

**ANALISIS POTENSI CUSTOMER CHURN MENGGUNAKAN  
ALGORITMA DECISION TREE (C4.5) PADA  
INDIBIZ TELKOM REGIONAL I**

**SKRIPSI**

**DISUSUN OLEH**

**NADHILAH SYAFITRI**

**NPM. 2209010194**



**UMSU**

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2026**

**ANALISIS POTENSI CUSTOMER CHURN  
MENGUNAKAN ALGORITMA DECISION TREE (C4.5)  
PADA INDIBIZ TELKOM REGIONAL I**

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana  
Komputer (S.Kom) dalam Program Studi Sistem Informasi pada Fakultas  
Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah  
Sumatera Utara**

**NADHILAH SYAFITRI**

**NPM. 2209010194**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA  
MEDAN  
2026**

## LEMBAR PENGESAHAN

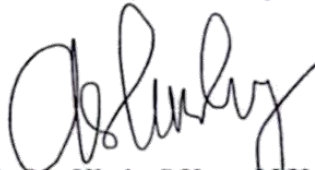
Judul Skripsi : Analisis Potensi Customer Churn Menggunakan  
Algoritma Decision Tree (C4.5) Pada Indibiz  
Telkom Regional I

Nama Mahasiswa : Nadhilah Syafitri

NPM : 2209010194

Program Studi : Sistem Informasi

Menyetujui  
Komisi Pembimbing



(Rizaldy Khair, S.Kom., M.Kom)  
NIDN. 0116098802

Ketua Program Studi



(Mahardika Abdi Prawira Tanjung, M.Kom)  
NIDN. 0117088902

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom)  
NIDN. 0127099201

## PERNYATAAN ORISINALITAS

### ANALISIS POTENSI CUSTOMER CHURN NGGUNAKAN ALGORITMA DECISION TREE (C4.5) PADA INDIBIZ TELKOM REGIONAL I

#### SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, Februari 2026

Yang membuat pernyataan



Nadhilah Syafitri

NPM. 2209010194

**PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN  
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Nadhilah Syafitri  
NPM : 2209010194  
Program Studi : Sistem Informasi  
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksektif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**ANALISIS POTENSI CUSTOMER CHURN MENGGUNAKAN ALGORITMA DECISION TREE (C4.5) PADA INDIBIZ TELKOM REGIONAL I**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksektif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, Februari 2026

Yang membuat pernyataan



Nadhilah Syafitri

NPM. 2209010194

## **RIWAYAT HIDUP**

### **DATA PRIBADI**

Nama Lengkap : Nadhilah Syafitri  
Tempat dan Tanggal Lahir : Medan, 27 November 2003  
Alamat Rumah : Jl. Padang Matinngi Kp. Jawa Ujung  
Telepon/Faks/HP : 08116227322  
E-mail : nadhilasyafitri13@gmail.com

### **DATA PENDIDIKAN**

SD : SDS PANGLIMA POLEM RANTAU PRAPAT TAMAT: 2015  
SMP : SMP NEGERI 1 RANTAU UTARA TAMAT: 2018  
SMA : SMA NEGERI 3 RANTAU UTARA TAMAT: 2021

## KATA PENGANTAR



*Assalamu 'alaikum warahmatullahi wabarakatuh*

Alhamdulillah rabbil'alamin. Puji syukur penulis panjatkan kehadiran Allah SWT, Tuhan Yang Maha Esa karena atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik. Shalawat berangkaikan salam penulis hadiahkan kepada junjungan umat muslim nabi besar Muhammad SAW, semoga kita mendapatkan syafaatnya di hari kemudian kelak. Aamiin.aamiin yaa rabbal 'alamiin.

Tugas akhir ini berjudul “ANALISIS POTENSI CUSTOMER CHURN MENGGUNAKAN ALGORITMA DECISION TREE (C4.5) PADA INDIBIZ TELKOM REGIONAL I” yang mana penelitian ini merupakan salah satu persyaratan untuk menyelesaikan jenjang strata satu (S1) bagi mahasiswa Program Studi Sistem Informasi Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, Medan.

Penulis menyadari bahwa skripsi ini tidak lepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Dengan segala kerendahan hati, penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Akrim, M.Pd. selaku Rektor Universitas Muhammadiyah Sumatera Utara.
2. Bapak Dr. Alkhowarizmi, M.Kom. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi.

3. Ibu Dr. Firaumi Rizky, M.Kom. selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi.
4. Bapak Mhd. Basri, S.Si.,M.Kom. selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.
5. Bapak Mahardika Abdi Prawira Tanjung, M.Kom selaku Ketua Prodi Sistem Informasi.
6. Bapak Mulkan Azhari, S.Kom.,M.Kom selaku Sekretaris Prodi Sistem Informasi.
7. Bapak Rizaldy Khair, S.Kom.,M.Kom. selaku dosen pembimbing tugas akhir yang telah berkontribusi membantu penulis dalam menyelesaikan tugas akhir. Serta memberikan ide, saran, kritik dan bimbingan kepada penulis dalam menyelesaikan tugas akhir.
8. Bapak Amrullah, M.Kom. selaku dosen pembimbing akademik penulis selama masa perkuliahan.
9. Bapak/Ibu Dosen FIKTI UMSU yang telah dengan sabar dan ikhlas memberikan ilmunya kepada penulis.
10. Pegawai biro administrasi dan seluruh staff fakultas ilmu komputer dan teknologi informasi yang telah memudahkan proses administrasi dan perkuliahan penulis.
11. Pihak Indibiz Telkom Regional I yang telah memberikan ruang untuk penulis belajar selama melakukan kerja praktik juga memberikan izin untuk meneliti dan mempermudah segala urusan penulis selama penelitian.
12. Teman-Teman seperjuangan angkatan 2022 FIKTI UMSU.

## LEMBAR PERSEMBAHAN

Alhamdulillah, rasa syukur penulis ucapkan atas terselesaikannya tugas akhir ini. Tentunya penyelesaian tugas akhir ini tidak lepas dari keterlibatan, dukungan, dan bantuan baik moril maupun materiil dari berbagai pihak. Oleh karena itu izinkan penulis mempersembahkan tugas akhir ini kepada :

1. Yang tercinta, Ibu Hj. Sri Arnita, ibu penulis yang do'a-doa'nya selalu menyertai proses penulis dan selalu mensupport kegiatan- kegiatan penulis
2. Yang tersayang, Alm. Bapak Matzen, ayah penulis yang selalu menanamkan kekuatan, kedisiplinan dalam diri penulis. Hiduplah dengan bahagia di jannah-Nya. Al-fatihah
3. dr. Siti Shea gustari dan Dwi Miftah Fatihya, S.Psi. kakak-kakak penulis yang selalu mensupport, serta menjadi teladan bagi penulis
4. E1 sistem informasi 2022 (E1 pening), teman-teman satu perjuangan penulis. Terimakasih atas canda, tawa, serta kerjasamanya selama 4 tahun sudah memberikan warna yang berbeda dalam proses belajar selama perkuliahan
5. Rekan-rekan kerja praktik di Indibiz, rekan-rekan magang selama di PTPN IV Regional II, serta rekan satu tim KKN penulis yang sudah menjadi tim yang kompak, dan supportif selama pelaksanaan program.
6. Pemilik NPM 2209010168 yang telah menemani dan membantu penulis dalam penyusunan skripsi dan juga sedang sama-sama berjuang untuk meraih gelar S.Kom. semoga dipermudah dan dilancarkan segala urusan kedepannya

**ANALISIS POTENSI CUSTOMER CHURN MENGGUNAKAN  
ALGORITMA DECISION TREE (C4.5) PADA  
INDIBIZ TELKOM REGIONAL I**

**ABSTRAK**

Penelitian ini membahas penerapan algoritma Decision Tree C4.5 dalam analisis potensi customer churn pada pelanggan bisnis Indibiz Telkom Regional I. Tujuan utama penelitian ini adalah mengembangkan sistem klasifikasi otomatis yang mampu mengidentifikasi pelanggan berpotensi churn berdasarkan atribut seperti payment status, total ticket, dan total complain. Metode yang digunakan meliputi tahap pengumpulan dan preprocessing data, perhitungan entropy dan information gain, serta pembentukan pohon keputusan menggunakan algoritma C4.5. Hasil penelitian menunjukkan bahwa atribut total ticket memiliki nilai gain ratio tertinggi sehingga menjadi faktor paling dominan dalam proses klasifikasi customer churn. Sistem berbasis web yang dibangun mampu mengelompokkan pelanggan ke dalam kategori churn dan tidak churn secara otomatis dengan tingkat akurasi 68%. Penerapan algoritma C4.5 terbukti efektif dalam membantu perusahaan memahami pola churn serta mendukung pengambilan keputusan dalam menentukan strategi retensi pelanggan.

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

**ANALYSIS OF CUSTOMER CHURN POTENTIAL USING  
THE C4.5 DECISION TREE ALGORITHM AT  
INDIBIZ TELKOM REGIONAL I**

***ABSTRACT***

*This study discusses the implementation of the C4.5 Decision Tree algorithm in analyzing customer churn potential among business customers of Indibiz Telkom Regional I. The primary objective of this research is to develop an automated classification system capable of identifying customers who are likely to churn based on attributes such as payment status, total tickets, and total complaints. The research methodology includes data collection, preprocessing, entropy and information gain calculation, and decision tree construction using the C4.5 algorithm. The results indicate that the total ticket attribute has the highest gain ratio value, making it the most dominant factor in determining churn potential. The developed web-based system is able to automatically classify customers into churn and non-churn categories with an accuracy rate of 68%. The implementation of the C4.5 algorithm proves to be effective in identifying churn patterns and supporting managerial decision-making in developing customer retention strategies.*

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

# DAFTAR ISI

|   |                              |
|---|------------------------------|
| <b>LEMBAR PENGESAHAN</b> .....                | Error! Bookmark not defined. |
| <b>PERNYATAAN ORISINALITAS</b> .....          | <b>i</b>                     |
| <b>PERNYATAAN PERSETUJUAN PUBLIKASI</b> ..... | <b>iii</b>                   |
| <b>KATA PENGANTAR</b> .....                   | <b>v</b>                     |
| <b>LEMBAR PERSEMBAHAN</b> .....               | <b>vii</b>                   |
| <b>ABSTRAK</b> .....                          | <b>viii</b>                  |
| <b>BAB I PENDAHULUAN</b> .....                | <b>1</b>                     |
| 1.1 Latar Belakang Masalah .....              | 1                            |
| 1.2 Rumusan Masalah.....                      | 2                            |
| 1.3 Batasan Masalah .....                     | 3                            |
| 1.4 Tujuan Penelitian .....                   | 3                            |
| 1.5 Manfaat Penelitian .....                  | 4                            |
| <b>BAB II LANDASAN TEORI</b> .....            | <b>5</b>                     |
| 2.1 Data Mining .....                         | 5                            |
| 2.2 Customer Churn .....                      | 5                            |
| 2.3 Klasifikasi .....                         | 6                            |
| 2.4 Algoritma Decision Tree.....              | 6                            |
| 2.5 Algoritma Decision Tree C.45 .....        | 6                            |
| 2.5.1 Entropy.....                            | 7                            |
| 2.5.2 Information Gain.....                   | 7                            |
| 2.5.3 Split Information.....                  | 8                            |
| 2.5.4 Gain Ratio .....                        | 9                            |
| 2.6 Flowchart .....                           | 9                            |
| 2.7 Unified Modelling Language (UML) .....    | 11                           |

|   |           |
|---|-----------|
| 2.7.1 Use Case Diagram.....                         | 12        |
| 2.7.2 Activity Diagram .....                        | 13        |
| 2.7.3 Sequence Diagram .....                        | 14        |
| 2.8 Tools Yang Digunakan.....                       | 17        |
| 2.8.1 Bahasa Pemrograman Python.....                | 17        |
| 2.8.2 Framework Flask .....                         | 17        |
| 2.8.3 Data Base MySQL.....                          | 18        |
| 2.8.4 Library Pendukung.....                        | 18        |
| 2.8.5 Perangkat Pendukung.....                      | 19        |
| 2.9 Penelitian Terkait .....                        | 19        |
| <b>BAB III ANALISA DAN PERANCANGAN SISTEM .....</b> | <b>24</b> |
| 3.1 Jenis dan pendekatan penelitian .....           | 24        |
| 3.2 Objek dan Lokasi Penelitian .....               | 24        |
| 3.3 Jenis dan Sumber Data.....                      | 25        |
| 3.4 Teknik Pengumpulan Data.....                    | 25        |
| 3.5 Metode Analisis Data.....                       | 25        |
| 3.6 Algoritma Sistem .....                          | 26        |
| 3.7 Pemodelan dan perancangan sistem.....           | 26        |
| 3.7.1 Flowchart.....                                | 28        |
| 3.7.2 Activity Diagram .....                        | 30        |
| 3.7.3 Use Case Diagram.....                         | 34        |
| 3.7.4 Sequence Diagram .....                        | 35        |
| 3.7.5 Class Diagram.....                            | 38        |
| 3.8 Implementasi Metode C.45.....                   | 40        |
| 3.8.1 Hasil perhitungan algoritma C4.5.....         | 40        |

|  |           |
|--|-----------|
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b>           | <b>49</b> |
| 4.1 Hasil Penelitian.....                          | 49        |
| 4.1.1 Struktur pohon keputusan akhir.....          | 49        |
| 4.1.2 Visualisasi pohon keputusan .....            | 50        |
| 4.1.3 Implementasi sistem .....                    | 50        |
| 4.1.4 Uji coba sistem .....                        | 58        |
| 4.2 Pembahasan .....                               | 60        |
| 4.2.1 Dataset penelitian .....                     | 60        |
| 4.2.2 Analisis kinerja model .....                 | 62        |
| 4.2.3 Analisis faktor penentu customer churn ..... | 69        |
| 4.2.4 Analisis implementasi sistem .....           | 71        |
| <b>BAB V KESIMPULAN DAN SARAN.....</b>             | <b>74</b> |
| 5.1 Kesimpulan.....                                | 74        |
| 5.2 Saran .....                                    | 75        |
| <b>DAFTAR PUSTAKA .....</b>                        | <b>77</b> |

## DAFTAR TABEL

|   |    |
|---|----|
| Tabel 2.1 Simbol & Notasi Flowchart .....                 | 10 |
| Tabel 2.2 Simbol & Notasi Use Case Diagram.....           | 12 |
| Tabel 2.3 Simbol & Notasi Activity Diagram .....          | 14 |
| Tabel 2.4 Simbol & Notasi Sequence Diagram .....          | 15 |
| Tabel 2.5 Simbol & Notasi Class Diagram.....              | 16 |
| Tabel 2.6 Penelitian Terdahulu .....                      | 20 |
| Tabel 4.1 Distribusi atribut payment status.....          | 42 |
| Tabel 4.2 Distribusi atribut total ticket .....           | 44 |
| Tabel 4.3 Distribusi atribut total complain .....         | 45 |
| Tabel 4.4 Perbandingan gain ratio.....                    | 46 |
| Tabel 4.5 Split tahap pertama.....                        | 46 |
| Tabel 4.6 Uji atribut total complain entropy rendah.....  | 47 |
| Tabel 4.7 Uji atribut total complain entropy tinggi ..... | 48 |
| Tabel 4.8 Uji coba sistem.....                            | 59 |
| Tabel 4.9 Sampel data .....                               | 61 |
| Tabel 4.10 Bentuk confusion matrix .....                  | 64 |

## DAFTAR GAMBAR

|  |    |
|--|----|
| Gambar 3.1 Flowchart sistem klasifikasi customer churn.....                | 28 |
| Gambar 3.2 Activity Diagram Admin .....                                    | 30 |
| Gambar 3.3 Activity Diagram Manager .....                                  | 32 |
| Gambar 3.4 Use Case Diagram .....  | 34 |
| Gambar 3.5 Sequence Diagram Proses Analisis Customer Chur oleh Admin ..... | 35 |
| Gambar 3.6 Sequence Diagram Melihat Report oleh Manager .....              | 37 |
| Gambar 3.7 Class Diagram .....   | 38 |
| Gambar 4.1 Visualisasi pohon keputusan.....                                | 50 |
| Gambar 4.2 Halaman login .....   | 50 |
| Gambar 4.3 Halaman registrasi.....   | 51 |
| Gambar 4.4 Dashboard admin.....  | 52 |
| Gambar 4.5 Dashboard manager .....   | 52 |
| Gambar 4.6 Bagian tengah dashboard.....                                    | 52 |
| Gambar 4.7 Bagian bawah dashboard .....                                    | 53 |
| Gambar 4.8 Halaman data pelanggan .....                                    | 53 |
| Gambar 4.9 Halaman analisis .....  | 54 |
| Gambar 4.10 Halaman analisis customer.....                                 | 55 |
| Gambar 4.11 Tampilan rules analsis .....                                   | 55 |
| Gambar 4.12 Halaman report.....  | 56 |
| Gambar 4.13 Bagian tengah halaman report .....                             | 56 |
| Gambar 4.14 Bagian akhir halaman report.....                               | 57 |
| Gambar 4.15 Halaman profil.....  | 57 |
| Gambar 4.16 pengujian confusion matrix.....                                | 65 |
| Gambar 4.17 Classification report .....                                    | 68 |

## DAFTAR RUMUS

|                                     |   |
|-------------------------------------|---|
| Rumus 2.1 Entropy .....             | 7 |
| Rumus 2.2 information Gain.. .....  | 8 |
| Rumus 2.3 Split Information . ..... | 8 |
| Rumus 2.4 Gain Ratio . .....        | 9 |

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Perkembangan teknologi informasi dan digitalisasi layanan telah mendorong perusahaan penyedia layanan berbasis digital untuk semakin berfokus pada kepuasan dan loyalitas pelanggan. Di tengah persaingan yang semakin ketat, perusahaan tidak hanya dituntut untuk memperoleh pelanggan baru, tetapi juga mampu mempertahankan pelanggan yang sudah ada. Salah satu permasalahan yang sering dihadapi oleh perusahaan adalah terjadinya customer churn, yaitu kondisi ketika pelanggan berhenti menggunakan atau tidak melanjutkan layanan dalam periode tertentu (Andy Hermawan et al., 2024)

Customer churn dapat memberikan dampak negatif bagi perusahaan, terutama dari sisi pendapatan dan keberlangsungan bisnis. Menurut (Andy Hermawan et al., 2024) Kehilangan pelanggan secara terus-menerus dapat meningkatkan biaya operasional, karena perusahaan harus mengeluarkan sumber daya yang lebih besar untuk menarik pelanggan baru dibandingkan mempertahankan pelanggan lama. Oleh karena itu, diperlukan suatu pendekatan yang mampu membantu perusahaan dalam mengidentifikasi pelanggan yang berpotensi mengalami churn sejak dini.

