

# Analisis Potensi Customer Churn Menggunakan Algoritma Decision Tree (C4.5) Pada Indibiz Telkom Regional I

*Analysis of Customer Churn Potential Using the Decision Tree (C4.5) Algorithm at Indibiz Telkom Regional I*

Nadhilah Syafitri\*<sup>1</sup>, Rizaldy Khair<sup>2</sup>(\* corespondent author)

<sup>1,2</sup>Program Studi Sistem Informasi, Universitas Muhammadiyah Sumatera Utara

E-mail: <sup>1</sup>nadhilasyafitri13@gmail.com, <sup>2</sup>rizaldykhair@umsu.ac.id

## Abstrak

Penelitian ini bertujuan untuk menganalisis potensi customer churn pada pelanggan bisnis Indibiz Telkom Regional I dengan menerapkan algoritma Decision Tree C4.5. Masalah utama yang dihadapi perusahaan adalah tingginya risiko kehilangan pelanggan bisnis yang berdampak pada stabilitas pendapatan. Dalam penelitian ini, dikembangkan sebuah sistem klasifikasi otomatis yang memanfaatkan atribut perilaku pelanggan seperti payment status, total ticket, dan total complain. Metodologi penelitian meliputi tahap pengumpulan data, preprocessing, penghitungan nilai entropy dan information gain, serta pembentukan pohon keputusan. Hasil penelitian mengungkapkan bahwa atribut total ticket memiliki nilai gain ratio tertinggi, yang menunjukkan bahwa frekuensi gangguan layanan merupakan faktor paling dominan dalam memicu churn. Pengujian sistem berbasis web yang dibangun menunjukkan tingkat akurasi sebesar 68% dalam mengklasifikasikan pelanggan ke dalam kategori churn dan tidak churn. Implementasi algoritma C4.5 ini terbukti efektif dalam memetakan pola perilaku pelanggan dan dapat digunakan sebagai instrumen pendukung keputusan bagi manajemen untuk menentukan strategi retensi pelanggan secara lebih tepat sasaran.

**Kata Kunci:** Customer Churn, Decision Tree C4.5, Data Mining.

## Abstract

This study analyzes the potential for customer churn among Indibiz Telkom Regional I business customers by implementing the Decision Tree C4.5 algorithm. The primary issue addressed is the high risk of losing business customers, which impacts revenue stability. This research develops an automated classification system utilizing customer behavior attributes such as payment status, total tickets, and total complaints. The research methodology includes data collection, preprocessing, calculating entropy and information gain, and constructing a decision tree. The results reveal that the "total ticket" attribute has the highest gain ratio, indicating that the frequency of service disruptions is the most dominant factor in triggering churn. Testing of the developed web-based system demonstrated an accuracy rate of 68% in classifying customers into churn and non-churn categories. The implementation of the C4.5 algorithm proves effective in mapping customer behavior patterns and serves as a decision support instrument for management to determine more targeted customer retention strategies.

**Keywords:** Customer Churn, Decision Tree C4.5, Data Mining

## 1. PENDAHULUAN

Perkembangan teknologi informasi dan digitalisasi layanan telah mendorong perusahaan penyedia layanan berbasis digital untuk semakin berfokus pada kepuasan dan loyalitas pelanggan. Di era transformasi digital saat ini, layanan internet dan solusi bisnis bukan lagi sekadar kebutuhan tambahan, melainkan infrastruktur utama bagi pelaku usaha. Hal ini menyebabkan persaingan antar *Internet Service Provider* (ISP) semakin tajam. Perusahaan tidak hanya dituntut untuk melakukan akuisisi atau memperoleh pelanggan baru melalui kampanye pemasaran yang masif, tetapi juga harus memiliki strategi yang kuat untuk mempertahankan (*retention*) pelanggan yang sudah ada.

Salah satu tantangan utama yang dihadapi oleh perusahaan telekomunikasi besar seperti Telkom Indonesia melalui unit bisnis Indibiz adalah fenomena customer churn. Churn didefinisikan sebagai kondisi di mana pelanggan memutuskan untuk berhenti menggunakan layanan atau tidak melanjutkan kontrak langganan dalam periode tertentu [2]. Dalam segmen B2B (Business to Business) seperti Indibiz, kehilangan satu pelanggan bisnis memiliki dampak finansial yang jauh lebih besar dibandingkan pelanggan ritel, mengingat nilai kontrak dan potensi kerja sama jangka panjang yang hilang.

Indibiz Telkom Regional I sebagai salah satu wilayah operasional strategis memiliki tantangan unik dalam melayani karakter pelanggan bisnis di wilayah Sumatera. Selama ini, pemantauan pelanggan yang berpotensi berhenti berlangganan masih bersifat reaktif, yaitu baru diketahui saat pelanggan sudah menunggak atau mengajukan keputusan. Dengan memanfaatkan atribut seperti status pembayaran, jumlah tiket gangguan (*total ticket*), dan frekuensi keluhan (*total complain*), penelitian ini bermaksud membangun sistem prediktif. Melalui algoritma C4.5, diharapkan faktor dominan seperti banyaknya gangguan layanan dapat dipetakan secara sistematis untuk mendukung pengambilan keputusan manajemen dalam memberikan penanganan khusus bagi pelanggan berisiko tinggi.

Proses analisis potensi *customer churn* pada pelanggan bisnis Indibiz Telkom Regional I diawali dengan pertanyaan mendasar mengenai bagaimana tahapan teknis implementasi algoritma *Decision Tree* (C4.5) dapat diterapkan secara efektif pada data pelanggan yang ada. Hal ini mencakup seluruh rangkaian proses *data mining*, mulai dari *preprocessing* hingga pembentukan model klasifikasi. Selanjutnya, penelitian ini berupaya menjawab persoalan mengenai atribut atau faktor-faktor spesifik apa saja yang memiliki pengaruh paling signifikan dalam menentukan probabilitas seorang pelanggan bisnis untuk berhenti berlangganan. Terakhir, rumusan masalah ini difokuskan pada bagaimana hasil analisis yang diperoleh dari sistem tersebut dapat ditransformasikan menjadi informasi yang strategis bagi pihak perusahaan, khususnya dalam membantu manajemen menentukan prioritas penanganan pelanggan guna menekan angka *churn*.

Tujuan utama dari penelitian ini adalah untuk melakukan analisis mendalam mengenai potensi *customer churn* pada segmen pelanggan bisnis dengan memanfaatkan kekuatan prediktif dari algoritma *Decision Tree* (C4.5). Melalui pendekatan ini, penelitian ditargetkan mampu mengidentifikasi secara akurat faktor-faktor dominan dan pola perilaku pelanggan yang menjadi pemicu utama terjadinya keputusan layanan. Lebih jauh lagi, penelitian ini bertujuan untuk menghasilkan sebuah sistem informasi fungsional yang dapat digunakan oleh pihak manajemen Indibiz Telkom Regional I sebagai instrumen pendukung keputusan. Sistem tersebut diharapkan mampu memberikan rekomendasi mengenai prioritas pelanggan yang memerlukan strategi retensi khusus berdasarkan hasil klasifikasi otomatis yang dihasilkan oleh algoritma.

