untuk manipulasi dan analisis data. Pandas menyediakan struktur data DataFrame yang memudahkan operasi data seperti filtering, grouping, dan aggregation.

2.7.4 Matplotlib dan Seaborn

Matplotlib dan Seaborn adalah library visualisasi data di Python. Matplotlib menyediakan kontrol tingkat rendah untuk plotting, sedangkan Seaborn menyediakan interface tingkat tinggi untuk visualisasi statistik yang menarik.

2.7.5 Jupyter Notebook

Jupyter Notebook adalah aplikasi web open-source yang memungkinkan pembuatan dan berbagi dokumen yang berisi kode, visualisasi, dan teks naratif. Sangat berguna untuk eksplorasi data dan dokumentasi proses analisis.

BAB 3

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode kuantitatif dengan pendekatan eksperimental. Penelitian ini menggunakan data historis transaksi harian Rumah Billiard untuk membangun model prediksi berbasis algoritma Decision Tree C4.5. Pendekatan ini dipilih karena memungkinkan pengukuran objektif terhadap performa model dan memberikan hasil yang dapat divalidasi secara statistik.

3.2 Rencana Penelitian

Penelitian ini dilaksanakan selama 4 bulan dengan tahapan yang terstruktur. Rincian rencana kegiatan penelitian disajikan pada Tabel 3.1:

Tabel 3.1 Timeline Penelitian

Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4
Studi Literatur & Pengumpulan Data	✓			
Preprocessing & Feature Engineering	✓	✓		
Pembangunan Model C4.5		✓	✓	
Evaluasi & Tuning Model			✓	
Dokumentasi & Penyusunan Laporan			✓	✓

3.3 Tahapan Penelitian

Penelitian ini dilaksanakan melalui beberapa tahapan sistematis:

Tahap 1: Identifikasi Masalah dan Studi Literatur

- Identifikasi permasalahan bisnis terkait strategi diskon
- Studi literatur tentang algoritma C4.5 dan aplikasinya
- Analisis penelitian terdahulu (SLR)

Tahap 2: Pengumpulan Data

- Pengumpulan data tutup hari dari sistem existing
- Validasi kelengkapan dan konsistensi data
- Dokumentasi format dan struktur data

Tahap 3: Preprocessing Data

- Data cleaning (handling missing values, outliers)
- Feature engineering (perhitungan variabel turunan)
- Kategorisasi target variable (efektivitas harian)
- Encoding variabel kategorikal

Tahap 4: Pembangunan Model

- Splitting data (training 80%, testing 20%)
- Implementasi algoritma C4.5
- Training model dengan data training
- Hyperparameter tuning

Tahap 5: Evaluasi Model

- Testing model dengan data testing
- Evaluasi menggunakan confusion matrix
- Perhitungan metrik (akurasi, presisi, recall, F1-score)
- Cross-validation untuk validasi robustness

Tahap 6: Interpretasi dan Rekomendasi

- Visualisasi pohon keputusan
- Ekstraksi aturan bisnis
- Penyusunan rekomendasi budget diskon
- Dokumentasi hasil penelitian

3.4 Lokasi dan Objek Penelitian

Lokasi Penelitian:

Rumah Billiard [Nama Tempat Billiard], [Alamat Lengkap]

Objek Penelitian:

Objek penelitian ini adalah data transaksi harian Rumah Billiard yang mencakup informasi tentang total pendapatan, total diskon, pendapatan sewa meja, pendapatan F&B, dan metode pembayaran. Data ini akan digunakan untuk membangun model prediksi efektivitas strategi diskon terhadap pendapatan harian.

3.5 Teknik Pengumpulan Data

3.5.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data tutup hari (closing) dari sistem kasir Rumah Billiard periode [Januari - Desember 2024]. Data ini mencakup catatan harian yang komprehensif tentang operasional bisnis.

3.5.2 Variabel Data

Data Mentah yang Dikumpulkan:

- Tanggal transaksi
- Total Pendapatan Harian (Rp)
- Total Diskon Harian (Rp)
- Total Penjualan Sewa Meja Harian (Rp)
- Total Penjualan Makanan dan Minuman Harian (Rp)
- Total Pembayaran Tunai (Rp)
- Total Pembayaran Non-Tunai (Rp)

3.5.3 Kriteria Data

1. Data lengkap tanpa missing values kritis
2. Periode minimal 6-12 bulan untuk menangkap pola musiman
3. Data tervalidasi dan konsisten
4. Jumlah minimal 200-300 record untuk training yang memadai

3.6 Preprocessing Data

3.6.1 Data Cleaning

1. Identifikasi dan handling missing values
2. Deteksi dan treatment outliers menggunakan IQR method
3. Validasi konsistensi data (misalnya: Total Pendapatan = Sewa Meja + F&B)
4. Penghapusan duplikasi data