Pada Penelitian (Setia Budi et al., 2024) Pemanfaatan data pelanggan melalui teknik data mining menjadi salah satu solusi yang dapat digunakan untuk mengatasi permasalahan tersebut. Dengan metode klasifikasi, data pelanggan dapat dianalisis untuk menemukan pola-pola tertentu yang berkaitan dengan

perilaku churn. Salah satu algoritma yang banyak digunakan dalam klasifikasi adalah Decision Tree C4.5 karena mampu menghasilkan model yang mudah dipahami serta dapat menjelaskan faktor-faktor yang memengaruhi hasil klasifikasi

Dalam konteks perusahaan penyedia layanan digital, termasuk Indibiz Telkom Regional I, analisis potensi customer churn menjadi hal yang penting untuk dilakukan. Oleh karena itu, penelitian ini dilakukan untuk menganalisis potensi customer churn pada pelanggan bisnis Indibiz Telkom Regional I menggunakan algoritma Decision Tree C4.5 dengan memanfaatkan data customer yang telah dianonimkan, sehingga hasil analisis dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan.

Hasil analisis customer churn yang diperoleh tidak hanya digunakan sebagai informasi teknis, tetapi juga disajikan dalam bentuk laporan (report) yang dapat dimanfaatkan oleh pihak manajemen sebagai bahan pendukung dalam pengambilan keputusan dan penentuan prioritas pelanggan.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana proses analisis potensi customer churn pada pelanggan bisnis menggunakan algoritma Decision Tree (C4.5) pada Indibiz Telkom Regional I?
2. Atribut apa saja yang berpengaruh dalam menentukan potensi customer churn pelanggan bisnis?
3. Bagaimana hasil analisis potensi customer churn yang dihasilkan oleh

sistem untuk membantu pihak perusahaan dalam menentukan prioritas pelanggan?

### **1.3 Batasan Masalah**

Agar penelitian ini lebih terarah dan tidak menyimpang dari tujuan yang telah ditetapkan, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya membahas analisis potensi customer churn pada Indibiz Telkom Regional I
2. Data yang digunakan dalam penelitian ini merupakan data pelanggan Indibiz Telkom Regional I yang telah dianonimkan. Identitas pribadi pelanggan seperti nama dan nomor pelanggan disamarkan untuk menjaga kerahasiaan data (privacy) sesuai dengan etika penelitian.
3. Metode yang digunakan adalah data mining dengan teknik klasifikasi.
4. Algoritma yang digunakan dalam penelitian ini adalah Decision Tree (C4.5).
5. Atribut yang digunakan meliputi id customer, customer name, order date, package name, payment status, phone number, total ticket, total complain.
6. Sistem dapat diakses oleh Staf (Admin) sebagai pengelola data teknis dan Manager (Pimpinan) sebagai pengamat hasil analisis dan laporan

### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis potensi customer churn pada pelanggan bisnis menggunakan algoritma Decision Tree (C4.5).
2. Mengidentifikasi faktor-faktor yang memengaruhi potensi terjadinya customer

churn berdasarkan atribut pelanggan.

3. Menghasilkan sistem yang dapat membantu pihak perusahaan dalam menentukan prioritas pelanggan berdasarkan hasil analisis customer chain

### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

#### 1. Bagi Penulis

Penelitian ini dapat menambah wawasan dan pemahaman mengenai penerapan data mining, khususnya algoritma Decision Tree (C4.5), dalam menganalisis potensi customer churn.

#### 2. Bagi Institusi Akademik

Penelitian ini dapat dijadikan sebagai referensi dan bahan kajian bagi penelitian selanjutnya yang berkaitan dengan data mining dan analisis customer churn.

#### 3. Bagi Pihak Perusahaan

Penelitian ini dapat memberikan gambaran awal mengenai potensi customer churn pelanggan bisnis serta faktor-faktor yang memengaruhinya sebagai bahan pertimbangan dalam evaluasi layanan dan pengambilan keputusan. sistem ini juga memberikan kemudahan dalam pemantauan kondisi pelanggan melalui fitur laporan (reporting), sehingga dapat menjadi pendukung keputusan (*Decision Support*) dalam menentukan strategi retensi pelanggan yang tepat.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Data Mining**

Secara konseptual, data mining adalah metode penemuan tren atau informasi berharga yang tersimpan di dalam basis data berskala besar. Esensi dari aktivitas ini adalah untuk menyarikan pemahaman atau wawasan baru yang krusial sebagai fondasi strategis dalam menentukan arah kebijakan perusahaan. Data mining sering digunakan dalam berbagai bidang, seperti bisnis, pendidikan, kesehatan, dan industri, untuk menganalisis data historis dan memprediksi kecenderungan di masa mendatang.(Setia Budi *et al.*, 2024)

Dalam penelitian ini, data mining digunakan untuk menganalisis data pelanggan bisnis guna mengidentifikasi potensi terjadinya customer churn berdasarkan pola-pola tertentu yang terbentuk dari data tersebut.

#### **2.2 Customer Churn**

Customer churn merupakan keadaan ketika pelanggan tidak lagi melanjutkan penggunaan layanan perusahaan pada kurun waktu tertentu. Pada penelitian (Antoh *et al.*, 2025) menjelaskan bahwa Churn merupakan salah satu indikator penting dalam mengukur tingkat loyalitas pelanggan terhadap suatu layanan. Persentase churn yang tinggi dapat menyebabkan penurunan pendapatan dan memengaruhi keberlanjutan operasional perusahaan.

Analisis customer churn bertujuan untuk mengidentifikasi pelanggan yang berpotensi berhenti berlangganan sehingga perusahaan dapat melakukan tindakan pencegahan lebih awal. Dengan mengetahui faktor-faktor penyebab churn,

perusahaan dapat meningkatkan kualitas layanan dan mempertahankan pelanggan yang sudah ada.

### **2.3 Klasifikasi**

Dalam ranah data mining, klasifikasi berfungsi sebagai metode untuk memetakan data ke dalam kategori-kategori spesifik sesuai dengan ciri atau variabel yang melekat pada data tersebut. Proses klasifikasi dilakukan dengan membangun sebuah model berdasarkan data latih (*Data Training*) yang telah diketahui kelasnya, kemudian model tersebut digunakan untuk memprediksi kelas data baru.(Anas *et al.*, 2025). Dalam penelitian ini, teknik klasifikasi digunakan untuk mengelompokkan pelanggan bisnis ke dalam kategori berpotensi churn atau tidak churn.

### **2.4 Algoritma Decision Tree**

Decision Tree atau pohon keputusan merupakan salah satu metode klasifikasi yang menyajikan model dalam bentuk struktur pohon. Setiap node pada pohon keputusan merepresentasikan atribut, sedangkan cabang menunjukkan nilai dari atribut tersebut, dan daun menunjukkan hasil klasifikasi.(Ermawati *et al.*, 2025)

Algoritma Decision Tree memiliki kelebihan berupa kemudahan dalam interpretasi hasil karena aturan yang dihasilkan berbentuk *if-then*. Hal ini membuat Decision Tree banyak digunakan dalam analisis yang membutuhkan penjelasan hasil secara jelas dan mudah dipahami.

### **2.5 Algoritma Decision Tree C.45**

C4.5 merupakan pengembangan dari algoritma *Decision Tree* sebelumnya yang menggunakan konsep *gain ratio* dalam pemilihan atribut terbaik sebagai *node* pada pohon keputusan. Algoritma ini mampu menangani atribut bertipe diskrit maupun kontinu serta dapat mengatasi data yang memiliki nilai kosong.

C4.5 menghasilkan model klasifikasi dalam bentuk pohon keputusan yang mudah dipahami karena direpresentasikan dalam aturan if-then.(Antoh *et al.*, 2025). Dalam penelitian ini, algoritma C4.5 digunakan untuk membangun model klasifikasi potensi customer churn berdasarkan atribut pelanggan bisnis.

### 2.5.1 Entropy

Penelitian oleh Ermawati *et al.*, (2025) menjelaskan bahwa *Entropy* berperan sebagai pengukur derajat keacakan atau tingkat ketidakpastian informasi di dalam sekumpulan data.. Semakin beragam data dalam satu kelas, maka nilai *entropy* akan semakin besar. Sebaliknya, jika data berada dalam satu kelas yang sama, maka nilai entropy akan bernilai nol.

Secara matematis, entropy dirumuskan sebagai berikut:

$$Entropy (S) = - \sum_{i=1}^n P_i \log_2 P_i \dots\dots\dots(2.1)$$

*Rumus 2. 1 Entropy*

Keterangan :

$S$  : Dataset

$P_i$  : Proporsi data pada kelas ke-i

$n$  : Jumlah kelas

### 2.5.2 Information Gain

Information Gain digunakan untuk mengukur seberapa besar pengurangan nilai entropy setelah data dibagi berdasarkan suatu atribut tertentu. Atribut dengan nilai information gain yang tinggi dianggap lebih baik dalam memisahkan data ke dalam kelas-kelas tertentu.(Yuliani *et al.* 2025)

Rumus *Information Gain* adalah sebagai berikut :

$$Information\ Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \dots\dots (2.2)$$

*Rumus 2. 2 Information Gain*

Keterangan :

A : Atribut

S : Himpunan Seluruh Data

|S| : Jumlah Seluruh Data

|S<sub>i</sub>| : Jumlah data pada subset ke-i

n : Jumlah nilai pada atribut A

### 2.5.3 Split Information

Split Information merupakan ukuran yang digunakan dalam algoritma Decision Tree C4.5 untuk mengetahui tingkat pembagian data yang dihasilkan oleh suatu atribut. Split Information menggambarkan seberapa besar suatu atribut membagi data ke dalam beberapa subset berdasarkan nilai-nilai yang dimilikinya. Atribut dengan jumlah nilai yang banyak cenderung menghasilkan nilai Split Information yang besar karena membagi data ke dalam banyak bagian.

$$Split\ Information = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \left( \frac{|S_i|}{|S|} \right) \dots\dots\dots (2.3)$$

*Rumus 2.3 Split Information*

Keterangan :

S : Keseluruhan Data

A : Atribut

S<sub>i</sub> : Subset data ke-I

$|S|$  : Jumlah Seluruh Data

$|S_i|$  : Jumlah data pada subset ke-i

#### 2.5.4 Gain Ratio

Gain Ratio merupakan pengembangan dari Information Gain yang digunakan dalam algoritma Decision Tree C4.5 untuk menentukan atribut terbaik dalam pembentukan pohon keputusan. Gain Ratio diperoleh dengan membandingkan nilai Information Gain terhadap nilai Split Information. Penggunaan Gain Ratio bertujuan untuk menghindari kecenderungan Information Gain dalam memilih atribut dengan jumlah nilai yang banyak, sehingga pemilihan atribut menjadi lebih objektif.

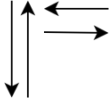



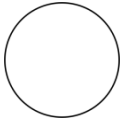

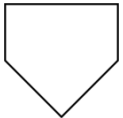



$$\text{Gain Ratio} = \frac{\text{Information Gain}}{\text{Split Information}} \dots\dots\dots (2.4)$$


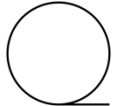
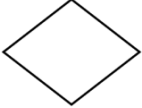



*Rumus 2.4 Gain Ratio*

#### 2.6 Flowchart

Flowchart atau bagan alur adalah diagram yang menampilkan langkah langkah dan keputusan untuk melakukan sebuah proses dari suatu program. Setiap langkah digambarkan dalam bentuk diagram dan dihubungkan dengan garis atau arah panah. flowchart cocok untuk menggambarkan algoritma, Proses perhitungan, dan tahapan teknis pengolahan data (Setiawan & Putra, 2024)

Tabel 2.1 Simbol &amp; Notasi Flowchart

| SIMBOL  | KETERANGAN   | SIMBOL   | KETERANGAN  |
|---|--|--|---|
|    | <p><b>Flow Direction symbol</b></p> <p>Digunakan untuk menghubungkan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga connecting line.</p> |    | <p><b>Manual Input Symbol</b></p> <p>Simbol untuk pemasukan data secara manual on-line keyboard</p>   |
|    | <p><b>Terminator Symbol</b></p> <p>Yaitu simbol untuk permulaan (start) atau akhir (stop) dari suatu kegiatan</p>  |    | <p><b>Preparation Symbol</b></p> <p>Simbol untuk mempersiapkan penyimpanan yang akan digunakan sebagai tempat pengolahan di dalam Storage</p> |
|  | <p><b>Connector Symbol</b></p> <p>Yaitu simbol untuk keluar - masuk atau penyambungan proses dalam lembar / halaman yang sama.</p>                                 |  | <p><b>Predefine Proses Symbol</b></p> <p>Simbol untuk pelaksanaan suatu bagian (sub-program)/prosedur</p>                                     |
|  | <p><b>Connector Symbol</b></p> <p>Yaitu simbol untuk keluar - masuk atau penyambungan proses dalam lembar / Halaman yang berbeda</p>                               |  | <p><b>Display Symbol</b></p> <p>Simbol yang menyatakan peralatan output yang digunakan yaitu layar, plotter, printer dan sebagainya</p>       |
|  | <p><b>Processing Symbol</b></p> <p>Simbol yang menunjukkan pengolahan yang dilakukan oleh komputer</p>   |  | <p><b>Disk and On-line Storage</b></p> <p>Simbol yang menyatakan input yang berasal dari disk atau disimpan ke disk</p>                       |

|   |  |  |   |
|---|--|--|---|
|  | <p><b>Manual Operation</b></p> <p>Simbol yang menunjukkan pengolahan yang tidak dilakukan oleh komputer</p>                        |  | <p><b>Magnetic Tape Unit</b></p> <p>Simbol yang menyatakan input berasal dari pita magnetik atau output disimpan ke pita magnetik</p> |
|  | <p><b>Decision Symbol</b></p> <p>Simbol pemilihan proses berdasarkan kondisi yang ada</p>  |  | <p><b>Punch Card Symbol</b></p> <p>Simbol yang menyatakan bahwa input berasal dari kartu atau output ditulis ke kartu</p>             |
|  | <p><b>Input-Output Symbol</b></p> <p>Simbol yang menyatakan proses input dan output tanpa tergantung dengan jenis peralatannya</p> |  | <p><b>Document</b></p> <p>Simbol yang menyatakan input berasal dari dokumen dalam bentuk kertas atau output dicetak ke kertas</p>     |

## 2.7 Unified Modelling Language (UML)

Unified Modeling Language (UML) merupakan bahasa pemodelan standar yang digunakan untuk memvisualisasikan, merancang, dan mendokumentasikan sistem secara terstruktur. Pada penelitian (Ary *et al.*, 2023) menjelaskan bahwa UML menyajikan representasi grafis dari suatu sistem sehingga memudahkan pemahaman terhadap fungsi, proses, serta interaksi yang terjadi di dalam sistem. UML banyak digunakan dalam pengembangan sistem informasi dan penelitian yang berkaitan dengan pemodelan proses, termasuk penelitian berbasis data mining.