Data Mining dan Analisis Churn Landasan teori dalam penelitian ini berakar pada konsep data mining yang didefinisikan sebagai proses sistematis untuk mengekstraksi pengetahuan atau pola yang tidak trivial dari sekumpulan data besar [8]. Dalam domain manajemen pelanggan, teknik ini diaplikasikan untuk menganalisis customer churn, yaitu sebuah fenomena kritis di mana pelanggan mengakhiri hubungan bisnis dengan perusahaan dalam jangka waktu tertentu [3]. Pemahaman mendalam mengenai churn sangat penting bagi perusahaan layanan digital seperti Telkom, karena biaya untuk mempertahankan pelanggan lama jauh lebih efisien dibandingkan biaya akuisisi pelanggan baru. Oleh karena itu, teknik klasifikasi dalam data mining digunakan untuk membagi data ke dalam kelas-kelas tertentu, dalam hal ini kategori "Churn" dan "Tidak Churn", berdasarkan karakteristik historis yang dimiliki oleh setiap pelanggan [1].

Algoritma Decision Tree C4.5 Sebagai instrumen analisis utama, algoritma Decision Tree C4.5 digunakan untuk membangun model prediksi dalam bentuk struktur pohon yang merepresentasikan aturan-aturan keputusan [6]. Algoritma ini bekerja dengan membagi dataset secara rekursif hingga mencapai tingkat homogenitas tertinggi di setiap cabangnya. Proses pemilihan atribut terbaik dalam pembentukan node pada pohon keputusan didasarkan pada perhitungan matematis yang komprehensif, dimulai dari nilai Entropy untuk mengukur tingkat ketidakteraturan data, diikuti dengan perhitungan Information Gain untuk melihat efektivitas atribut tersebut. Untuk menghindari bias pada atribut yang memiliki banyak nilai unik, algoritma C4.5 menerapkan Split Information yang kemudian menghasilkan nilai Gain Ratio [6]. Atribut dengan Gain Ratio tertinggi akan terpilih sebagai akar atau cabang utama, yang memungkinkan perusahaan untuk melihat secara visual faktor teknis maupun administratif apa yang paling memicu pelanggan untuk berhenti menggunakan layanan.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* untuk menghasilkan informasi yang objektif dan terukur dari data pelanggan. Sifat penelitian ini adalah deskriptif dan eksperimental, di mana peneliti berupaya menggambarkan karakteristik data pelanggan serta melakukan pengujian langsung terhadap penerapan algoritma *Decision Tree* C4.5 dalam mengklasifikasikan potensi *customer churn*. Lokasi penelitian dilakukan dengan mengambil studi kasus pada Indibiz Telkom Regional I menggunakan data sekunder yang telah melalui proses anonimisasi untuk menjaga kerahasiaan identitas pelanggan. Teknik pengumpulan data dilakukan melalui dua cara utama, yaitu studi pustaka untuk mendalami teori algoritma C4.5 dan penyusunan dataset simulasi yang dirancang dengan atribut-atribut relevan seperti *payment status*, *total ticket*, dan *total complain*.

Metode analisis data dijalankan secara sistematis mengikuti tahapan teknis algoritma C4.5 yang meliputi *preprocessing* data untuk pembersihan, diikuti dengan perhitungan nilai *entropy* guna mengukur tingkat ketidakteraturan data. Selanjutnya, sistem melakukan perhitungan *information gain* dan *gain ratio* pada setiap atribut untuk menentukan *node* atau akar pada pohon keputusan. Proses ini dilakukan secara berulang hingga terbentuk struktur pohon keputusan yang utuh dan menghasilkan aturan keputusan (*if-then*) yang mudah diinterpretasikan. Pada tahap akhir, model yang telah terbentuk diimplementasikan ke dalam sistem berbasis web menggunakan bahasa pemrograman Python dan *framework* Flask, yang memungkinkan pihak manajemen untuk melakukan klasifikasi otomatis dan memantau laporan potensi *churn* pelanggan secara efisien

### 3. HASIL DAN PEMBAHASAN

Hasil perhitungan algoritma Decision Tree C4.5 menunjukkan bahwa atribut Total Ticket memiliki nilai *Gain Ratio* tertinggi sebesar 0,193, sehingga dipilih sebagai *root node*. Hal ini mengindikasikan bahwa jumlah tiket layanan merupakan faktor paling dominan dalam menentukan potensi *customer churn*, di mana tingginya intensitas interaksi teknis mencerminkan potensi ketidakpuasan pelanggan [1], [6]. Atribut Total Complain juga memberikan pengaruh signifikan dengan nilai *Gain Ratio* 0,186 sebagai pemisah lanjutan, sementara Payment Status memiliki pengaruh paling rendah sebesar 0,034 [3]. Pola utama yang terbentuk menunjukkan bahwa kombinasi *total ticket* dan *total complain* yang tinggi secara konsisten mengklasifikasikan pelanggan ke dalam kategori *churn* [4].

Implementasi model ini diwujudkan dalam sistem informasi berbasis web yang mencakup halaman *dashboard*, data pelanggan, analisis, hingga laporan (*report*). Sistem dirancang secara transparan dengan menampilkan aturan keputusan (*decision rules*) secara eksplisit, sehingga berbeda dengan model *black-box* lainnya [5], [10]. Fitur visualisasi pada *dashboard* dan laporan distribusi pelanggan memudahkan pihak operasional maupun manajerial dalam melakukan monitoring secara sistematis [8]. Pembagian hak akses antara Admin dan Manajer memastikan bahwa sistem ini dapat mendukung proses pengambilan keputusan strategis dalam upaya retensi pelanggan secara lebih cepat dan akurat [7], [11]. Pengujian terhadap sistem menunjukkan bahwa integrasi algoritma C4.5 ke dalam aplikasi web mampu memberikan gambaran risiko *churn* yang mudah dipahami oleh pengguna [9].

#### 3.1 Struktur pohon keputusan akhir

Root : Total ticket

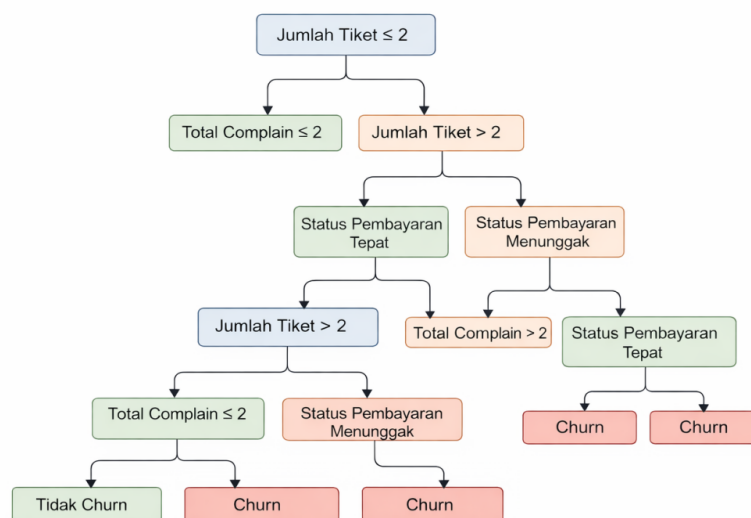
Jika Total ticket Rendah:

- Jika Total Complain Rendah → Tidak Churn
- Jika Total Complain Tinggi → Analisis payment status
- Jika payment status = Tepat → Tidak Churn
- Jika payment status = Menunggak → Churn

Jika Total ticket Tinggi:

- Jika Total Complain Tinggi → Churn
- Jika Total Complain Rendah → Analisis payment status
- Jika payment status = Tepat → Churn
- Jika payment status = Menunggak → Churn

#### 4.1.2 Visualisasi pohon keputusan

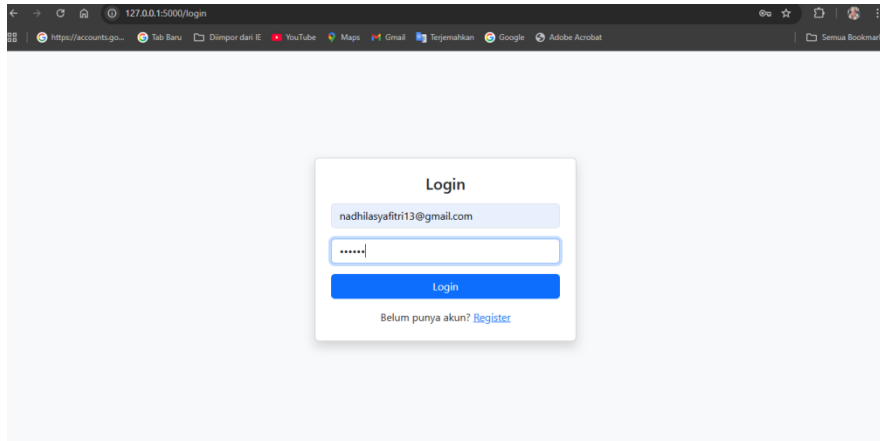


Gambar 4.1 Visualisasi pohon keputusan

Gambar diatas menampilkan visualisasi pohon keputusan yang terbentuk berdasarkan proses perhitungan algoritma C4.5. Visualisasi menunjukkan struktur pemisahan atribut dimulai dari Total Ticket sebagai root node, kemudian dilanjutkan dengan atribut Total Complain dan Payment Status sebagai cabang lanjutan.

### 4.1.3 Implementasi sistem

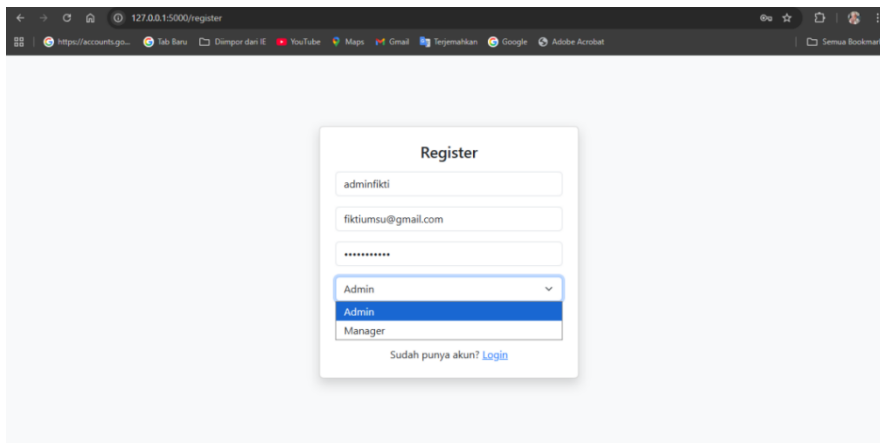
#### a. Tampilan login



Gambar 4.2 Halaman login

Halaman login digunakan sebagai proses autentikasi pengguna sebelum mengakses sistem. Pada halaman ini, pengguna diminta untuk memasukkan username dan password yang telah terdaftar di dalam sistem.

#### b. Halaman registrasi

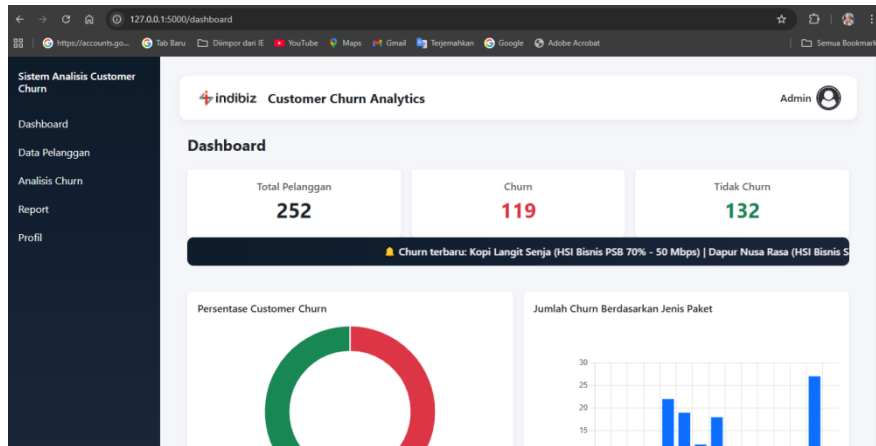


Gambar 4.3 Halaman registrasi

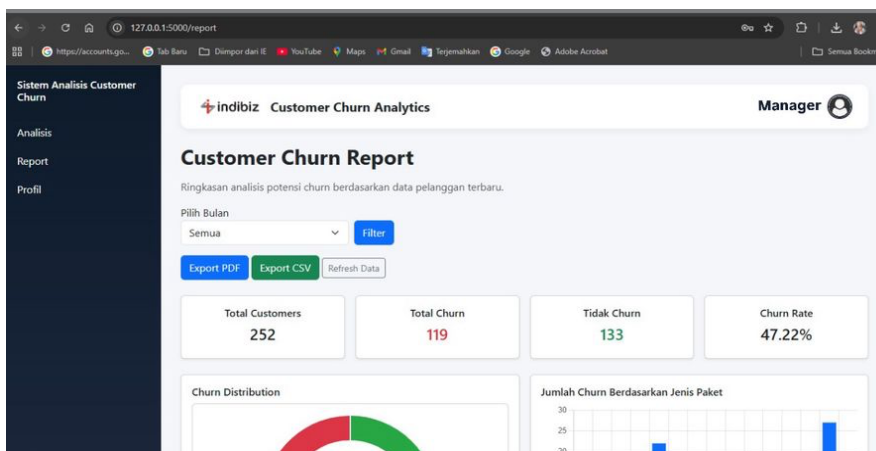
Halaman registrasi digunakan untuk melakukan pendaftaran akun baru. Pengguna diminta mengisi data yang diperlukan untuk membuat akun agar dapat mengakses sistem.