3.6.2 Feature Engineering

Pembuatan variabel turunan dari data mentah:

1. Persentase Diskon Harian:
`= (Total Diskon / Total Pendapatan) × 100%`

2. Rasio Pendapatan Sewa Meja:
`= (Total Sewa Meja / Total Pendapatan) × 100%`

3. Rasio Pendapatan F&B:
`= (Total F&B / Total Pendapatan) × 100%`

4. Hari dalam Minggu:
`= Ekstrak dari Tanggal (Senin, Selasa, ..., Minggu)`

3.6.3 Kategorisasi Target Variable

Kategori Efektivitas Pendapatan Harian berdasarkan kuartil:

Tidak Efektif: Total Pendapatan < Q1 (Kuartil 1)
Cukup Efektif: Q1 ≤ Total Pendapatan ≤ Q3
Sangat Efektif: Total Pendapatan > Q3 (Kuartil 3)

3.6.4 Encoding

- Label Encoding untuk variabel Hari (Senin=0, Selasa=1, ..., Minggu=6)
- One-Hot Encoding jika diperlukan untuk kategori dengan lebih dari 2 nilai

3.7 Implementasi Algoritma C4.5

3.7.1 Pembagian Data

Dataset dibagi menjadi data training (80%) dan data testing (20%) menggunakan stratified sampling untuk memastikan distribusi kelas yang seimbang di kedua set.

3.7.2 Parameter Model

Parameter utama yang akan di-tuning:

- **max_depth:** Kedalaman maksimal pohon (untuk menghindari overfitting)
- **min_samples_split:** Jumlah minimal sampel untuk split node
- **min_samples_leaf:** Jumlah minimal sampel di leaf node
- **criterion:** Entropy atau Gini index

3.7.3 Training Model

Model Decision Tree C4.5 diimplementasikan menggunakan library scikit-learn dengan class DecisionTreeClassifier. Training dilakukan dengan data training yang telah dipreprocessing, menggunakan criterion='entropy' untuk information gain.

3.8 Evaluasi Model

3.8.1 Metrik Evaluasi

1. Confusion Matrix
Matriks yang menampilkan True Positive, True Negative, False Positive, dan False Negative untuk setiap kelas.

2. Akurasi
`Akurasi = (TP + TN) / (TP + TN + FP + FN)`
Target: ≥ 80%

3. Presisi
`Presisi = TP / (TP + FP)`

4. Recall
`Recall = TP / (TP + FN)`

5. F1-Score
`F1 = 2 × (Presisi × Recall) / (Presisi + Recall)`

3.8.2 Cross-Validation

K-Fold Cross-Validation dengan k=5 akan dilakukan untuk memvalidasi stabilitas dan robustness model. Metrik evaluasi akan dihitung untuk setiap fold dan dirata-rata untuk mendapatkan estimasi performa yang lebih reliable.

3.8.3 Analisis Feature Importance

Analisis feature importance akan dilakukan untuk mengidentifikasi variabel mana yang paling berpengaruh dalam prediksi efektivitas harian. Hasil ini akan divisualisasikan dan digunakan untuk menyusun rekomendasi bisnis.

3.9 Visualisasi dan Interpretasi

3.9.1 Visualisasi Pohon Keputusan

Pohon keputusan akan divisualisasikan menggunakan library graphviz untuk memudahkan interpretasi aturan klasifikasi. Visualisasi ini akan menunjukkan path keputusan dan splitting criteria di setiap node.

3.9.2 Ekstraksi Aturan Bisnis

Aturan klasifikasi akan diekstrak dari pohon keputusan dan diterjemahkan ke dalam rekomendasi bisnis yang actionable. Contoh: "Jika Hari = Weekend DAN Persentase Diskon > 10% MAKA Efektivitas = Sangat Efektif"

3.9.3 Rekomendasi Budget Diskon

Berdasarkan hasil model, akan disusun rekomendasi praktis mengenai budget diskon optimal untuk setiap hari dalam minggu, dengan mempertimbangkan pola komposisi pendapatan dan probabilitas mencapai kategori "Sangat Efektif".

3.10 Tools dan Teknologi

Tabel 3.2 Teknologi yang Digunakan

Komponen	Teknologi	Fungsi
Bahasa Pemrograman	Python 3.8+	Bahasa utama untuk implementasi
Machine Learning	Scikit-learn	Implementasi algoritma C4.5
Data Manipulation	Pandas, NumPy	Preprocessing dan manipulasi data
Visualisasi	Matplotlib, Seaborn	Visualisasi data dan hasil
Decision Tree Viz	Graphviz	Visualisasi pohon keputusan
Development Environment	Jupyter Notebook	Eksplorasi data dan dokumentasi