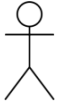

Dalam penelitian ini, UML digunakan untuk menggambarkan proses analisis customer churn pada pelanggan bisnis Indibiz Telkom Regional I menggunakan algoritma Decision Tree C4.5. Pemodelan dengan UML bertujuan untuk






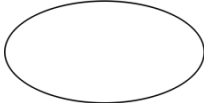

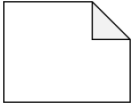
memberikan gambaran umum mengenai alur proses sistem dan aktivitas yang dilakukan dalam penelitian.

### 2.7.1 Use Case Diagram

Use case diagram adalah jenis diagram *Unified Modeling Language (UML)* yang digunakan dalam rekayasa perangkat lunak untuk secara visual merepresentasikan interaksi antara berbagai aktor (pengguna atau sistem eksternal) dan suatu sistem. Diagram ini menggambarkan bagaimana pengguna berinteraksi dengan sistem untuk mencapai tujuan tertentu. Aktor direpresentasikan sebagai figur manusia, dan use case diilustrasikan sebagai oval. Garis hubung menunjukkan hubungan dan interaksi antara aktor dan use case, menggambarkan fungsionalitas dan perilaku sistem dari perspektif pengguna (Nistrina & Lestari, 2024). Use case diagram juga dapat menggambarkan perbedaan peran antar aktor dalam sistem, seperti aktor operasional dan aktor pengambil keputusan.

Tabel 2.2 Simbol & Notasi Use Case Diagram

| NO | GAMBAR  | NAMA              | KETERANGAN   |
|----|---|-------------------|--|
| 1  |  | <i>Actor</i>      | Menspesifikasikan himpunan peran yang pengguna mainkan ketika dengan use case  |
| 2  |  | <i>Dependency</i> | Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri akan mempengaruhi elemen yang tidak mandiri (independent) |



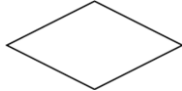

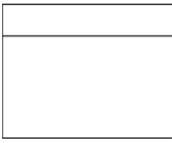
|    |   |                       |   |
|----|---|-----------------------|---|
| 3  |    | <i>Generalization</i> | Hubungan dimana objek anak (descendent) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (ancestor)                  |
| 4  |    | <i>Include</i>        | Menspesifikasikan bahwa use case sumber secara eksplisit.   |
| 5  |    | <i>Extend</i>         | Menspesifikasikan bahwa use case target memperluas perilaku dari use case sumber pada suatu titik yang diberikan.                                 |
| 6  |    | <i>Association</i>    | Apa yang menghubungkan antara objek satu dengan objek lainnya.  |
| 7  |   | <i>System</i>         | Menspesifikasikan paket yang menampilkan sistem secara terbatas.  |
| 8  |  | <i>Use Case</i>       | Deskripsi dari urutan aksi-akst yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor                               |
| 9  |  | <i>Collaboration</i>  | Interaksi aturan-aturan dan elemen lain yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah dan elemen-elemennya (sinergi). |
| 10 |  | <i>Note</i>           | elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya data komputasi  |

### 2.7.2 Activity Diagram

Activity Diagram adalah jenis diagram dalam Unified Modeling Language (UML) yang digunakan untuk menggambarkan aliran kerja atau

aktivitas dalam suatu sistem atau proses. Diagram ini menyajikan serangkaian kegiatan, tindakan, dan keputusan yang terjadi sepanjang waktu. (Nistrina & Lestari, 2024)

Tabel 2.3 Simbol & Notasi Activity Diagram

| NO | GAMBAR  | NAMA                       | KETERANGAN  |
|----|---|----------------------------|---|
| 1  |    | <i>Initial Node</i>        | Sebuah diagram aktivitas memiliki sebuah status awal                                |
| 2  |    | <i>Activity</i>            | Aktivitas yang dilakukan sistem, biasanya diawali dengan kata kerja                 |
| 3  |   | <i>Decision</i>            | Untuk menggambarkan suatu keputusan yang harus diambil pada kondisi tertentu        |
| 4  |   | <i>Line Connector Node</i> | Untuk menghubungkan satu simbol dengan simbol lainnya                               |
| 5  |  | <i>Activity Final Node</i> | Status akhir dari sebuah sistem   |
| 6  |  | <i>Swimlane</i>            | Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi |

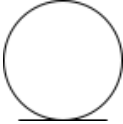
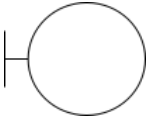

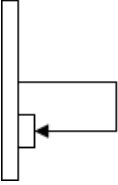

### 2.7.3 Sequence Diagram

Sequence Diagram merupakan salah satu jenis diagram dalam *Unified Modeling Language* (UML) yang digunakan untuk menggambarkan interaksi antar objek dalam suatu sistem berdasarkan urutan waktu. Diagram ini menunjukkan bagaimana objek saling berkomunikasi melalui pengiriman

pesan (message) untuk menjalankan suatu proses atau skenario tertentu dalam sistem perangkat lunak.

Menurut Booch *et al.* (2005), sequence diagram digunakan untuk memodelkan interaksi antar objek yang disusun berdasarkan urutan waktu, sehingga memudahkan dalam memahami dinamika perilaku sistem. Selain itu, sequence diagram juga berfungsi sebagai alat bantu dalam analisis dan perancangan sistem karena mampu menggambarkan alur komunikasi antar komponen secara detail.

Tabel 2.4 Simbol & Notasi Sequence Diagram

| Gambar  | Nama                  | Keterangan   |
|---|-----------------------|--|
|   | <i>Entity Class</i>   | Gambaran sistem sebagai landasan dalam menyusun basis data               |
|  | <i>Boundary Class</i> | Menangani komunikasi antar lingkungan sistem                             |
|  | <i>Control Class</i>  | Bertanggung jawab terhadap kelas-kelas terhadap objek yang berisi logika |
|  | <i>Recursive</i>      | Pesan untuk dirinya  |
|  | <i>Activation</i>     | Mewakili proses durasi aktivasi sebuah operasi                           |


|   |                  |  |
|---|------------------|--|
|  | <i>Life Line</i> | Komponen yang digambarkan garis putus terhubung dengan objek |
|---|------------------|--|

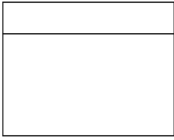
#### 2.7.4 Class Diagram

Class Diagram merupakan salah satu jenis diagram dalam Unified Modeling Language (UML) yang digunakan untuk menggambarkan struktur statis suatu sistem dengan menampilkan kelas-kelas, atribut, metode (method), serta hubungan antar kelas yang terdapat di dalam sistem tersebut. Dalam class diagram, setiap kelas merepresentasikan suatu entitas yang memiliki atribut sebagai data dan metode sebagai perilaku. Hubungan antar kelas dapat berupa asosiasi (association), pewarisan (inheritance), agregasi (aggregation), dan komposisi (composition), yang menggambarkan keterkaitan dan interaksi antar komponen dalam sistem.

Menurut Booch *et al.*, (2005), class diagram digunakan untuk memodelkan struktur sistem dengan menampilkan kelas-kelas serta relasi di antaranya, sehingga dapat memberikan gambaran yang jelas mengenai arsitektur sistem perangkat lunak. Dengan adanya class diagram, proses perancangan sistem dapat dilakukan secara lebih terstruktur dan terorganisir sebelum tahap implementasi.

Tabel 2.5 Simbol & Notasi Class Diagram

| Gambar  | Nama                  | Keterangan   |
|---|-----------------------|--|
|  | <i>Generalization</i> | Hubungan dimana objek anak ( <i>descendent</i> ) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk ( <i>ancestor</i> ) |

|   |              |   |
|---|--------------|---|
|  | <i>Class</i> | Himpunan dari objek-objek yang berbagai atribut serta operasi yang sama |
|---|--------------|---|

## 2.8 Tools Yang Digunakan

Pada penelitian ini, pembangunan sistem analisis customer churn menggunakan algoritma Decision Tree C4.5 didukung oleh berbagai tools dan teknologi yang berperan dalam proses pengolahan data, pengembangan sistem, serta penyajian hasil analisis. Pemilihan tools dan teknologi ini disesuaikan dengan kebutuhan penelitian serta kemudahan implementasi sistem pendukung keputusan berbasis web.

### 2.8.1 Bahasa Pemrograman Python

Bahasa pemrograman Python digunakan sebagai bahasa utama dalam pengembangan sistem. Python dipilih karena memiliki sintaks yang sederhana, mudah dipahami, serta didukung oleh berbagai library yang mendukung penerapan algoritma data mining dan pengolahan data. Dalam penelitian ini, Python digunakan untuk mengimplementasikan algoritma Decision Tree C4.5, melakukan preprocessing data pelanggan simulasi, menjalankan proses klasifikasi customer churn, mengelola logika sistem analisis customer churn.

### 2.8.2 Framework Flask

Flask merupakan framework web berbasis Python yang bersifat ringan (lightweight) dan fleksibel. Flask digunakan sebagai framework utama dalam pembangunan sistem berbasis web pada penelitian ini. Framework ini

memungkinkan pengembangan aplikasi web dengan struktur sederhana namun tetap terorganisir. Flask digunakan untuk mengelola alur aplikasi web (routing), Menghubungkan antarmuka pengguna dengan proses analisis di sisi server, menangani input ID pelanggan untuk proses analisis customer churn, menampilkan hasil klasifikasi dan laporan customer churn kepada pengguna.

### **2.8.3 Data Base MySQL**

Sistem basis data MySQL digunakan sebagai media penyimpanan data pelanggan simulasi. MySQL berfungsi untuk menyimpan data pelanggan yang digunakan sebagai data training dan data testing dalam proses analisis customer churn. Dalam penelitian ini, MySQL digunakan untuk menyimpan atribut pelanggan simulasi, mengelola data pelanggan berdasarkan ID pelanggan, mendukung proses pengambilan data untuk analisis klasifikasi customer churn.

### **2.8.4 Library Pendukung**

Untuk mendukung proses analisis dan pengembangan sistem, digunakan beberapa library pendukung, antara lain Scikit-learn, digunakan untuk membantu proses pembentukan model decision tree sebagai media simulasi penerapan algoritma C4.5. Pandas, digunakan untuk pengolahan dan manipulasi data pelanggan simulasi, NumPy, digunakan untuk mendukung proses perhitungan numerik. ReportLab, digunakan untuk pembuatan laporan hasil analisis customer churn dalam format PDF secara otomatis.

### **2.8.5 Perangkat Pendukung**

Beberapa perangkat pendukung lain yang digunakan dalam penelitian ini meliputi Visual Studio Code sebagai text editor dan IDE pengembangan sistem, Google Chrome sebagai web browser untuk pengujian sistem, phpMyAdmin sebagai alat bantu pengelolaan basis data MySQL, Draw.io / Microsoft Visio untuk pembuatan diagram UML dan flowchart, Microsoft Word untuk penyusunan laporan skripsi.

### **2.9 Penelitian Terkait**

Penelitian terkait mengenai analisis customer churn dan penerapan algoritma Decision Tree telah banyak dilakukan sebelumnya. Beberapa penelitian menunjukkan bahwa algoritma Decision Tree C4.5 mampu menghasilkan model klasifikasi yang cukup baik serta mudah diinterpretasikan dalam analisis perilaku pelanggan. Hasil dari penelitian-penelitian tersebut menjadi salah satu dasar dalam pemilihan metode dan algoritma yang digunakan dalam penelitian ini.

Berikut beberapa penelitian terdahulu:

Tabel 2. 6 Penelitian Terdahulu

| NO. | JUDUL   | PENELITI                     | HASIL PENELITIAN  |
|-----|---|------------------------------|---|
| 1.  | Prediksi Churn Pelanggan Telekomunikasi Dengan Optimalisasi Seleksi Fitur Dan Tuning Hyperparameter Pada Algoritma Klasifikasi C4.5 | Antoh <i>Et al.</i> (2025)   | Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 mampu digunakan akurasi yang lebih tinggi.  |
| 2   | Implementation of Machine Learning Models for Predicting Internet Service Provider Customer Churn                                   | Asyhari <i>et al.</i> (2025) | Penelitian terdahulu menghasilkan F1-score churn sebesar 0.52 dengan recall hanya 0.44 akibat bias data imbalanced pada kelas mayoritas non-churn. Kelemahan utama terletak pada rendahnya kemampuan deteksi pelanggan yang benar-benar churn                                     |
| 3   | Prediksi Customer Churn dengan Algoritma Decision Tree C4.5 Berdasarkan Segmentasi Pelanggan untuk Mempertahankan Pelanggan pada    | Wardani <i>et al.</i> (2018) | Penelitian terdahulu mengenai customer churn umumnya berfokus pada peningkatan akurasi model prediksi, namun masih terbatas penelitian yang menekankan pada analisis faktor-faktor yang memengaruhi customer churn secara interpretatif menggunakan algoritma Decision Tree C4.5, |

|   |   |                                  |  |
|---|---|----------------------------------|--|
|   | Perusahaan Retail   |                                  | sehingga penelitian ini dilakukan untuk memberikan pemahaman terhadap pola churn pada pelanggan Indibiz Telkom Regional I  |
| 4 | Penerapan Algoritma C4.5 untuk Klasifikasi Customer Churn pada Perusahaan Perbankan   | Riftiarraafi dan Ernawati (2024) | Pada penelitian tersebut belum secara spesifik menekankan penggunaan struktur pohon keputusan  |
| 5 | Klasifikasi Pelanggan Pada Customer Churn Prediction Models Menggunakan Decision Tree | Sinata <i>et al</i> (2025)       | Penelitian tersebut berfokus pada pembangunan model klasifikasi dan evaluasi kinerja model berdasarkan metrik akurasi, presisi, dan recall, sehingga menekankan efektivitas algoritma dalam menghasilkan prediksi customer churn yang akurat.  |
| 6 | Prediksi customer churn menggunakan algoritma decision tree                           | Putriani <i>et al.</i> (2024)    | Penelitian dilakukan untuk memprediksi customer churn dengan menggunakan algoritma Decision Tree dan metodologi SEMMA pada dataset publik berskala besar. Penelitian tersebut berfokus pada perbandingan performa beberapa algoritma klasifikasi, seperti Naive Bayes, K-Nearest Neighbor, dan Decision Tree, dengan tujuan utama memperoleh tingkat akurasi prediksi tertinggi. |

|   |   |                              |  |
|---|---|------------------------------|--|
|   |   |                              | Hasil penelitian menunjukkan bahwa algoritma Decision Tree memiliki performa paling optimal berdasarkan metrik akurasi, presisi, recall, dan F1-score  |
| 7 | Prediction and Analysis Customer Churn at Telkomsel Using a Machine Learning Approach                 | Asror dan Nuryana (2025)     | Penelitian tersebut bertujuan untuk menentukan algoritma dengan performa terbaik berdasarkan metrik evaluasi. Sebaliknya, penelitian penulis membandingkan algoritma.  |
| 8 | Comparison of support vector machine, random forest, and C4.5 Algorithms for Customer Loss Prediction | Maulana <i>et al.</i> (2025) | Penelitian tersebut menyimpulkan bahwa Random Forest memiliki akurasi tertinggi, namun algoritma tersebut Decision Tree (C4.5) untuk memahami faktor-faktor yang memengaruhi potensi customer churn.   |
| 9 | Machine Learning-Based Prediction of Telecom Customer Churn: Comparative Model Analysis               | Zhang (2025)                 | Model XGBoost yang digunakan pada penelitian terdahulu memiliki tingkat akurasi yang tinggi, namun bersifat black-box sehingga sulit untuk ditafsirkan secara langsung. Penelitian tersebut hanya membahas feature importance tanpa menyajikan aturan keputusan yang eksplisit. Dalam penelitian |

|    |   |                               |  |
|----|---|-------------------------------|--|
|    |   |                               | dianalisis secara lebih jelas.   |
| 10 | A Decade of Churn Prediction Techniques in TelCo Domain: A Survey | Barsotti <i>et al.</i> (2024) | Penelitian menggunakan algoritma yang bersifat kompleks dan kurang mudah diinterpretasikan sehingga menyulitkan penerapannya dalam pengambilan keputusan manajerial. |

## **BAB III**

### **ANALISA DAN PERANCANGAN SISTEM**

#### **3.1 Jenis dan pendekatan penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining. Pendekatan kuantitatif dipilih karena penelitian berfokus pada pengolahan dan analisis data numerik untuk menghasilkan informasi yang objektif dan terukur. Metode data mining digunakan untuk menggali pola dan pengetahuan dari data pelanggan yang berkaitan dengan potensi terjadinya customer churn.

Penelitian ini bersifat deskriptif dan eksperimental. Bersifat deskriptif karena bertujuan untuk menggambarkan karakteristik data pelanggan dan faktor-faktor yang memengaruhi customer churn. Bersifat eksperimental karena dilakukan pengujian penerapan algoritma Decision Tree C4.5 dalam mengklasifikasikan pelanggan ke dalam kategori churn dan tidak churn.