#### c. Halaman dashboard

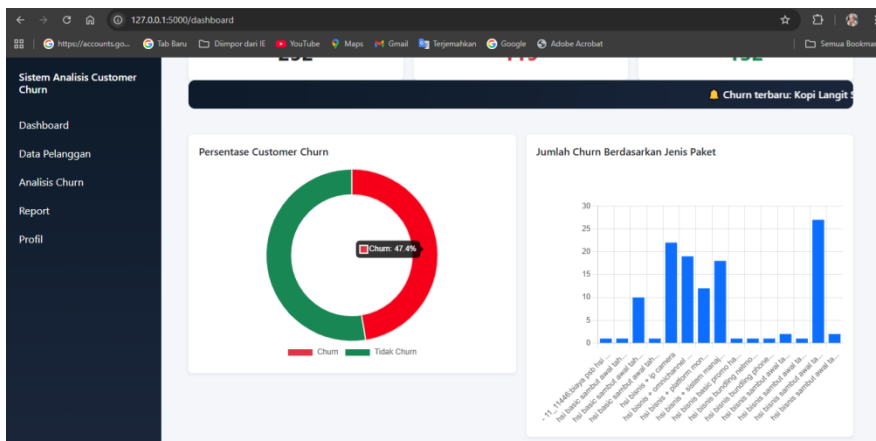
Halaman dashboard menampilkan ringkasan informasi terkait jumlah pelanggan, jumlah churn, dan tidak churn. Dashboard juga menampilkan visualisasi data dalam bentuk grafik dan diagram untuk memberikan gambaran umum kondisi pelanggan. Pada halaman dashboard juga menampilkan data customer dalam tabel pagination.



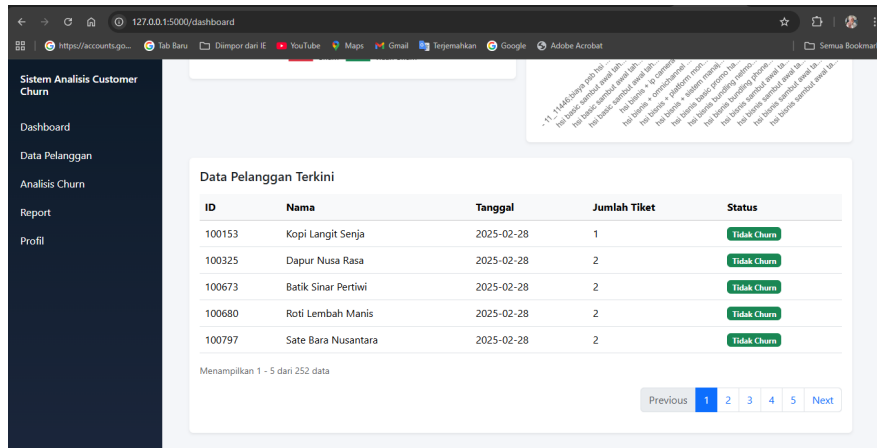
Gambar 4.4 Dashboard admin



Gambar 4.5 Dashboard manager

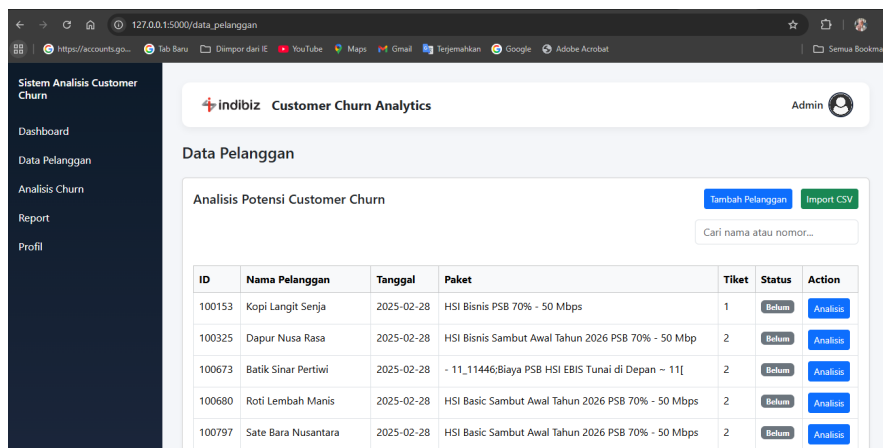


Gambar 4.6 Bagian tengah dashboard



Gambar 4.7 Bagian bawah dashboard

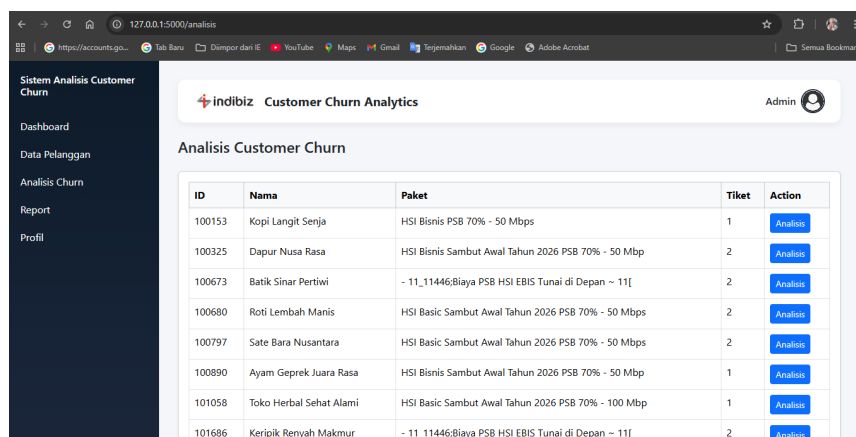
d. Halaman Data Pelanggan



Gambar 4.8 Halaman Data pelanggan

Halaman data pelanggan menampilkan daftar pelanggan yang tersimpan dalam sistem. Data ditampilkan dalam bentuk tabel dan dilengkapi dengan fitur pagination untuk memudahkan