#### **3.2 Objek dan Lokasi Penelitian**

Objek penelitian ini adalah data-data pelanggan bisnis Indibiz Telkom Regional I. Data yang digunakan bukan data pelanggan asli, melainkan data simulasi yang disusun berdasarkan karakteristik umum pelanggan bisnis Indibiz Telkom Regional I.

Penelitian ini dilakukan di lingkungan akademik dengan studi kasus Indibiz Telkom Regional I, dengan tujuan memberikan gambaran konseptual mengenai penerapan algoritma Decision Tree C4.5 dalam analisis potensi customer churn.

### **3.3 Jenis dan Sumber Data**

Jenis data yang digunakan dalam penelitian ini adalah data sekunder berupa data pelanggan bisnis Indibiz Telkom Regional I. Data yang digunakan merupakan data asli perusahaan yang telah melalui proses anonimisasi dengan menyamarkan identitas pelanggan untuk menjaga kerahasiaan dan keamanan data.

Data pelanggan yang digunakan dalam penelitian ini telah memperoleh izin penggunaan untuk kepentingan akademik dan tidak menampilkan informasi sensitif yang dapat mengungkap identitas pelanggan secara langsung.

Sumber data berasal dari literatur dan jurnal ilmiah yang membahas customer churn dan data mining, referensi terkait layanan dan karakteristik pelanggan bisnis Indibiz, Data simulasi yang dirancang peneliti sebagai data training dan data testing.

### **3.4 Teknik Pengumpulan Data**

Teknik pengumpulan data dalam penelitian ini dilakukan melalui studi pustaka, yaitu pengumpulan data dan informasi dari buku, jurnal, dan penelitian terdahulu yang berkaitan dengan customer churn dan algoritma Decision Tree C4.5 dan Penyusunan data simulasi, yaitu pembuatan dataset pelanggan bisnis dengan atribut-atribut yang relevan terhadap potensi customer churn.

### **3.5 Metode Analisis Data**

Metode analisis data yang digunakan adalah teknik klasifikasi dengan algoritma Decision Tree C4.5. Tahapan analisis data dalam penelitian ini meliputi:

1. Data selection, yaitu pemilihan atribut data pelanggan yang relevan.
2. Data preprocessing, yaitu proses pembersihan dan penyesuaian data simulasi

agar siap digunakan dalam analisis.

3. Data transformation, yaitu perubahan data ke dalam format yang sesuai untuk proses klasifikasi.
4. Penerapan algoritma C4.5 untuk membangun model pohon keputusan.
5. Analisis hasil klasifikasi customer churn berdasarkan aturan yang dihasilkan.

### **3.6 Algoritma Sistem**

Algoritma yang digunakan dalam penelitian ini adalah algoritma Decision Tree C4.5. Algoritma ini dipilih karena mampu menghasilkan model klasifikasi yang mudah dipahami serta dapat menampilkan aturan keputusan dalam bentuk if-then. Proses kerja algoritma C4.5 dalam sistem ini meliputi beberapa tahapan sebagai berikut:

1. Mengumpulkan data berupa data pelanggan simulasi
2. Menghitung nilai entropy untuk mengetahui tingkat ketidakpastian data.
3. Menghitung nilai information gain dan gain ratio untuk setiap atribut.
4. Menentukan atribut dengan nilai gain ratio tertinggi sebagai node pada pohon keputusan.
5. Membentuk pohon keputusan hingga seluruh data dapat diklasifikasikan.
6. Menggunakan pohon keputusan yang terbentuk untuk menganalisis data pelanggan berdasarkan nomor pelanggan yang dimasukkan.

Hasil dari proses ini berupa klasifikasi status customer churn beserta faktor-faktor yang mempengaruhinya.

### **3.7 Pemodelan dan perancangan sistem**

Pemodelan dan perancangan sistem dilakukan untuk memberikan gambaran mengenai alur kerja sistem yang digunakan dalam penelitian ini. Perancangan sistem bertujuan untuk menjelaskan tahapan proses analisis customer churn

secara terstruktur, mulai dari pengelolaan data hingga diperolehnya hasil klasifikasi. Sistem yang dirancang bersifat prototype dan digunakan sebagai media simulasi untuk mendukung proses analisis dalam penelitian.

Sistem menggunakan database simulasi yang disusun oleh peneliti sebagai sumber data. Data pelanggan disimpan dalam database tersebut dengan menggunakan nomor pelanggan fiktif sebagai identitas data. Nomor pelanggan ini berfungsi sebagai kunci utama dalam proses pencarian dan pengambilan data pelanggan ketika analisis dilakukan. Data yang disimpan terdiri dari beberapa atribut yang merepresentasikan karakteristik umum pelanggan layanan bisnis, yang digunakan sebagai variabel dalam proses analisis customer churn.

Alur kerja sistem dimulai ketika Admin memasukkan nomor internet pelanggan (ID Pelanggan) pada menu analisis. Sistem kemudian melakukan proses pencarian data pelanggan berdasarkan nomor internet pelanggan yang dimasukkan dan mengambil data atribut yang sesuai dari database simulasi. Data tersebut selanjutnya diproses menggunakan algoritma Decision Tree C4.5 untuk melakukan analisis dan menentukan klasifikasi potensi customer churn.

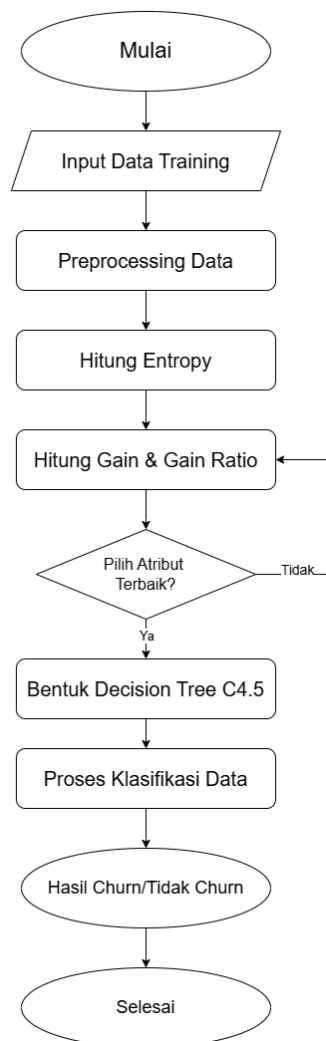
Setelah proses analisis selesai, sistem menampilkan hasil klasifikasi pelanggan ke dalam kategori churn atau tidak churn. Selain itu, sistem juga menampilkan informasi mengenai faktor-faktor yang memengaruhi hasil klasifikasi tersebut berdasarkan aturan yang dihasilkan dari pohon keputusan. Informasi ini digunakan sebagai bahan pembahasan untuk mengidentifikasi pola-pola yang berkaitan dengan terjadinya customer churn pada pelanggan layanan bisnis.

Selain pengguna sebagai operator sistem, sistem juga melibatkan pihak manajemen sebagai aktor yang memanfaatkan output sistem. Hasil analisis

customer churn disajikan dalam bentuk report yang dapat digunakan oleh manager atau pimpinan sebagai bahan evaluasi dan pendukung pengambilan keputusan.

Untuk memperjelas proses dan struktur sistem yang dirancang, pemodelan sistem digambarkan menggunakan Use Case Diagram, Activity Diagram, serta Flowchart untuk menggambarkan alur perhitungan dari algoritma Decision Tree C4.5 Diagram-diagram tersebut digunakan untuk menjelaskan alur proses sistem, interaksi antara pengguna dan sistem, serta struktur data yang digunakan dalam penelitian ini.

### 3.7.1 Flowchart



Gambar 3.1 Flowchart sistem klasifikasi customer churn

Flowchart pada Gambar tersebut menggambarkan alur kerja sistem dalam melakukan analisis customer churn menggunakan algoritma Decision Tree C4.5 secara bertahap dan terstruktur. Flowchart ini menunjukkan urutan proses mulai dari tahap awal hingga diperolehnya hasil klasifikasi customer churn.

Proses diawali dengan mulai, kemudian dilanjutkan dengan input data training yang berisi kumpulan data pelanggan simulasi. Data training ini digunakan sebagai dasar pembentukan model klasifikasi customer churn. Setelah data diinput, sistem melakukan tahap preprocessing data, yang bertujuan untuk memastikan kualitas data dengan melakukan pembersihan data, penyesuaian format, serta penanganan data yang tidak lengkap agar siap digunakan dalam proses analisis.

Selanjutnya, sistem melakukan perhitungan nilai entropy untuk mengukur tingkat ketidakpastian data terhadap kelas churn dan tidak churn. Berdasarkan nilai entropy tersebut, sistem kemudian melakukan perhitungan information gain dan gain ratio untuk setiap atribut. Tahap ini bertujuan untuk menentukan atribut yang paling berpengaruh dalam memisahkan data ke dalam kelas churn dan tidak churn.

Pada tahap berikutnya, sistem melakukan proses pemilihan atribut terbaik. Jika atribut terbaik belum ditemukan, maka sistem akan kembali ke tahap perhitungan gain dan gain ratio secara berulang. Apabila atribut terbaik telah diperoleh, maka sistem melanjutkan proses dengan membentuk struktur Decision Tree C4.5 sebagai model klasifikasi.

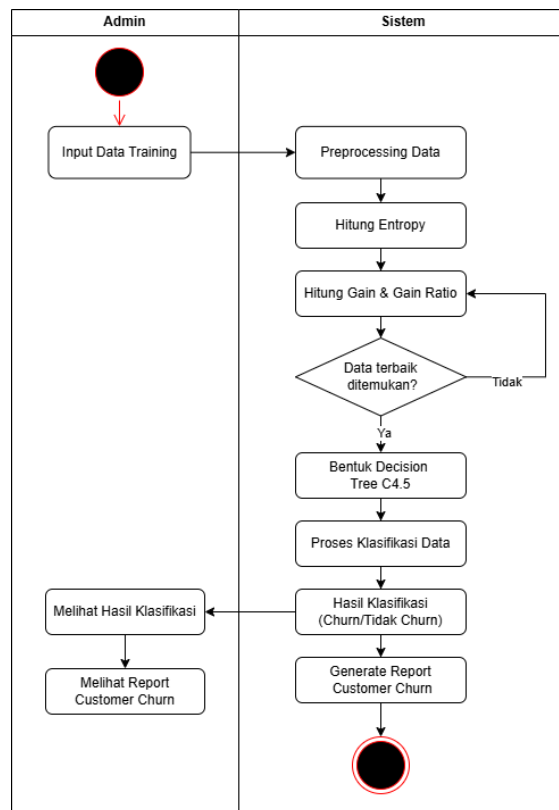
Setelah pohon keputusan terbentuk, sistem melakukan proses klasifikasi data, yaitu pengelompokan data pelanggan berdasarkan aturan

keputusan yang dihasilkan oleh pohon keputusan C4.5. Hasil dari proses ini berupa status pelanggan, apakah termasuk dalam kategori churn atau tidak churn.

Tahap akhir dari flowchart ini adalah penyajian hasil klasifikasi churn/tidak churn, kemudian proses diakhiri dengan selesai. Flowchart ini menunjukkan bahwa algoritma Decision Tree C4.5 digunakan secara sistematis untuk membangun model klasifikasi customer churn yang dapat mendukung analisis dan pengambilan keputusan pada pelanggan bisnis Indibiz Telkom Regional I.

### 3.7.2 Activity Diagram

#### a) admin



Gambar 3.2 Activity Diagram Admin

Activity diagram pada gambar tersebut menunjukkan activity diagram user admin yang menggambarkan alur aktivitas sistem dalam melakukan

analisis customer churn menggunakan algoritma Decision Tree C4.5. Diagram ini terdiri dari dua swimlane, yaitu Admin dan Sistem, yang menunjukkan pembagian peran antara pengguna dan sistem dalam menjalankan proses analisis.

Proses dimulai dari sisi Admin yang berperan sebagai pengguna utama sistem. Aktivitas diawali dengan melakukan input data training, yaitu data pelanggan simulasi yang digunakan sebagai dasar dalam pembentukan model klasifikasi. Data yang telah diinput kemudian dikirim ke sistem untuk diproses lebih lanjut.

Pada bagian Sistem, data yang diterima terlebih dahulu melalui tahap preprocessing data yang bertujuan untuk membersihkan data, menangani data yang tidak lengkap, serta menyesuaikan format data agar sesuai dengan kebutuhan algoritma. Selanjutnya, sistem melakukan perhitungan nilai entropy untuk mengukur tingkat ketidakpastian data terhadap kelas target, yaitu churn dan tidak churn.

Tahap berikutnya adalah perhitungan nilai information gain dan gain ratio untuk setiap atribut. Perhitungan ini bertujuan untuk menentukan atribut terbaik yang paling berpengaruh dalam proses pemisahan data. Sistem kemudian melakukan proses pengambilan keputusan untuk menentukan apakah atribut terbaik telah ditemukan. Jika belum, maka proses perhitungan akan diulang hingga diperoleh atribut yang optimal. Jika telah ditemukan, sistem akan melanjutkan ke tahap berikutnya.

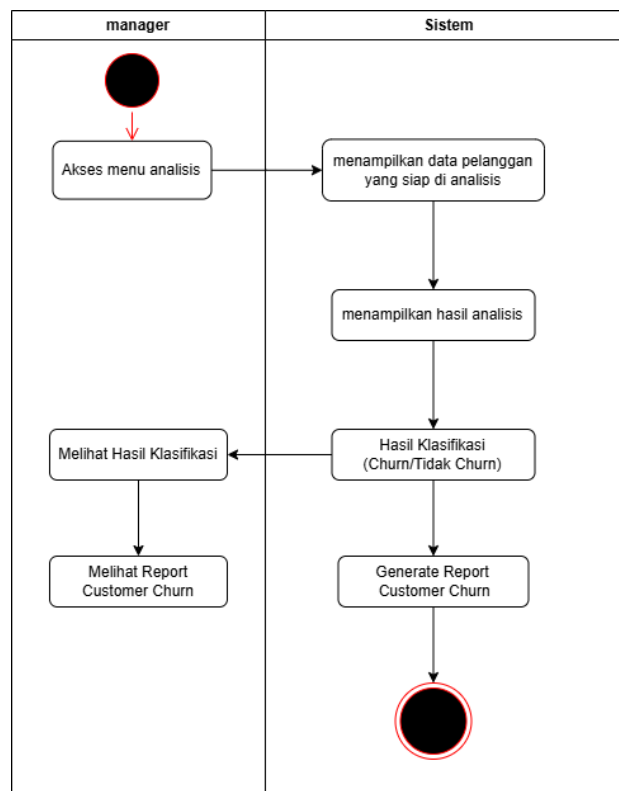
Selanjutnya, sistem membentuk model Decision Tree C4.5 berdasarkan atribut terbaik yang telah dipilih. Model yang terbentuk

kemudian digunakan dalam proses klasifikasi data pelanggan ke dalam kategori churn atau tidak churn berdasarkan aturan keputusan yang dihasilkan.

Hasil klasifikasi yang diperoleh kemudian ditampilkan kepada Admin. Admin dapat melihat hasil klasifikasi tersebut melalui fitur yang tersedia dalam sistem, dan selanjutnya sistem akan melakukan proses generate report customer churn sebagai bentuk penyajian hasil analisis.

Proses berakhir setelah laporan customer churn berhasil dihasilkan oleh sistem. Laporan ini selanjutnya dapat digunakan sebagai bahan evaluasi serta pendukung dalam pengambilan keputusan terkait strategi retensi pelanggan.

#### b) Manager



Gambar 3.3 Activity Diagram Manager

Activity diagram manager pada gambar tersebut menggambarkan alur aktivitas yang dilakukan oleh manager dalam mengakses dan melihat hasil analisis customer churn pada sistem. Diagram ini terdiri dari dua swimlane, yaitu Manager dan Sistem, yang menunjukkan pembagian peran antara pengguna dan sistem dalam proses penyajian hasil analisis.

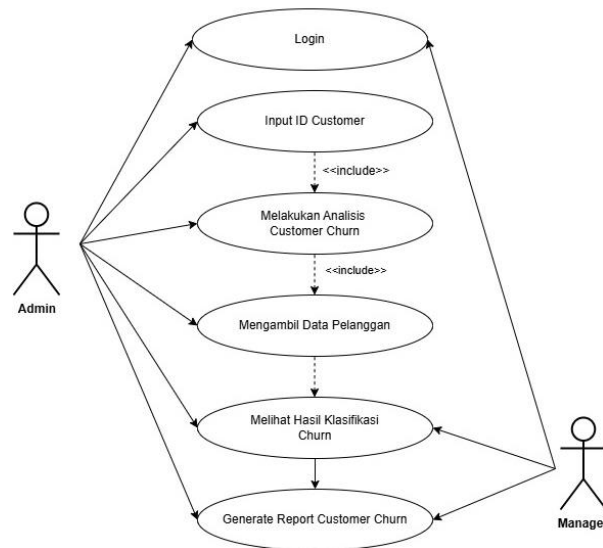
Proses dimulai dari sisi Manager dengan melakukan akses menu analisis pada sistem. Aktivitas ini bertujuan untuk membuka halaman yang menampilkan data pelanggan yang telah siap untuk dianalisis. Permintaan tersebut kemudian diproses oleh sistem. Pada bagian Sistem, sistem akan menampilkan data pelanggan yang telah melalui proses analisis sebelumnya. Selanjutnya, sistem menampilkan hasil analisis yang telah dilakukan menggunakan algoritma Decision Tree C4.5.

Hasil analisis tersebut disajikan dalam bentuk hasil klasifikasi customer churn, yaitu kategori churn atau tidak churn. Informasi ini kemudian dikirimkan kepada Manager untuk ditampilkan melalui fitur melihat hasil klasifikasi. Manager selanjutnya dapat melihat hasil klasifikasi tersebut, kemudian melanjutkan ke tahap melihat report customer churn sebagai bentuk laporan hasil analisis yang telah dihasilkan oleh sistem.

Pada tahap akhir, sistem melakukan proses generate report customer churn berdasarkan hasil klasifikasi yang telah diperoleh. Proses berakhir setelah laporan berhasil dihasilkan dan ditampilkan kepada Manager. Dengan demikian, activity diagram ini menggambarkan alur aktivitas Manager dalam mengakses hasil analisis tanpa terlibat langsung dalam proses perhitungan atau pengolahan data, melainkan hanya berfokus pada

penerimaan informasi hasil analisis sebagai dasar dalam pengambilan keputusan.

### 3.7.3 Use Case Diagram



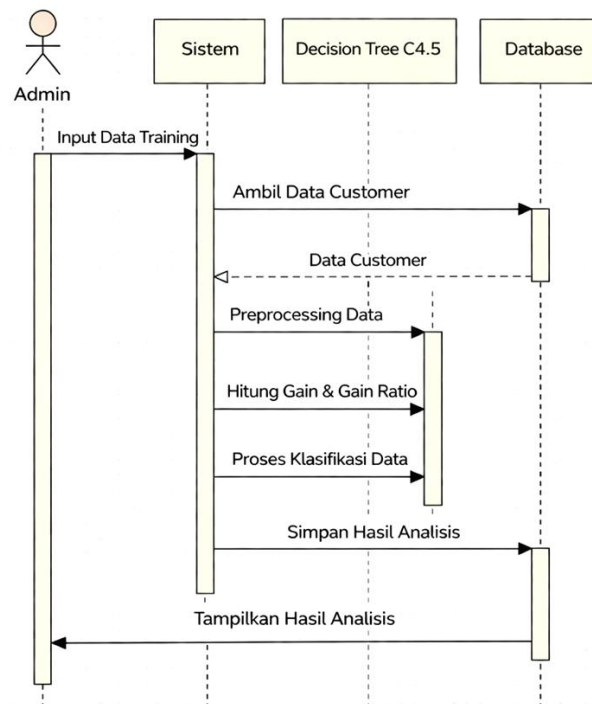
*Gambar 3.4 Use Case Diagram*

Use Case Diagram pada Gambar diatas menggambarkan interaksi antara aktor dengan sistem analisis customer churn yang dirancang. Diagram ini melibatkan dua aktor, yaitu Admin dan Manager, yang memiliki peran berbeda dalam penggunaan sistem. Admin berperan sebagai pengguna yang menjalankan sistem secara operasional, mulai dari melakukan input ID customer sebagai kata kunci pencarian data pelanggan, hingga menjalankan proses analisis customer churn. Proses analisis customer churn mencakup pengambilan data pelanggan dari basis data dan menghasilkan klasifikasi status pelanggan churn atau tidak churn. Pada use case ini terdapat relasi include yang menunjukkan bahwa proses input ID customer dan pengambilan data pelanggan merupakan bagian yang tidak terpisahkan dari proses analisis customer churn.

Manager tidak terlibat dalam proses input data maupun analisis, namun hanya memiliki akses untuk melihat hasil klasifikasi customer churn serta hasil report customer churn. Report customer churn ini disajikan sebagai output akhir sistem yang berfungsi sebagai bahan evaluasi dan pendukung pengambilan keputusan bagi pihak manajemen. Dengan adanya use case diagram ini, diharapkan dapat memberikan gambaran yang jelas mengenai fungsi sistem, alur interaksi antar aktor, serta peran sistem dalam mendukung proses analisis dan pengambilan keputusan terkait customer churn.

### 3.7.4 Sequence Diagram

a) Admin



*Gambar 3.5 Sequence Diagram Proses Analisis Customer Churn oleh Admin*

Sequence diagram pada gambar tersebut menggambarkan alur interaksi antara aktor dan komponen sistem dalam proses analisis customer churn

menggunakan algoritma Decision Tree C4.5. Diagram ini melibatkan beberapa komponen utama, yaitu Admin, Sistem, Decision Tree C4.5, dan Database.

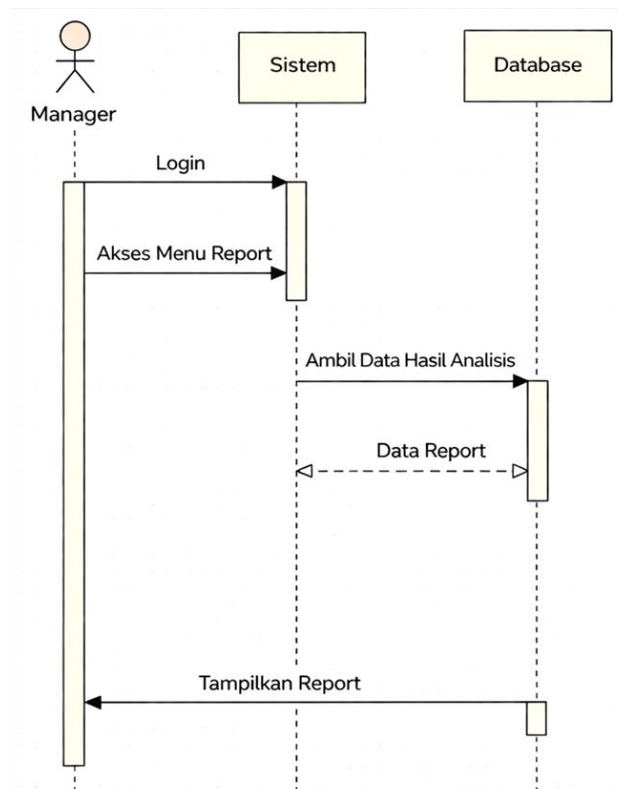
Proses dimulai dari Admin yang melakukan input data training ke dalam sistem. Data yang dimasukkan berupa data pelanggan yang akan digunakan sebagai dasar dalam proses analisis customer churn. Setelah data diterima, Sistem akan melakukan tahap preprocessing data untuk menyesuaikan dan membersihkan data agar siap digunakan dalam proses perhitungan.

Selanjutnya, Sistem melakukan proses perhitungan dengan metode Decision Tree C4.5, yang meliputi perhitungan nilai information gain dan gain ratio untuk menentukan atribut terbaik dalam pembentukan model klasifikasi. Proses ini dilakukan oleh komponen Decision Tree C4.5 sebagai inti dari analisis.

Setelah proses perhitungan selesai, Sistem melanjutkan ke tahap klasifikasi data, yaitu mengelompokkan data pelanggan ke dalam kategori churn atau tidak churn berdasarkan model yang telah terbentuk. Hasil klasifikasi tersebut kemudian dikirimkan kembali kepada Sistem. Data hasil analisis selanjutnya disimpan ke dalam Database sebagai bentuk penyimpanan hasil klasifikasi yang dapat digunakan kembali pada proses berikutnya.

Dengan demikian, sequence diagram ini menggambarkan alur komunikasi antar komponen sistem secara berurutan, mulai dari input data oleh Admin, proses analisis menggunakan algoritma C4.5, hingga penyimpanan hasil klasifikasi ke dalam database.

## b) Manager



*Gambar 3.6 Sequence Diagram Melihat Report oleh Manager*

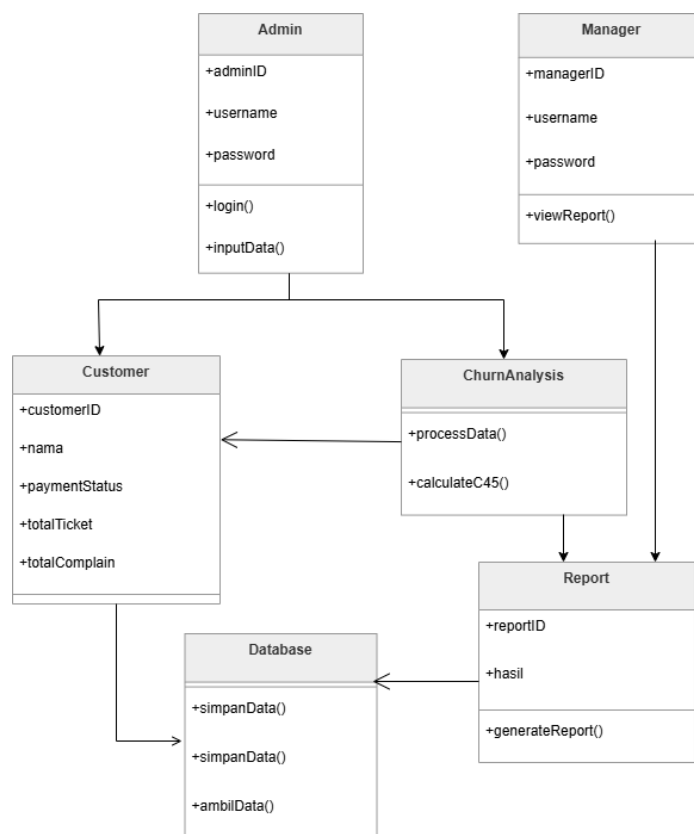
Sequence diagram pada gambar tersebut menggambarkan alur interaksi antara Manager dengan sistem dalam proses melihat hasil laporan customer churn. Diagram ini melibatkan tiga komponen utama, yaitu Manager, Sistem, dan Database.

Proses dimulai dari Manager dengan melakukan login ke dalam sistem sebagai tahap awal untuk mengakses fitur yang tersedia. Setelah berhasil login, Manager mengakses menu report untuk melihat hasil analisis customer churn yang telah dihasilkan sebelumnya. Selanjutnya, Sistem menerima permintaan dari Manager dan melakukan pengambilan data report dari Database. Data yang diambil merupakan hasil analisis customer churn yang telah diproses sebelumnya menggunakan algoritma Decision Tree C4.5.

Setelah data berhasil diambil, Sistem kemudian menampilkan report kepada Manager dalam bentuk informasi hasil klasifikasi customer churn. Manager dapat melihat hasil tersebut sebagai bahan evaluasi dan pengambilan keputusan.

Dengan demikian, sequence diagram ini menggambarkan alur komunikasi antara Manager dan sistem yang berfokus pada proses pengambilan dan penyajian data hasil analisis, tanpa melibatkan proses perhitungan atau pengolahan data secara langsung oleh Manager.

### 3.7.5 Class Diagram



*Gambar 3.7 Class Diagram*

Class diagram pada gambar tersebut menggambarkan struktur sistem analisis customer churn yang dibangun dalam penelitian ini. Diagram ini terdiri dari enam kelas utama, yaitu Admin, Manager, Customer, ChurnAnalysis, Report, dan Database yang saling berhubungan dalam mendukung proses analisis dan penyajian hasil.

Kelas Admin berperan sebagai pengguna utama yang memiliki hak akses untuk mengelola data serta menjalankan proses analisis. Admin memiliki atribut berupa adminID, username, dan password, serta method login() dan inputData() yang digunakan untuk mengelola data pelanggan dalam sistem.

Kelas Manager berperan sebagai pengguna yang hanya memiliki akses untuk melihat hasil laporan. Manager memiliki atribut managerID, username, dan password, serta method viewReport() yang digunakan untuk mengakses laporan hasil analisis customer churn.

Kelas Customer merepresentasikan data pelanggan yang digunakan sebagai input dalam proses analisis. Kelas ini memiliki atribut seperti customerID, nama, paymentStatus, totalTicket, dan totalComplain. Data pelanggan ini dikelola oleh Admin dan menjadi dasar dalam proses analisis churn.

Kelas Churn Analysis merupakan komponen inti dalam sistem yang berfungsi untuk melakukan proses analisis menggunakan algoritma Decision Tree C4.5. Kelas ini memiliki method processData() dan calculateC45() yang digunakan untuk mengolah data pelanggan dan menghasilkan klasifikasi customer churn.

Kelas Report berfungsi untuk menyimpan dan menyajikan hasil analisis dalam bentuk laporan. Kelas ini memiliki atribut reportID dan hasil, serta method generateReport() yang digunakan untuk menghasilkan laporan berdasarkan hasil klasifikasi.

Kelas Database berperan sebagai media penyimpanan data dalam sistem. Database menyimpan data pelanggan serta hasil analisis yang telah dihasilkan. Kelas ini memiliki method simpanData() dan ambilData() yang digunakan dalam proses penyimpanan dan pengambilan data.

Hubungan antar kelas dalam diagram ini menunjukkan bahwa Admin mengelola data Customer serta menjalankan proses pada ChurnAnalysis. Selanjutnya, ChurnAnalysis menggunakan data Customer untuk menghasilkan Report. Manager berinteraksi dengan Report untuk melihat hasil analisis. Data pelanggan dan laporan yang dihasilkan disimpan dalam Database sebagai penyimpanan terpusat.

### 3.8 Implementasi Metode C.45

#### 3.8.1 Hasil perhitungan algoritma C4.5

Pada tahap ini dilakukan perhitungan manual algoritma Decision Tree C4.5 untuk menentukan atribut yang paling berpengaruh dalam mengklasifikasikan pelanggan ke dalam kategori churn dan tidak churn. Dataset yang digunakan berjumlah 250 data pelanggan yang telah dianonimkan.

a. Perhitungan Entropy awal

$$\text{Rumus Entropy : } \mathbf{Entropy (S)} = - \sum_{i=1}^n P_i \log_2 P_i$$

Diketahui :

Total Churn (Ya) = 119

Total Tidak Churn (Tidak) = 131

Total = 250

1) Hitung Probabilitas

$$P_{ya} = \frac{119}{250} = 0,476$$

$$P_{tidak} = \frac{131}{250} = 0,524$$

2) Hitung  $\log_2$  masing-masing

Pakai rumus :  $\log_2(x) = \frac{\ln(x)}{\ln(2)}$

$$\begin{aligned} \bullet \log_2(p_{ya}) &= \frac{\ln(0,476)}{0,693} \rightarrow \log_2(0,476) = \frac{-0,742}{0,693} \\ &\rightarrow \log_2(0,476) = -1,070 \end{aligned}$$

$$\begin{aligned} \bullet \log_2(p_{tidak}) &= \frac{\ln(0,524)}{0,693} \rightarrow \log_2(0,524) = \frac{-0,646}{0,693} \\ &\rightarrow \log_2(0,524) = -0,933 \end{aligned}$$

3) Masukkan ke rumus *Entropy*

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

$$Entropy(S) = -(p_{ya} \times \log_2(p_{ya}) + p_{tidak} \times \log_2(p_{tidak}))$$

$$\begin{aligned} Entropy(S) &= -(0,476 \times (-1,070) + 0,524 \times (-0,933)) \\ &= -(-0,509) + (-0,489) \end{aligned}$$

$$Entropy(S) = -(-0,998) \rightarrow 0,998$$

Nilai entropy mendekati 1 menunjukkan bahwa dataset masih bersifat heterogen sehingga diperlukan pemisahan atribut untuk membentuk pohon keputusan.

b. Perhitungan pada setiap atribut

1. Atribut payment status

Distribusi data :

Tabel 4.1 Distribusi atribut payment status

| Payment status | Tidak | Ya | Total |
|----------------|-------|----|-------|
| Tepat          | 88    | 54 | 142   |
| Menunggak      | 43    | 65 | 108   |

Hitung probabilitasnya :