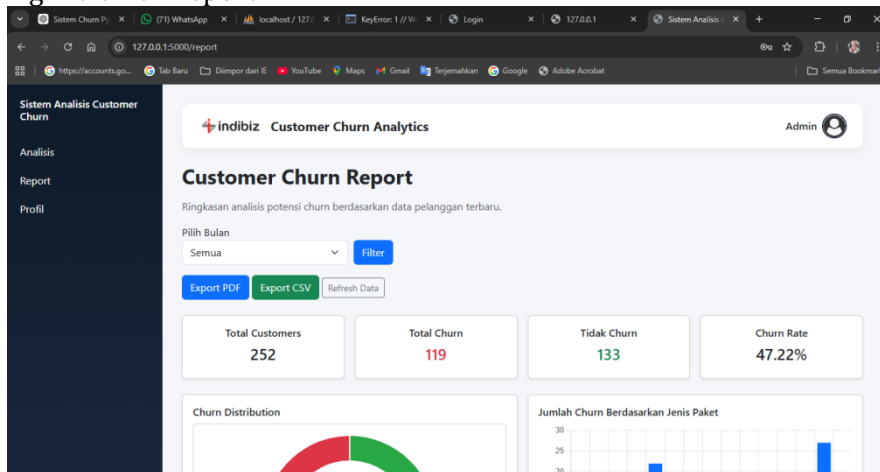
e. Halaman analisis



Gambar 4.9 Halaman analisis

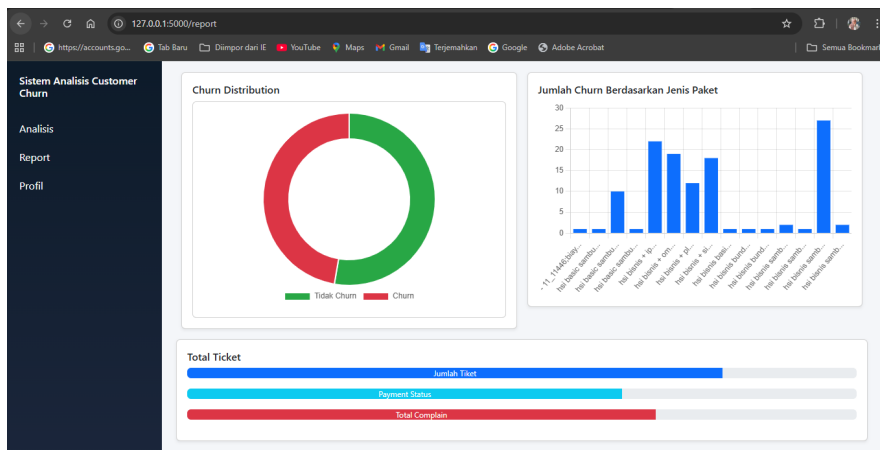


g. Halaman report

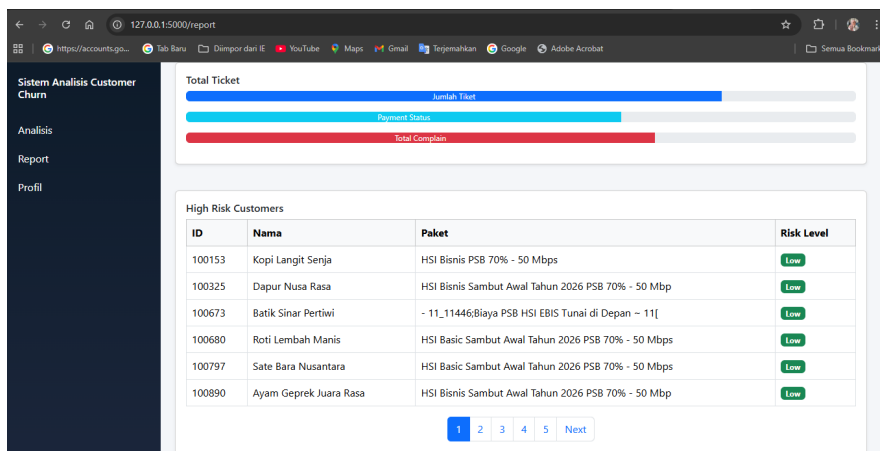


Gambar 4.12 Halaman report

Pada bagian atas halaman report menampilkan fitur filter data berdasarkan bulan sehingga data dapat menyesuaikan dengan hasil filter. Dibawahnya terdapat button export PDF dan CSV untuk mengekspor report.



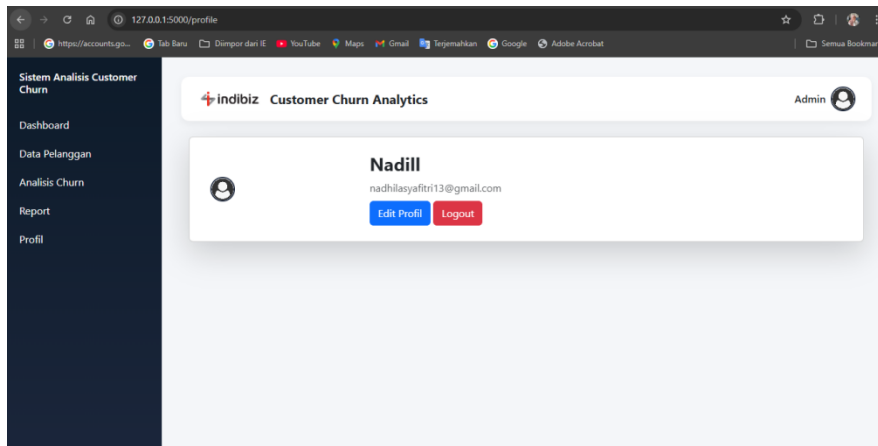
Gambar 4.13 Bagian tengah halaman report



Gambar 4.14 Bagian akhir halaman report

Halaman report menampilkan ringkasan hasil analisis customer churn dalam bentuk statistik dan visualisasi distribusi pelanggan churn dan tidak churn.

h. Halaman Profil



Gambar 4.15 Halaman profil

Pada halaman Profil ditampilkan informasi data akun pengguna yang sedang aktif menggunakan sistem. Informasi yang ditampilkan meliputi nama pengguna dan alamat email yang terdaftar pada sistem. Halaman ini berfungsi sebagai media untuk melihat identitas akun serta memastikan data pengguna telah tersimpan dengan benar di dalam database.

**4.1.4 Uji coba sistem**

Proses pengujian sistem merupakan tahapan penting dalam siklus pengembangan sistem analisis potensi customer churn yang bertujuan untuk mengevaluasi tingkat kesesuaian antara rancangan sistem dengan implementasi yang telah dibangun. Pengujian ini dilakukan untuk memastikan bahwa seluruh fitur dan fungsi sistem mampu berjalan secara optimal sesuai dengan kebutuhan pengguna serta spesifikasi yang telah ditentukan pada tahap perancangan.

Seluruh rangkaian pengujian dilaksanakan pada lingkungan lokal (local environment) sebagai tahap awal sebelum sistem diimplementasikan secara lebih luas. Pendekatan ini memungkinkan proses identifikasi dan perbaikan kesalahan dilakukan secara lebih efektif sehingga kualitas sistem yang dihasilkan dapat terjaga

Tabel 4.8 Uji coba sistem

Fitur	Skenario Pengujian	Output yang diharapkan	Status
Login	Input email dan password valid	Sistem mengarahkan ke dashboard	Berhasil
Login	Password salah	Sistem menampilkan pesan error	Berhasil
Login	Field kosong	Sistem menolak proses login	Berhasil
Register	Input data lengkap	Data tersimpan di database	Berhasil
Register	Field tidak lengkap	Sistem menolak registrasi	Berhasil