- *Entropy* Tepat

$$p_{ya} = \frac{54}{142} = 0.380$$

$$p_{tidak} = \frac{88}{142} = 0.620$$

Hitung log nya :

$$\log_2(p_{ya}) = -1.395$$

$$\log_2(p_{tidak}) = -0.689$$

Masukkan ke rumus Entropy :

$$Entropy \text{ (Tepat)} = -(p_{ya} \times \log_2(p_{ya}) + p_{tidak} \times \log_2(p_{tidak}))$$

$$\begin{aligned} Entropy \text{ (Tepat)} &= -(0.380 \times (-1.395) + 0.620 \times (-0.689)) \\ &= -(-0.530) + (-0.427) \end{aligned}$$

$$Entropy \text{ (Tepat)} = 0.958$$

- *Entropy* Menunggak

$$p_{ya} = \frac{65}{108} = 0.602$$

$$p_{tidak} = \frac{43}{108} = 0.398$$

Hitung log nya :

$$\log_2(p_{ya}) = -0.736$$

$$\log_2(p_{tidak}) = -1.329$$

Masukkan ke rumus Entropy :

$$\text{Entropy (Menunggak)} = -(p_{ya} \times \log_2(p_{ya}) + p_{tidak} \times \log_2(p_{tidak}))$$

$$\begin{aligned} \text{Entropy (Menunggak)} &= -(0.602 \times (-0.736) + 0.398 \times (-1.329)) \\ &= -(-0.443) + (-0.529) \end{aligned}$$

$$\text{Entropy (Menunggak)} = 0.972$$

- **Hitung Gain**

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

$$\text{Bagian 1} \quad \frac{142}{250} = 0.568$$

$$0.568 \times 0.958 = 0.544$$

$$\text{Bagian 2} \quad \frac{108}{250} = 0.432$$

$$0.432 \times 0.971 = 0.420$$

$$\text{Jumlahkan} \quad 0.544 + 0.420 = 0.964$$

Kurangkan dari *Entropy Total*

$$\text{Gain} = 0.998 - 0.964$$

$$\rightarrow \text{Gain} = 0.034$$

- **Hitung Split Information**

$$\text{Split Information} = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \left( \frac{|S_i|}{|S|} \right)$$

$$\text{Split Information} = - \left( \frac{142}{250} \right) \log_2 \left( \frac{142}{250} \right) + \left( \frac{108}{250} \right) \log_2 \left( \frac{108}{250} \right)$$

$$\text{Split Information} = 0.989$$

- **Hitung Gain Ratio**

$$\text{Gain Ratio} = \frac{\text{Information Gain}}{\text{Split Information}}$$

$$\text{Gain Ratio} = \frac{0.034}{0.989}$$

$$\text{Gain Ratio} = \mathbf{0.034}$$

2. Atribut total ticket

Atribut dikategorikan :

- $\leq 2 = \text{Rendah}$
- $\geq 2 = \text{Tinggi}$

Distribusi :

Tabel 4.2 Distribusi atribut total ticket

| Total ticket | Tidak | Ya | Total |
|--------------|-------|----|-------|
| Rendah       | 97    | 28 | 125   |
| Tinggi       | 34    | 91 | 125   |

$$\text{Entropy Rendah} = 0.764$$

$$\text{Entropy Tinggi} = 0.846$$

- Hitung *Gain* :

$$\text{Gain} = 0.998 - \left( \frac{125}{250} \times 0.764 \right) + \left( \frac{125}{250} \times 0.846 \right)$$

$$\text{Gain} = \mathbf{0.193}$$

- Hitung *Split Information*

$$\text{Split Information} = - \left( \frac{125}{250} \right) \log_2 \left( \frac{125}{250} \right) + \left( \frac{125}{250} \right) \log_2 \left( \frac{125}{250} \right)$$

$$\text{Split Information} = \mathbf{1}$$

- Hitung *Gain Ratio*

$$\text{Gain Ratio} = \frac{\text{Information Gain}}{\text{Split Information}}$$

$$\text{Gain Ratio} = \frac{0.193}{1}$$

$$\text{Gain Ratio} = \mathbf{0.193}$$

### 3. Atribut total complain

Dikategorikan :

- $\leq 2 =$  Rendah
- $\geq 2 =$  Tinggi

Tabel 4.3 Distribusi atribut total complain

| Total Complain | Tidak | Ya | Total |
|----------------|-------|----|-------|
| Rendah         | 101   | 33 | 134   |
| Tinggi         | 30    | 86 | 116   |

$$\text{Entropy Rendah} = 0.805$$

$$\text{Entropy Tinggi} = 0.824$$

- Hitung *Gain* :

$$\text{Gain} = 0.998 - \left( \frac{134}{250} \times 0.805 \right) + \left( \frac{116}{250} \times 0.824 \right)$$

$$\text{Gain} = 0.185$$

- Hitung *Split Information*

$$\text{Split Information} = - \left( \frac{134}{250} \right) \log_2 \left( \frac{134}{250} \right) + \left( \frac{116}{250} \right) \log_2 \left( \frac{116}{250} \right)$$

$$\text{Split Information} = 0.997$$

- Hitung *Gain Ratio*

$$\text{Gain Ratio} = \frac{\text{Information Gain}}{\text{Split Information}}$$

$$\text{Gain Ratio} = \frac{0.185}{0.997}$$

$$\text{Gain Ratio} = 0.186$$

#### c. Perbandingan Gain Ratio

Berdasarkan perhitungan terhadap atribut:

- Payment Status

- Total Ticket
- Total Complain

Tabel 4.4 Perbandingan gain ratio

| Atribut        | Gain  | Split Info | Gain Ratio |
|----------------|-------|------------|------------|
| Payment status | 0.034 | 0.989      | 0.034      |
| Total ticket   | 0.193 | 1.000      | 0.193      |
| Total Complain | 0.185 | 0.997      | 0.186      |

Berdasarkan tabel di atas, atribut **Total ticket** memiliki nilai gain tertinggi yaitu 0.193.

#### d. Penentuan Root Node

Karena total ticket memiliki nilai Information Gain tertinggi yaitu 0.193, maka atribut tersebut dipilih sebagai root node dalam pembentukan pohon keputusan algoritma C4.5

#### e. Split tahap pertama (Penentuan root)

Atribut yang diuji : payment status, total ticket, total complain

Ringkasan Gain Ratio :

Tabel 4.5 Split tahap pertama

| Atribut        | Gain Ratio |
|----------------|------------|
| Payment status | 0.034      |
| Total ticket   | 0.193      |
| Total Complain | 0.186      |

Karena total ticket memiliki Gain Ratio tertinggi (0.193), maka dipilih sebagai root node. Struktur awal, total ticket dikategorikan dengan  $\leq 2 =$  Rendah,  $\geq 2 =$  Tinggi

Split tahap kedua

a. Cabang total ticket = rendah

Distribusi:

Tidak = 97

Ya = 28

Total = 125

1. Entropy cabang rendah

$$Entropy = -(0.776 \log_2 0.776 + 0.224 \log_2 0.224)$$

$$Entropy = 0.764$$

Uji atribut total complain

Tabel 4.6 Uji atribut total complain entropy rendah

| Complain | Tidak | Ya | Total |
|----------|-------|----|-------|
| Rendah   | 80    | 15 | 95    |
| Tinggi   | 17    | 13 | 30    |

Entropy rendah = 0.629

Entropy tinggi = 0.987

Information gain :

$$Gain = 0.764 - \left( \frac{95}{125} \times 0.629 \right) + \left( \frac{30}{125} \times 0.987 \right)$$

$$Gain = 0.148$$

Split Information :

$$Split Information = - \left( \frac{95}{125} \right) \log_2 \left( \frac{95}{125} \right) + \left( \frac{30}{125} \right) \log_2 \left( \frac{30}{125} \right)$$

$$Split Information = 0.795$$

Gain Ratio :

$$Gain Ratio = \frac{0.148}{0.795}$$

$$\mathbf{Gain\ Ratio = 0.186}$$

Karena masih terdapat campuran kelas, maka Total Complain dipilih sebagai split kedua pada cabang ini

b. Cabang total ticket = tinggi

Distribusi:

Tidak Churn = 35

Churn = 90

Total = 125

1. Entropy cabang tinggi

$$\mathbf{Entropy = 0.846}$$

Uji atribut total complain

Tabel 4.7 Uji atribut total complain entropy tinggi

| <b>Complain</b> | <b>Tidak</b> | <b>Ya</b> | <b>Total</b> |
|-----------------|--------------|-----------|--------------|
| Rendah          | <b>15</b>    | <b>20</b> | <b>35</b>    |
| Tinggi          | <b>20</b>    | <b>70</b> | <b>90</b>    |

Entropy rendah = 0.985

Entropy tinggi = 0.741

*Information gain :*

$$\mathbf{Gain = 0.162}$$

*Split Information :*

$$\mathbf{Split\ Information = 0.863}$$

**Gain Ratio :**

$$Gain\ Ratio = 0.187$$

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Hasil Penelitian**

##### **4.1.1 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

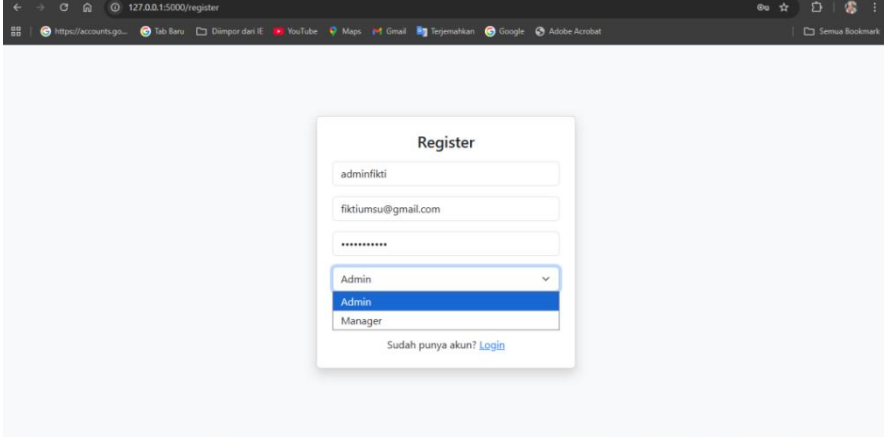
Jika Total ticket Tinggi:

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



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

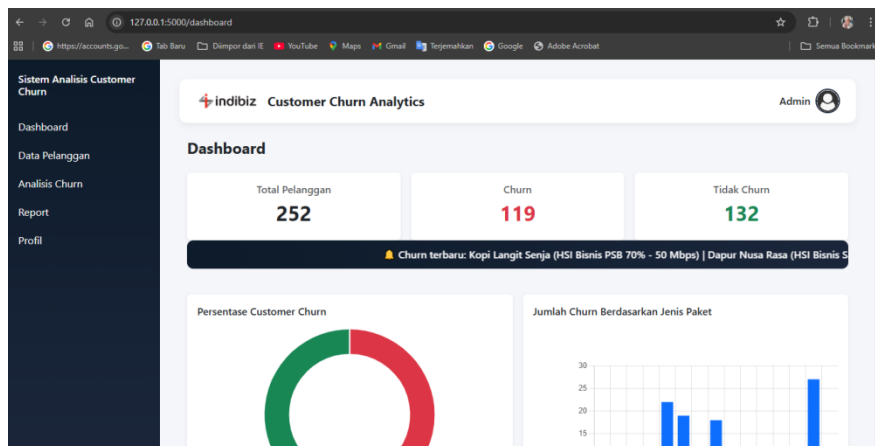
A screenshot of a web browser displaying a registration form. The browser's address bar shows the URL 'https://accounts.google.com/...'. The form is titled 'Register' and contains the following fields: a text input for the username 'adminfikti', a text input for the email 'fiktiums@gmail.com', a password input field with masked characters '\*\*\*\*\*', and a dropdown menu for selecting a role. The dropdown menu is open, showing 'Admin' as the selected option and 'Manager' as another available option. Below the form, there is a link that says 'Sudah punya akun? Login'.

Gambar 4.3 Halaman registrasi

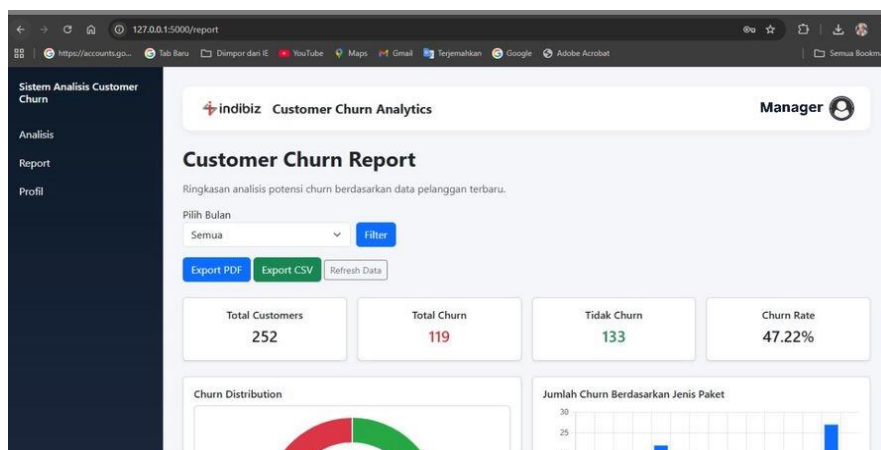
Halaman registrasi digunakan untuk melakukan pendaftaran akun baru. Pengguna diminta mengisi data yang diperlukan untuk membuat akun agar dapat mengakses sistem.

#### c. Halaman dashboard

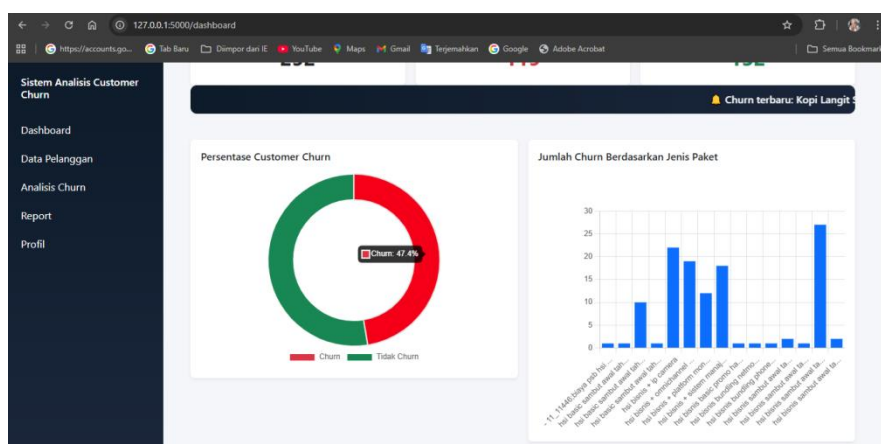
Halaman dashboard menampilkan ringkasan informasi terkait jumlah pelanggan, jumlah churn, dan tidak churn. Dashboard juga menampilkan visualisasi data dalam bentuk grafik dan diagram untuk memberikan gambaran umum kondisi pelanggan. Pada halaman dashboard juga menampilkan data customer dalam tabel pagination.



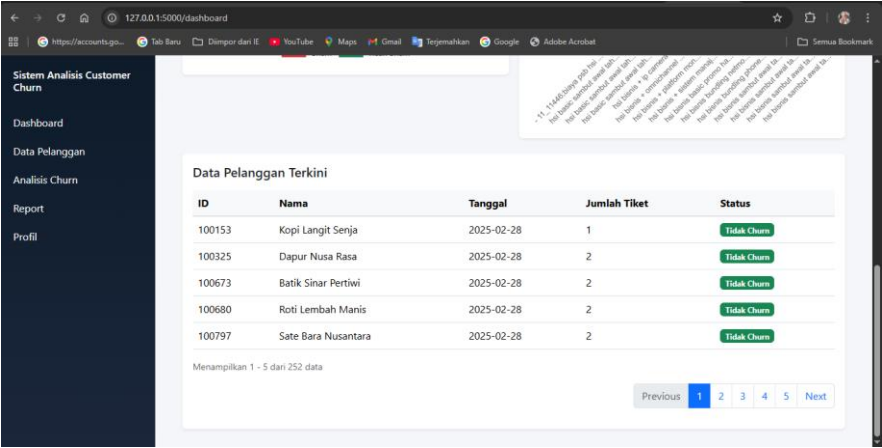
Gambar 4.4 Dashboard admin



Gambar 4.5 Dashboard manager



Gambar 4.6 Bagian tengah dashboard



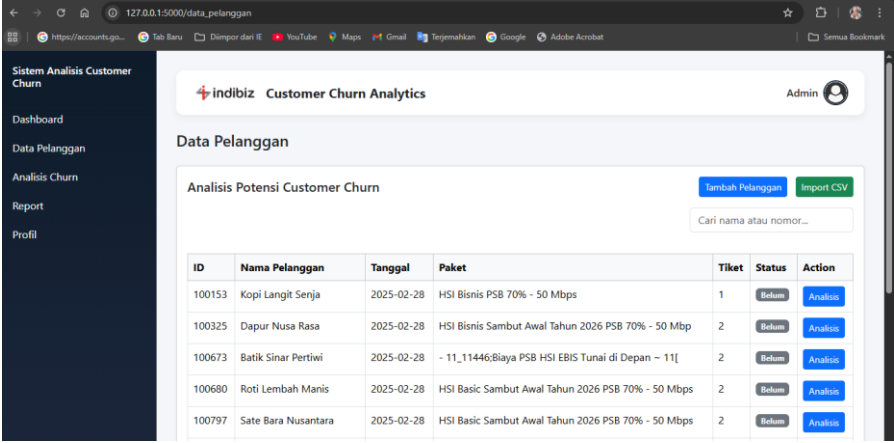
The screenshot shows a web browser window with the URL 127.0.0.1:5000/dashboard. The page title is 'Sistem Analisis Customer Churn'. The left sidebar contains navigation links: Dashboard, Data Pelanggan, Analisis Churn, Report, and Profil. The main content area displays a table titled 'Data Pelanggan Terkini' with the following data:

| ID     | Nama                | Tanggal    | Jumlah Tiket | Status      |
|--------|---------------------|------------|--------------|-------------|
| 100153 | Kopi Langit Senja   | 2025-02-28 | 1            | Tidak Churn |
| 100325 | Dapur Nusa Rasa     | 2025-02-28 | 2            | Tidak Churn |
| 100673 | Batik Sinar Pertiwi | 2025-02-28 | 2            | Tidak Churn |
| 100680 | Roti Lembah Manis   | 2025-02-28 | 2            | Tidak Churn |
| 100797 | Sate Bara Nusantara | 2025-02-28 | 2            | Tidak Churn |

Below the table, it indicates 'Menampilkan 1 - 5 dari 252 data' and includes pagination controls for pages 1 through 5.

Gambar 4.7 Bagian bawah dashboard

#### d. Halaman Data Pelanggan



The screenshot shows a web browser window with the URL 127.0.0.1:5000/data\_pelanggan. The page title is 'Sistem Analisis Customer Churn'. The left sidebar contains navigation links: Dashboard, Data Pelanggan, Analisis Churn, Report, and Profil. The main content area displays a table titled 'Data Pelanggan' with the following data:

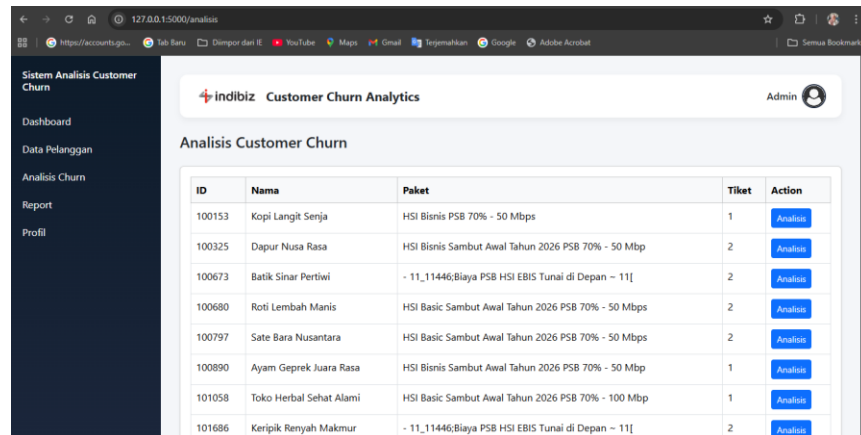
| ID     | Nama Pelanggan      | Tanggal    | Paket  | Tiket | Status | Action   |
|--------|---------------------|------------|--|-------|--------|----------|
| 100153 | Kopi Langit Senja   | 2025-02-28 | HSI Bisnis PSB 70% - 50 Mbps                       | 1     | Belum  | Analisis |
| 100325 | Dapur Nusa Rasa     | 2025-02-28 | HSI Bisnis Sambut Awal Tahun 2026 PSB 70% - 50 Mbp | 2     | Belum  | Analisis |
| 100673 | Batik Sinar Pertiwi | 2025-02-28 | - 11_11446:Biaya PSB HSI EBIS Tunai di Depan ~ 11[ | 2     | Belum  | Analisis |
| 100680 | Roti Lembah Manis   | 2025-02-28 | HSI Basic Sambut Awal Tahun 2026 PSB 70% - 50 Mbps | 2     | Belum  | Analisis |
| 100797 | Sate Bara Nusantara | 2025-02-28 | HSI Basic Sambut Awal Tahun 2026 PSB 70% - 50 Mbps | 2     | Belum  | Analisis |

The page also includes a search bar with the placeholder text 'Cari nama atau nomor...', a 'Tambah Pelanggan' button, and an 'Import CSV' button.

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 navigasi.

### e. Halaman analisis



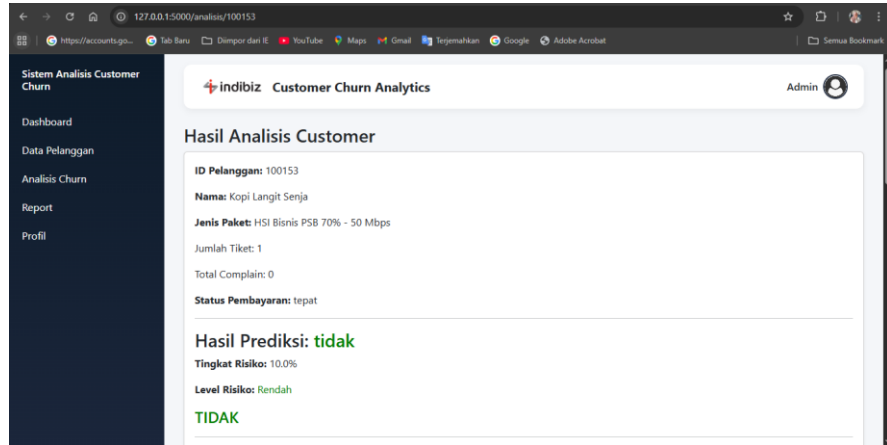
| ID     | Nama                    | Paket  | Tiket | Action   |
|--------|-------------------------|--|-------|----------|
| 100153 | Kopi Langit Senja       | HSI Bisnis PSB 70% - 50 Mbps                       | 1     | Analisis |
| 100325 | Dapur Nusa Rasa         | HSI Bisnis Sambut Awal Tahun 2026 PSB 70% - 50 Mbp | 2     | Analisis |
| 100673 | Batik Sinar Pertiwi     | - 11_11446;Blaya PSB HSI EBIS Tunai di Depan - 11[ | 2     | Analisis |
| 100680 | Roti Lembah Manis       | HSI Basic Sambut Awal Tahun 2026 PSB 70% - 50 Mbps | 2     | Analisis |
| 100797 | Sate Bara Nusantara     | HSI Basic Sambut Awal Tahun 2026 PSB 70% - 50 Mbps | 2     | Analisis |
| 100890 | Ayam Geprek Juara Rasa  | HSI Bisnis Sambut Awal Tahun 2026 PSB 70% - 50 Mbp | 1     | Analisis |
| 101058 | Toko Herbal Sehat Alami | HSI Basic Sambut Awal Tahun 2026 PSB 70% - 100 Mbp | 1     | Analisis |
| 101686 | Keripik Renyah Makmur   | - 11_11446;Blaya PSB HSI EBIS Tunai di Depan - 11[ | 2     | Analisis |

Gambar 4.9 Halaman analisis

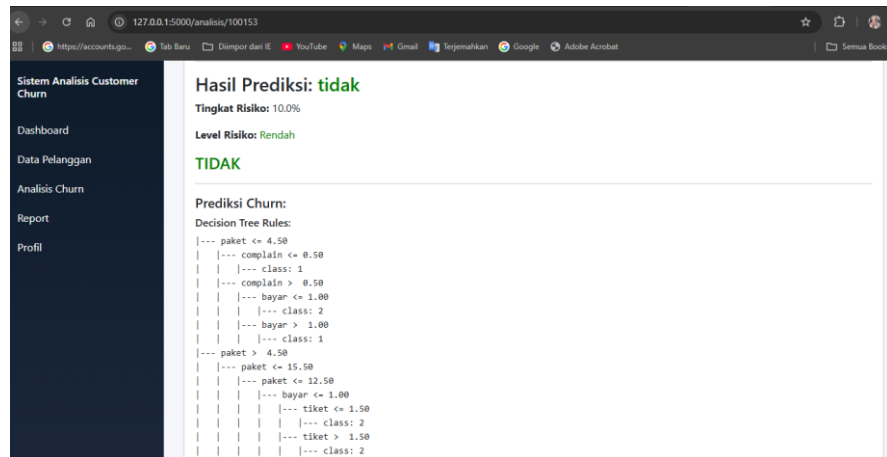
Halaman analisis digunakan oleh Admin untuk melakukan proses klasifikasi customer churn berdasarkan ID pelanggan yang dimasukkan. Sistem kemudian menampilkan hasil klasifikasi beserta aturan keputusan (decision rules) yang digunakan dalam proses analisis.

### f. Halaman analisis customer

Pada halaman hasil Analisis Customer Churn ditampilkan informasi detail mengenai pelanggan yang telah dilakukan proses klasifikasi menggunakan algoritma Decision Tree. Halaman ini memuat data pelanggan seperti ID pelanggan, nama pelanggan, jenis paket yang digunakan, total ticket, total complain, serta payment status. Informasi tersebut digunakan sebagai dasar dalam proses prediksi churn.



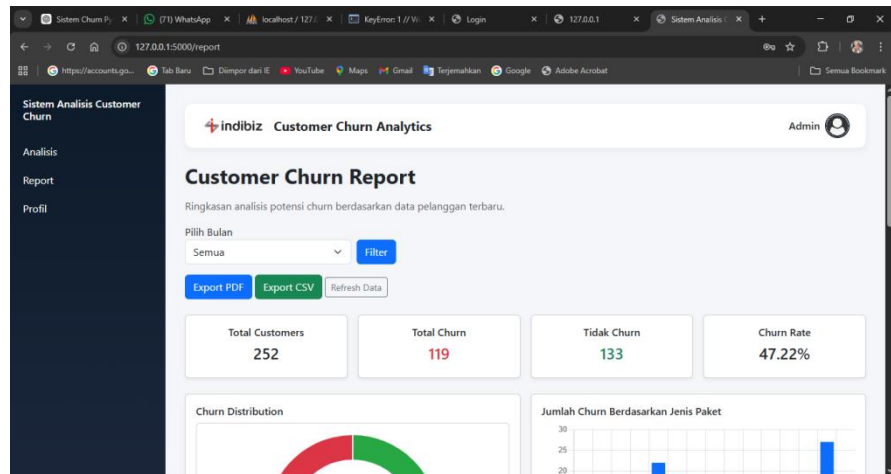
Gambar 4.10 Halaman analisis



Gambar 4.11 Tampilan rules analisis

Di bagian bawah halaman juga ditampilkan aturan (rules) Decision Tree yang dihasilkan dari proses pelatihan model. Aturan ini menunjukkan percabangan keputusan berdasarkan variabel-variabel seperti jenis paket, total ticket, total complain, dan payment status.

## g. Halaman report

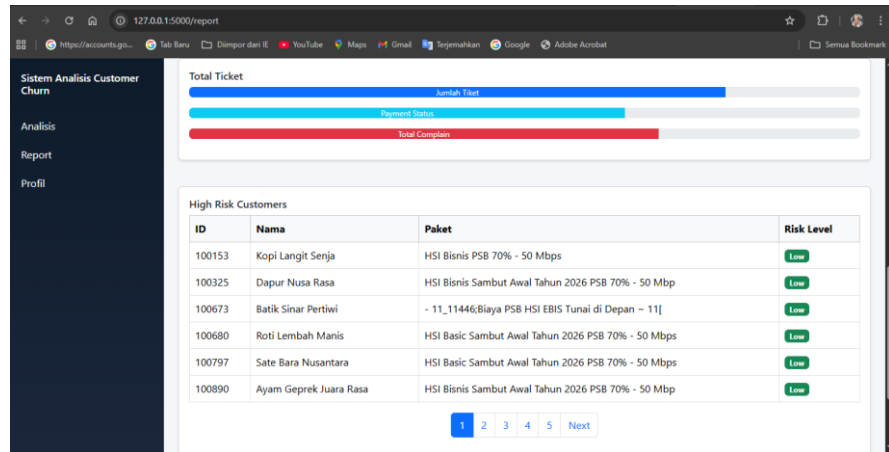


Gambar 4.12 Halaman report

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.



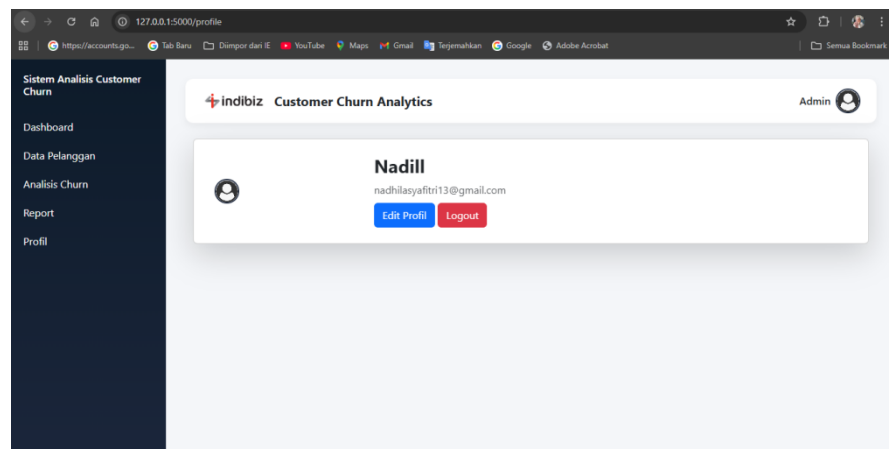
Gambar 4.13 Bagian tengah halaman report



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.

Melalui proses pengujian ini, dilakukan pemeriksaan terhadap kinerja sistem dalam menerima, memproses, dan menghasilkan output berupa prediksi churn pelanggan. Pengujian difokuskan pada aspek fungsionalitas sistem, termasuk validasi input data, proses analisis, serta akurasi dan kelayakan hasil yang ditampilkan. Dengan demikian, sistem diharapkan mampu memberikan hasil analisis yang relevan dan dapat digunakan sebagai dasar pengambilan keputusan

Selain pengujian terhadap fungsi sistem, dilakukan pula peninjauan terhadap kesiapan perangkat keras dan perangkat lunak yang digunakan dalam proses pengembangan dan implementasi. Hal ini bertujuan untuk memastikan bahwa sistem dapat dijalankan dalam lingkungan operasional yang stabil serta meminimalkan potensi gangguan teknis selama penggunaan.

##### 1. Hardware

Merk : Laptop HP ProBook 430 G5

RAM : Installed RAM 8.00 GB

Processor: Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 1.99 GHz

SSD 250 GB

System type : 64-bit operating system, x64-based processor

## 2. Software

Code editor : Visual Studio Code

Web server : Xampp

Databse : MySQL, phpMyAdmin

Browser : Google Chrome

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

| <b>Fitur</b> | <b>Skenario Pengujian</b>      | <b>Output yang diharapkan</b>    | <b>Status</b> |
|--------------|--------------------------------|----------------------------------|---------------|
| 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  $\pm 250$  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      | Ticket | Complain |
|--------|------------------------|------------|--|-----------|-------|------------|--------|----------|
| 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             | 2025-      | HSI Bisnis                                   | menunggak | ya    | 0812000214 | 1      | 0        |

|        |                        |            |   |           |       |            |   |   |
|--------|------------------------|------------|---|-----------|-------|------------|---|---|
|        | Mandiri                | 09-29      | Sambut Awal Tahun 2026 PSB 70% - 50 Mbps            |           |       |            |   |   |
| 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 |
| 112795 | CV Sejahtera Jaya      | 2025-10-08 | HSI Bisnis Sambut Awal Tahun 2026 PSB 70% - 50 Mbps | tepat     | tidak | 0812000223 | 1 | 0 |
| 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    | Tidak Churn | Churn |
| Aktual      |             |       |
| 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

$$accuracy = \frac{TP + TN}{Total}$$

$$accuracy = \frac{14 + 20}{50}$$

$$accuracy = \frac{34}{50}$$

$$accuracy \mathbf{0.68} \rightarrow \mathbf{68\%}$$

#### 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.

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{14 + 4}{14}$$

$$= \frac{14}{18}$$

$$Precision = 0.778 \rightarrow 77.8\%$$

## 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.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{14}{14 + 12}$$

$$Recall = \frac{14}{26}$$

$$Recall = 0.538 \rightarrow 53.8\%$$

## 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.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 = 2 \times \frac{0.778 \times 0.538}{0.778 + 0.538}$$

$$F1 = \frac{0.836}{1.316}$$

$$= 0.6356 \rightarrow 0.636$$

$$\mathbf{F1 = 0.6356 \rightarrow 0.636}$$

#### 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.

#### 4.2.3 Analisis faktor penentu customer churn

Berdasarkan hasil perhitungan algoritma Decision Tree C4.5 pada subbab sebelumnya, diperoleh bahwa atribut Total Ticket memiliki nilai Gain Ratio tertinggi sebesar 0,193 sehingga dipilih sebagai root node dalam pembentukan pohon keputusan. Hal ini menunjukkan bahwa jumlah tiket layanan yang dibuat oleh pelanggan menjadi faktor paling dominan dalam menentukan potensi customer churn.