Grafik Churn	Menampilkan persentase churn	Grafik donut tampil dengan benar	Berhasil
Grafik Paket	Menampilkan churn berdasarkan jenis paket	Grafik batang tampil dengan benar	Berhasil
Pagination	Pindah halaman data	Data berpindah sesuai halaman	Berhasil
Tambah Pelanggan	Input data pelanggan baru	Data tersimpan di database	Berhasil
Import CSV	Upload file CSV	Data berhasil diimpor	Berhasil
Pencarian	Input keyword nama/nomor	Data tampil sesuai pencarian	Berhasil
Analisis	Klik tombol analisis	Sistem menampilkan hasil prediksi	Berhasil
Klasifikasi	Proses prediksi pelanggan	Status churn tampil	Berhasil
Probabilitas	Menampilkan persentase risiko	Nilai probabilitas tampil	Berhasil
Level Risiko	Kategori rendah/sedang/tinggi	Level risiko tampil sesuai aturan	Berhasil
Update Status	Penyimpanan hasil ke database	Status_churn terupdate	Berhasil
Filter Bulan	Memilih bulan tertentu	Data tampil sesuai periode	Berhasil
Export PDF	Klik tombol export PDF	File PDF terunduh	Berhasil
Export CSV	Klik tombol export CSV	File CSV terunduh	Berhasil
Role Admin	Login sebagai admin	Semua menu dapat diakses	Berhasil
Role Manager	Login sebagai manager	Hanya dapat melihat analisis dan report	Berhasil

## 4.2 Pembahasan

### 4.2.1 Dataset penelitian

Dataset yang digunakan dalam penelitian ini merupakan data pelanggan bisnis Indibiz Telkom Regional I yang diperoleh untuk kepentingan akademik. Data yang digunakan adalah data asli perusahaan, namun telah melalui proses anonimisasi untuk menjaga kerahasiaan dan keamanan informasi pelanggan.

Proses anonimisasi dilakukan dengan menyamarkan atribut yang bersifat sensitif seperti nama pelanggan dan ID pelanggan. Identitas pelanggan tidak ditampilkan dalam bentuk asli, tetapi dianonimkan sehingga tidak dapat ditelusuri kembali kepada individu tertentu. Hal ini dilakukan sesuai dengan prinsip etika penelitian dan perlindungan data.

Total dataset yang digunakan dalam penelitian ini berjumlah ±250 data pelanggan. Dataset tersebut digunakan sebagai data training dalam proses pembentukan model klasifikasi menggunakan algoritma Decision Tree C4.5.

Untuk menjaga efisiensi penulisan dan perlindungan data, tidak seluruh dataset ditampilkan dalam laporan ini. Berikut ditampilkan sebagian data sebagai representasi dataset penelitian yang telah dianonimkan.

Tabel 4.9 Sampel data

ID	Customer name	Date	Package name	Payment	Label	Phone	Tick et	Co mpl ain
----	---------------	------	--------------	---------	-------	-------	---------	------------

112654	CV Sejahtera Jaya	2025-09-24	HSI Bisnis + IP Camera	menunggak	ya	0812000209	1	0
112656	PT Nusantara Bersama	2025-09-25	HSI Bisnis + Omnichannel Engagement Platform	tepat	ya	0812000210	1	0
112661	CV Indonesia Teknologi	2025-09-26	HSI Bisnis + Sistem Manajemen Sekolah	tepat	ya	0812000211	1	0
112663	PT Maju Abadi	2025-09-27	HSI Bisnis + IP Camera	menunggak	tidak	0812000212	7	0
112680	CV Global Sejahtera	2025-09-28	HSI Bisnis + Platform Monitoring Jaringan	tepat	tidak	0812000213	1	0
112709	PT Digital Mandiri	2025-09-29	HSI Bisnis Sambut Awal Tahun 2026 PSB 70% - 50 Mbps	menunggak	ya	0812000214	1	0
112727	CV Cipta Nusantara	2025-09-30	HSI Bisnis + Omnichannel Engagement Platform	tepat	ya	0812000215	1	1
112730	PT Mitra Indonesia	2025-10-01	HSI Bisnis + Sistem Manajemen Sekolah	tepat	ya	0812000216	1	0
112734	CV Pilar Prima	2025-10-02	HSI Bisnis Sambut Awal Tahun 2026 PSB 70% - 50 Mbps	menunggak	tidak	0812000217	1	1
112736	PT Solusi Global	2025-10-03	HSI Bisnis + IP Camera	tepat	tidak	0812000218	1	1
112739	CV Graha Digital	2025-10-04	HSI Bisnis + Omnichannel Engagement Platform	menunggak	ya	0812000219	4	0
112789	PT Artha Cipta	2025-10-05	HSI Bisnis + Sistem Manajemen Sekolah	tepat	ya	0812000220	1	0
112791	CV Prima Makmur	2025-10-06	HSI Bisnis + IP Camera	tepat	ya	0812000221	1	0
112792	PT Mandiri Sentosa	2025-10-07	HSI Bisnis + Platform Monitoring Jaringan	menunggak	tidak	0812000222	1	0
11279	CV Sejahtera	2025-	HSI Bisnis	tepat	tidak	081200022	1	0

5	Jaya	10-08	Sambut Awal Tahun 2026 PSB 70% - 50 Mbps			3		
112797	PT Nusantara Bersama	2025-10-09	HSI Bisnis + Omnichannel Engagement Platform	menunggak	ya	0812000224	1	1
112802	CV Indonesia Teknologi	2025-10-10	HSI Bisnis + Sistem Manajemen Sekolah	tepat	ya	0812000225	1	0
112803	PT Maju Abadi	2025-10-11	HSI Bisnis Sambut Awal Tahun 2026 PSB 70% - 50 Mbps	tepat	ya	0812000226	1	0
112808	CV Global Sejahtera	2025-10-12	HSI Bisnis + IP Camera	menunggak	tidak	0812000227	1	0

#### 4.2.2 Analisis kinerja model

Analisis kinerja atau evaluasi model menjadi bagian penting dalam penelitian ini karena tujuan utama adalah memastikan bahwa model tidak hanya mampu membentuk pola klasifikasi, tetapi juga memiliki tingkat ketepatan yang baik dalam mendeteksi pelanggan yang berpotensi churn. Dengan melakukan evaluasi, dapat diketahui kekuatan dan kelemahan Evaluasi model merupakan tahap untuk mengukur dan menilai kinerja algoritma yang telah dibangun dalam mengklasifikasikan data. Pada penelitian ini, evaluasi model dilakukan untuk mengetahui seberapa baik algoritma Decision Tree C4.5 dalam memprediksi pelanggan churn dan tidak churn berdasarkan data yang digunakan.

Tahap evaluasi dilakukan setelah proses pembentukan pohon keputusan selesai. Model yang telah dilatih menggunakan data training kemudian diuji menggunakan data testing untuk melihat kesesuaian antara hasil prediksi dan kondisi sebenarnya (label aktual). Hasil pengujian tersebut kemudian dianalisis menggunakan beberapa metrik evaluasi. Dalam penelitian ini, metode evaluasi yang digunakan meliputi Confusion Matrix, Accuracy, Precision, Recall, F1-score, Classification Report.

##### a. Confusion matrix

Confusion matrix merupakan salah satu metode evaluasi performa model klasifikasi yang digunakan untuk membandingkan hasil prediksi model dengan label aktual pada data. Matriks ini disusun dalam bentuk tabel dua dimensi, di mana baris menunjukkan kelas sebenarnya (*actual class*) dan kolom menunjukkan kelas hasil prediksi (*predicted class*).

Ukuran confusion matrix bergantung pada jumlah kelas yang digunakan dalam model klasifikasi. Jika jumlah kelas lebih dari dua (multiclass), maka matriks yang dihasilkan berukuran  $n \times n$ , sesuai dengan jumlah kelas. Namun, pada penelitian ini klasifikasi yang dilakukan bersifat biner (binary classification), yaitu hanya terdiri dari dua kelas, Churn (Ya), Tidak Churn (Tidak) Karena hanya terdapat dua kelas, maka confusion matrix yang digunakan berukuran  $2 \times 2$ .

Bentuk confusion matriks dalam penelitian ini:

Tabel 4.10 Bentuk confusion matrix

Prediksi Aktual	Tidak Churn	Churn
Tidak Churn	TN	FP
Churn	FN	TP

Keterangan :

- **True Positive (TP):** pelanggan yang benar-benar churn dan diprediksi churn oleh model.
- **True Negative (TN):** pelanggan yang benar-benar tidak churn dan diprediksi tidak churn oleh model.
- **False Positive (FP):** pelanggan yang sebenarnya tidak churn tetapi diprediksi churn oleh model.
- **False Negative (FN):** pelanggan yang sebenarnya churn tetapi diprediksi tidak churn oleh model.

Bentuk matriks dari sklearn :

```
[[TN FP]
 [FN TP]
```

Hasil dari pengujian pada Terminal Python :

```
Confusion Matrix:
[[20  4]
 [12 14]]
```

Gambar 4.16 pengujian confusion matrix

Interpretasi:

20 data merupakan True Negative (Tidak Churn diprediksi Tidak Churn).

4 data merupakan False Positive (Tidak Churn diprediksi Churn).

12 data merupakan False Negative (Churn diprediksi Tidak Churn).

14 data merupakan True Positive (Churn diprediksi Churn).

Total data testing sebanyak 50 data, di mana 34 data berhasil diklasifikasikan dengan benar dan 16 data mengalami kesalahan prediksi.

Berdasarkan hasil pengujian dengan pembagian data 80:20, diperoleh akurasi sebesar 68%. Model berhasil mengklasifikasikan 34 dari 50 data uji dengan benar. Nilai precision sebesar 77,8% menunjukkan bahwa sebagian besar pelanggan yang diprediksi churn benar-benar merupakan pelanggan churn. Nilai recall sebesar 53,8% menunjukkan bahwa model telah mampu mendeteksi lebih dari setengah pelanggan churn yang sebenarnya.

Nilai F1-score sebesar 63,6% menunjukkan keseimbangan yang cukup baik antara precision dan recall. Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik dalam memprediksi customer churn, meskipun masih terdapat ruang untuk peningkatan terutama dalam meningkatkan kemampuan deteksi pelanggan churn.

b. Perhitungan Accuracy

*Accuracy* adalah ukuran yang menunjukkan persentase keseluruhan prediksi yang benar dibandingkan dengan total data pengujian. Dalam penelitian ini, accuracy digunakan untuk mengetahui tingkat ketepatan model secara keseluruhan dalam mengklasifikasikan

pelanggan churn dan tidak churn. Namun, accuracy saja belum cukup untuk menggambarkan performa model secara detail, sehingga diperlukan metrik tambahan. Perhitungan *accuracy* dilakukan dengan menggunakan rumus

$$\begin{aligned} accuracy &= \frac{TP + TN}{Total} \\ accuracy &= \frac{50}{14 + 20} \\ accuracy &= \frac{50}{34} \\ accuracy &= \frac{50}{50} \\ \mathbf{accuracy\ 0.68} &\rightarrow \mathbf{68\%} \end{aligned}$$

c. Perhitungan Precision, Recall, dan F-1 Score

### 1. Precision

Precision adalah ukuran yang menunjukkan seberapa banyak prediksi churn yang benar-benar merupakan churn. Dalam konteks penelitian customer churn, precision digunakan untuk mengetahui seberapa akurat sistem ketika memprediksi pelanggan akan churn. Nilai precision yang tinggi menunjukkan bahwa sebagian besar pelanggan yang diprediksi churn memang benar-benar churn.

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Precision &= \frac{14}{14 + 4} \\ &= \frac{14}{18} \\ \mathbf{Precision} &= \mathbf{0.778} \rightarrow \mathbf{77.8\%} \end{aligned}$$

### 2. Recall

Recall adalah ukuran yang menunjukkan seberapa banyak pelanggan churn yang berhasil dideteksi oleh model. Dalam penelitian ini, recall digunakan untuk mengukur kemampuan model dalam mendeteksi pelanggan yang benar-benar churn. Nilai recall yang tinggi menunjukkan bahwa model mampu menangkap sebagian besar pelanggan yang berpotensi berhenti berlangganan.

$$\begin{aligned} Recall &= \frac{TP}{TP + FN} \\ Recall &= \frac{14}{14 + 12} \\ Recall &= \frac{14}{26} \\ \mathbf{Recall} &= \mathbf{0.538} \rightarrow \mathbf{53.8\%} \end{aligned}$$

### 3. F-1 Score

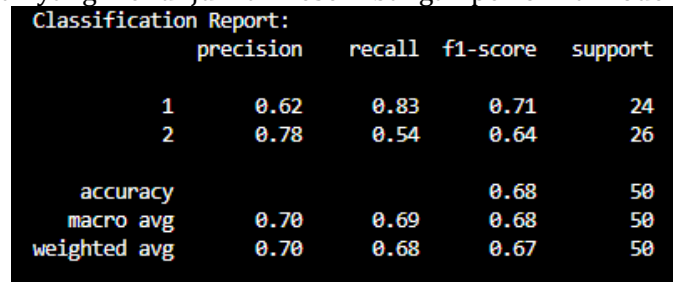
F1-score merupakan rata-rata harmonik antara precision dan recall yang digunakan untuk mengukur keseimbangan antara keduanya. Dalam penelitian ini, F1-score digunakan untuk memberikan gambaran performa model secara lebih seimbang, terutama ketika terdapat perbedaan antara nilai precision dan recall. F1-score membantu menilai apakah model memiliki keseimbangan yang baik dalam memprediksi churn dan mendeteksi churn secara akurat.

$$\begin{aligned} F1 &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\ F1 &= 2 \times \frac{0.778 \times 0.538}{0.778 + 0.538} \\ F1 &= \frac{0.836}{1.316} \\ &= 0.6356 \rightarrow 0.636 \end{aligned}$$

**F1 = 0.6356 → 0.636**

d. Classification Report

Evaluasi model juga dilakukan menggunakan classification report yang menyajikan metrik precision, recall, f1-score, dan support untuk masing-masing kelas target, yaitu Tidak Churn dan Churn. Precision digunakan untuk mengukur tingkat ketepatan model dalam memprediksi suatu kelas, recall menilai kemampuan model dalam mendeteksi seluruh data aktual pada kelas tersebut, sedangkan f1-score merupakan rata-rata harmonis antara precision dan recall yang menunjukkan keseimbangan performa model.



Classification Report:				
	precision	recall	f1-score	support
1	0.62	0.83	0.71	24
2	0.78	0.54	0.64	26
accuracy			0.68	50
macro avg	0.70	0.69	0.68	50
weighted avg	0.70	0.68	0.67	50

Gambar 4.17 Classification report

Berdasarkan hasil classification report, model menunjukkan performa yang cukup baik pada kedua kelas. Pada kelas Tidak Churn, diperoleh f1-score sebesar 0.71 dengan recall yang cukup tinggi, yaitu 0.83. Sementara itu, pada kelas Churn diperoleh precision sebesar 0.78 dan f1-score sebesar 0.64, meskipun nilai recall masih berada pada tingkat sedang, yaitu 0.54. Secara keseluruhan, model menghasilkan akurasi sebesar 68% dengan nilai macro f1-score sebesar 0.68.

Hasil ini menunjukkan bahwa model Decision Tree C4.5 cukup andal dalam mengklasifikasikan pelanggan churn dan tidak churn, meskipun masih terdapat ruang untuk peningkatan terutama dalam meningkatkan kemampuan deteksi pelanggan churn.

## KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis potensi customer churn menggunakan algoritma Decision Tree (C4.5) pada Indibiz Telkom Regional I, menghasilkan beberapa kesimpulan. Yaitu, Penelitian ini berhasil menerapkan metode data mining dengan teknik klasifikasi menggunakan algoritma Decision Tree C4.5 dalam menganalisis potensi customer churn pada pelanggan bisnis Indibiz Telkom Regional I. Proses analisis dilakukan secara sistematis melalui tahapan preprocessing data, perhitungan entropy, information gain, split information, dan gain ratio, hingga pembentukan pohon keputusan sebagai model klasifikasi.

Hasil perhitungan manual dan implementasi sistem menunjukkan bahwa atribut total ticket merupakan faktor yang paling berpengaruh dalam menentukan potensi customer churn, diikuti oleh atribut total complain, dan kemudian payment status. Temuan ini mengindikasikan bahwa semakin tinggi intensitas keluhan dan interaksi pelanggan terhadap layanan, maka semakin besar kemungkinan pelanggan tersebut mengalami churn. Secara bisnis, hal ini dapat menjadi indikator awal dalam mengidentifikasi pelanggan dengan risiko tinggi.

Sistem yang dikembangkan berbasis web menggunakan Python dan framework Flask telah berhasil mengintegrasikan model klasifikasi ke dalam aplikasi yang dapat digunakan oleh dua jenis pengguna, yaitu Admin dan Manager. Sistem mampu menampilkan hasil klasifikasi dalam bentuk status churn (ya/tidak), nilai probabilitas, level risiko (rendah, sedang, tinggi), serta laporan dalam bentuk visualisasi grafik dan file ekspor (PDF dan CSV). Dengan demikian, sistem tidak hanya berfungsi sebagai alat analisis teknis, tetapi juga sebagai sistem pendukung keputusan (*Decision Support System*).

Berdasarkan hasil evaluasi model menggunakan confusion matrix dan classification report dengan pembagian data 80:20, diperoleh nilai akurasi sebesar 68%, precision sebesar 77,8%, recall sebesar 53,8%, dan F1-score sebesar 63,6%. Nilai precision yang cukup tinggi menunjukkan bahwa model cukup akurat dalam memprediksi pelanggan churn. Namun, nilai recall yang masih berada pada tingkat sedang menunjukkan bahwa model belum sepenuhnya optimal dalam mendeteksi seluruh pelanggan yang benar-benar berpotensi churn. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma Decision Tree C4.5 mampu memberikan model klasifikasi yang *interpretable* (mudah dipahami) melalui aturan keputusan (*decision rules*).

#### DAFTAR PUSTAKA

- [1] A. Anas, A. J. Zebua, and A. Akhmadi, "Klasifikasi Data Kelulusan Mahasiswa Menggunakan Algoritma C.45," *Jurnal Ilmiah Media Sisfo*, vol. 19, no. 2, pp. 109-115, 2025, doi: 10.33998/mediasisfo.2025.19.2.2498.
- [2] A. Hermawan, N. R. Jayanti, Z. Tabaruk, F. L. Y. Triadi, A. Saputra, and M. R. H. Syachrudin, "Membangun Model Prediksi Churn Pelanggan yang Akurat," *Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 6, pp. 67-81, 2024, doi: 10.61132/mercurius.v2i6.398.
- [3] S. Antoh, R. Herteno, I. Budiman, D. Kartini, and M. I. Mazdadi, "Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4.5," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 1, pp. 60-67, 2025, doi: 10.14710/vol15iss1pp60-67.
- [4] S. Antoh, R. Herteno, I. Budiman, D. Kartini, and M. I. Mazdadi, "Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4.5," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 1, pp. 60-67, 2025, doi: 10.14710/vol15iss1pp60-67.
- [5] L. P. Ary, S. Tjahyanti, P. S. Saputra, M. S. Gitakarma, T. Informasi, P. Sakti, T. Rekayasa, and S. Elektronika, "Pengembangan Sistem Informasi Pendaftaran PKL di Universitas Panji Sakti," *Jurnal Minfo Polgan*, vol. 12, no. 2, 2023, doi: 10.33395/jmp.v12i2.12545.
- [6] E. Ermawati, B. Irawan, and N. A. Ramdhan, "Klasifikasi Persediaan Obat Menggunakan Algoritma Decision Tree di Puskesmas Brebes," *Jurnal Ilmiah Teknik Mesin, Elektro dan Komputer*, vol. 5, no. 2, pp. 415-432, 2025, doi: 10.51903/juritek.v5i2.5061.
- [7] K. Nistrina and T. A. Lestari, "Desain Inovatif Sistem Informasi Profil Hotel Damanaka Pangalengan Berbasis Website Menggunakan UML dan Figma," n.d.
- [8] E. Setia Budi, A. R. Kadafi, Y. Kharismawan, R. Fadillah, and D. S. Putri, "Analisa Kepuasan Pelanggan terhadap Layanan Aplikasi E-Commerce Menggunakan Algoritma C4.5," *Resolusi: Rekayasa Teknik Informatika dan Informasi*, vol. 4, no. 6, p. 542, 2024. [Online]. Available: <https://djournals.com/resolusi>.
- [9] E. Setia Budi, A. R. Kadafi, Y. Kharismawan, R. Fadillah, and D. S. Putri, "Analisa Kepuasan Pelanggan terhadap Layanan Aplikasi E-Commerce Menggunakan Algoritma C4.5," *Resolusi: Rekayasa Teknik Informatika dan Informasi*, vol. 4, no. 6, p. 542, 2024. [Online]. Available: <https://djournals.com/resolusi>.
- [10] M. A. Setiawan and S. H. Putra, "Pemanfaatan Sistem Informasi Pelayanan Terpadu dalam Meningkatkan Kepuasan Mahasiswa Politeknik Ganesha," *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 1848-1857, 2024, doi: 10.33395/jmp.v13i2.14275.
- [11] G. Booch, J. Rumbaugh, and I. Jacobson, *The Unified Modeling Language User Guide*, 2nd ed. Addison-Wesley, 2005.