Secara umum, pelanggan dengan jumlah tiket yang tinggi menunjukkan adanya intensitas interaksi yang lebih besar terhadap layanan, yang dapat

mengindikasikan adanya permasalahan teknis atau ketidakpuasan terhadap layanan. Kondisi ini meningkatkan kemungkinan pelanggan untuk menghentikan penggunaan layanan (churn). Dengan demikian, atribut total ticket dapat dijadikan indikator awal dalam mendeteksi potensi churn pada pelanggan bisnis.

Atribut Total Complain juga memiliki pengaruh yang signifikan dalam proses klasifikasi, dengan nilai Gain Ratio sebesar 0,186. Dalam struktur pohon keputusan, atribut ini berperan sebagai pemisah lanjutan setelah total ticket. Hal ini menunjukkan bahwa tingkat keluhan pelanggan turut memengaruhi keputusan akhir dalam klasifikasi churn. Semakin tinggi jumlah complain yang diajukan, maka semakin besar kemungkinan pelanggan dikategorikan sebagai churn.

Sementara itu, atribut Payment Status memiliki nilai Gain Ratio paling rendah dibandingkan dua atribut lainnya, yaitu sebesar 0,034. Atribut ini tidak menjadi root node, tetapi tetap digunakan dalam cabang tertentu pada pohon keputusan. Hal ini menunjukkan bahwa status pembayaran memiliki pengaruh, namun bukan faktor utama dalam menentukan potensi churn pada dataset yang digunakan dalam penelitian ini.

Berdasarkan struktur pohon keputusan akhir, dapat disimpulkan bahwa pola utama yang terbentuk dalam model klasifikasi adalah bahwa pelanggan dengan total ticket tinggi dan total complain tinggi cenderung diklasifikasikan sebagai churn. Pola ini memberikan gambaran bahwa tingkat interaksi layanan dan keluhan pelanggan menjadi indikator penting dalam mengidentifikasi risiko churn pada pelanggan bisnis Indibiz Telkom Regional I.

#### 4.2.4 Analisis implementasi sistem

Implementasi sistem analisis customer churn dalam penelitian ini menunjukkan bahwa model klasifikasi Decision Tree C4.5 yang telah dibangun tidak hanya bersifat konseptual, tetapi dapat diterapkan secara operasional dalam bentuk aplikasi berbasis web. Sistem dirancang untuk mengintegrasikan proses klasifikasi dengan penyajian informasi yang terstruktur sehingga dapat digunakan oleh pihak operasional maupun manajerial.

Halaman dashboard berfungsi sebagai tampilan utama yang menyajikan ringkasan jumlah total pelanggan, jumlah pelanggan churn, dan tidak churn dalam bentuk visualisasi grafik dan diagram. Penyajian informasi secara visual memberikan gambaran umum kondisi pelanggan secara cepat dan sistematis. Melalui dashboard, pengguna dapat memahami distribusi data tanpa harus melakukan proses analisis secara manual, sehingga halaman ini berperan sebagai media monitoring awal terhadap potensi churn.

Halaman data pelanggan menampilkan daftar pelanggan dalam bentuk tabel yang dilengkapi dengan fitur pagination dan pencarian. Tampilan ini memungkinkan pengguna untuk menelusuri data pelanggan secara lebih terstruktur serta melihat atribut yang digunakan dalam proses klasifikasi, seperti total ticket, total complain, dan payment status. Dengan adanya keterbukaan data ini, sistem menunjukkan hubungan yang jelas antara data pelanggan dan hasil klasifikasi yang dihasilkan oleh model.

Pada halaman analisis, Admin dapat memasukkan ID pelanggan untuk menjalankan proses klasifikasi. Sistem kemudian memproses data berdasarkan aturan yang terbentuk dari pohon keputusan dan menampilkan hasil klasifikasi

berupa status churn atau tidak churn. Selain itu, sistem juga menampilkan aturan keputusan (decision rules) yang digunakan dalam menentukan hasil tersebut. Penyajian aturan keputusan ini menjadi keunggulan sistem karena memberikan transparansi terhadap proses pengambilan keputusan. Berbeda dengan model yang bersifat black-box, Decision Tree C4.5 menghasilkan struktur keputusan yang dapat dijelaskan secara eksplisit sehingga faktor-faktor penyebab churn dapat dipahami dengan jelas.

Halaman report menyajikan rekapitulasi hasil analisis dalam bentuk statistik dan visualisasi distribusi pelanggan churn dan tidak churn. Laporan ini memberikan gambaran agregat terhadap kondisi pelanggan secara keseluruhan dan dapat digunakan sebagai dasar evaluasi berkala. Dengan adanya laporan yang terstruktur, sistem tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga sebagai media penyajian informasi yang mendukung proses pengambilan keputusan.

Halaman profil menampilkan informasi akun pengguna yang sedang login serta mencerminkan adanya mekanisme pengelolaan hak akses dalam sistem. Pembagian peran antara Admin dan Manager menunjukkan bahwa sistem telah dirancang sesuai dengan kebutuhan organisasi, di mana Admin bertanggung jawab terhadap proses operasional analisis, sedangkan Manager berfokus pada pemantauan hasil dan laporan.

Secara keseluruhan, implementasi sistem menunjukkan bahwa model Decision Tree C4.5 dapat diintegrasikan ke dalam aplikasi berbasis web yang mendukung proses klasifikasi, monitoring, dan evaluasi pelanggan secara sistematis. Sistem yang dibangun tidak hanya menghasilkan prediksi churn, tetapi juga menyediakan visualisasi data dan aturan keputusan yang dapat

dipahami secara langsung, sehingga berpotensi berfungsi sebagai pendukung keputusan dalam membantu perusahaan mengidentifikasi pelanggan berisiko churn dan menentukan langkah tindak lanjut yang tepat.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 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*).

## 5.2 Saran

Berdasarkan hasil penelitian dan evaluasi yang telah dilakukan, maka terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian dan sistem di masa mendatang. Yaitu, Penelitian selanjutnya disarankan untuk menggunakan dataset dengan jumlah data yang lebih besar dan distribusi kelas yang lebih seimbang, sehingga performa model, khususnya dalam meningkatkan nilai recall pada kelas churn, dapat lebih optimal.

Perlu dilakukan optimasi parameter (*hyperparameter tuning*) pada algoritma Decision Tree C4.5 guna meningkatkan akurasi dan kestabilan model. Selain itu, dapat dilakukan teknik pruning yang lebih mendalam untuk menghindari overfitting.

Disarankan untuk melakukan perbandingan performa dengan algoritma klasifikasi lainnya seperti Random Forest, Support Vector Machine (SVM), Logistic Regression, atau XGBoost guna mengetahui metode yang paling efektif dalam mendeteksi customer churn pada data pelanggan bisnis.

Sistem dapat dikembangkan lebih lanjut dengan menambahkan fitur rekomendasi tindakan (*retention strategy recommendation*), sehingga pelanggan yang terdeteksi memiliki risiko churn tinggi dapat langsung diberikan tindakan preventif secara sistematis. Pengembangan sistem ke arah integrasi real-time dengan database operasional perusahaan sangat disarankan, sehingga analisis churn dapat dilakukan secara otomatis dan berkelanjutan tanpa harus melalui proses input manual.

Untuk meningkatkan kualitas pengambilan keputusan, penelitian selanjutnya juga dapat mempertimbangkan penambahan variabel lain yang lebih kompleks, seperti lama berlangganan (*tenure*), nilai kontrak layanan, tingkat penggunaan layanan, atau tingkat kepuasan pelanggan

## DAFTAR PUSTAKA

- Anas, A., Zebua, A. J., & Akhmadi, A. (2025). Klasifikasi Data Kelulusan Mahasiswa Menggunakan Algoritma C.45. *Jurnal Ilmiah Media Sisfo*, 19(2), 109–115. <https://doi.org/10.33998/mediasisfo.2025.19.2.2498>
- Andy Hermawan, Nila Rusiardi Jayanti, Zia Tabaruk, Faizal Lutfi Yoga Triadi, Aji Saputra, & M.Rahmat Hidayat Syachrudin. (2024). Membangun Model Prediksi Churn Pelanggan yang Akurat. *Merkurius : Jurnal Riset Sistem Informasi Dan Teknik Informatika*, 2(6), 67–81. <https://doi.org/10.61132/mercurius.v2i6.398>
- Antoh, S., Herteno, R., Budiman, I., Kartini, D., & Mazdadi, M. I. (2025a). Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4.5. *Jurnal Sistem Informasi Bisnis*, 15(1), 60–67. <https://doi.org/10.14710/vol15iss1pp60-67>
- Antoh, S., Herteno, R., Budiman, I., Kartini, D., & Mazdadi, M. I. (2025b). Prediksi Churn Pelanggan Telekomunikasi dengan Optimalisasi Seleksi Fitur dan Tuning Hyperparameter pada Algoritma Klasifikasi C4.5. *Jurnal Sistem Informasi Bisnis*, 15(1), 60–67. <https://doi.org/10.14710/vol15iss1pp60-67>
- Ary, L. P., Tjahyanti, S., Saputra, P. S., Gitakarma, M. S., Informasi, T., Sakti, P., Rekayasa, T., & Elektronika, S. (2023). Pengembangan Sistem Informasi Pendaftaran PKL Di Universitas Panji Sakti. *Jurnal Minfo Polgan*, 12(2). <https://doi.org/10.33395/jmp.v12i2.12545>
- Ermawati, E., Bambang Irawan, & Nur Ariesanto Ramdhan. (2025). Klasifikasi Persediaan Obat Menggunakan Algoritma Decision Tree di Puskesmas Brebes. *Jurnal Ilmiah Teknik Mesin, Elektro Dan Komputer*, 5(2), 415–432.

<https://doi.org/10.51903/juritek.v5i2.5061>

Nistrina, K., & Lestari, T. A. (n.d.). *Desain Inovatif Sistem Informasi Profil Hotel Damanaka Pangalengan Berbasis Website Menggunakan UML dan Figma*.

Setia Budi, E., Rahman Kadafi, A., Kharismawan, Y., Fadillah, R., & Sasqia Putri, D. (2024a). RESOLUSI : Rekayasa Teknik Informatika dan Informasi Analisa Kepuasan Pelanggan Terhadap Layanan Aplikasi E-Commerce Menggunakan Algoritma C4.5. *Media Online*), 4(6), 542. <https://djournals.com/resolusi>

Setia Budi, E., Rahman Kadafi, A., Kharismawan, Y., Fadillah, R., & Sasqia Putri, D. (2024b). RESOLUSI : Rekayasa Teknik Informatika dan Informasi Analisa Kepuasan Pelanggan Terhadap Layanan Aplikasi E-Commerce Menggunakan Algoritma C4.5. *Media Online*), 4(6), 542. <https://djournals.com/resolusi>

Setiawan, M. azi setiawan, & Putra, S. H. (2024). Pemanfaatan Sistem Informasi Pelayanan Terpadu Dalam Meningkatkan Kepuasan Mahasiswa Politeknik Ganesha. *Jurnal Minfo Polgan*, 13(2), 1848–1857. <https://doi.org/10.33395/jmp.v13i2.14275>

Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *The Unified Modeling Language User Guide* (2nd ed.). Addison-Wesley.

## LAMPIRAN

### 1. Surat Keterangan Penetapan Judul & Dosen Pembimbing



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 174/SK/BDAN-PT/AK-Pg/PT/10/2024  
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp (061) 6622400 - 66224567 Fax (061) 6625474 - 6631003

<http://itid.umsu.ac.id> [itid@umsu.ac.id](mailto:itid@umsu.ac.id) [umsumedan](https://www.facebook.com/umsumedan) [umsumedan](https://www.instagram.com/umsumedan) [umsumedan](https://www.youtube.com/umsumedan) [umsumedan](https://www.tiktok.com/umsumedan)

**PENETAPAN DOSEN PEMBIMBING**  
**PROPOSAL/SKRIPSI MAHASISWA**  
**NOMOR : 983/IL3-AU/UMSU-09/F/2025**

*Assalamu'alaikum Warahmatullahi Wabarakatuh*

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

**Program Studi** : Sistem Informasi  
**Pada tanggal** : 27 Oktober 2025

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

**Nama** : Nadhilah Syafitri  
**NPM** : 2209010194  
**Semester** : VII (Tujuh)  
**Program studi** : Sistem Informasi  
**Judul Proposal / Skripsi** : Analisis Pola Kunjungan Pelanggan Studiosaja Menggunakan Algoritma Apriori untuk Strategi Promosi dan Pengembangan Sistem Rekomendasi Paket

**Dosen Pembimbing** : Rizaldi Khair, S.Kom.,M.Kom

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. Proyek Proposal / Skripsi dinyatakan " BATAL " bila tidak selesai sebelum Masa Kadalua tanggal : 27 Oktober 2026
4. Revisi judul.....

*Wassalamu'alaikum Warahmatullahi Wabarakatuh.*

Ditetapkan di : Medan  
 Pada Tanggal : 05 Jumadil Awwal 1447 H  
 27 Oktober 2025M



Dekan  
  
**Dr. A. E. Khwarizmi, M.Kom.**  
 NIDN : 0127099201

*Cc. File*

## 2. Surat Perubahan Topik/Judul Penelitian



### PERUBAHAN TOPIK/JUDUL PENELITIAN

|                                       |   |
|---------------------------------------|---|
| Nama                                  | : Nadhilah Syafitri   |
| NPM                                   | : 2209010194  |
| Nomor agenda                          | : 983/II.3-AU/UMSU-09/F/2025  |
| Tanggal persetujuan                   | : 27 Oktober 2025   |
| Topik yang disetujui Program Studi    | : Analisis Pola Kunjungan Pelanggan Studiosaja Menggunakan Algoritma Apriori untuk Strategi Promosi Pada Sistem Rekomendasi Paket |
| Nama Dosen pembimbing                 | : Rizaldy Khair, S.Kom.,M.Kom   |
| Judul yang disetujui Dosen Pembimbing | : Analisis Potensi Customer Churn Menggunakan Algoritma Decision Tree (C.45) pada Indibiz Telkom Regional I                       |

Medan, 06 Januari 2026

Ketua Program Studi

**Dr. Firahmi Rizky, M.Kom.**  
NIDN : 0116079201

a.n.Dekan  
Wakil Dekan I

**Dr. Firahmi Rizky, M.Kom.**  
NIDN : 0116079201



### 3. Surat Izin Penelitian



Nomor : 82/II.3-AU/UMSU-09/F/2026  
 Lampiran : -  
 Perihal : **IZIN PENELITIAN**

Medan, 18 Rajab 1447 H  
 07 Januari 2026 M

Kepada Yth.  
**Bapak/Ibu Pimpinan**  
**Indibiz Telkom Regional 1**  
**Graha Merah Putih Jl. Putri Hijau, kesawan, kecamatan medan barat, kota medan**

Di Tempat

*Assalamu'alaikum Warahmatullahi Wabarakatuh*

Dengan hormat, sehubungan mahasiswa kami akan menyelesaikan studi, untuk itu kami memohon kesediaan Bapak / Ibu untuk memberikan kesempatan pada mahasiswa kami melakukan riset di **Perusahaan / Instansi** yang Bapak / Ibu pimpin, guna untuk penyusunan skripsi yang merupakan salah satu persyaratan dalam menyelesaikan Program Studi Strata Satu (S-1)

Adapun Mahasiswa/i di Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara tersebut adalah:

**Nama** : Nadhilah Syafitri  
**Npm** : 2209010194  
**Jurusan** : Sistem Informasi  
**Semester** : VIII (Delapan)  
**Judul** : Analisis Potensi Customer Churn Menggunakan Algoritma Decision Tree (C4.5) pada Indibiz Telkom Regional I  
**Email** : nadhilasyafitri13@gmail.com  
**Hp/Wa** : 08116227322

Demikianlah surat kami ini, atas perhatian dan kerjasamanya yang Bapak / Ibu berikan kami ucapkan terimakasih

*Wassalamu 'alaikum Warahmatullahi Wabarakatuh*



Cc.File

  
 Dekan  
 Universitas Muhammadiyah Sumatera Utara  
 Fakultas Ilmu Komputer dan Teknologi Informasi  
**Al-Khowarizmi, M.Kom.**  
 NIDN : 012709201

#### 4. Surat Pernyataan Telah Menyelesaikan Penelitian



#### **PT. Telkom Indonesia, Tbk. (Regional 1)**

Jl. Prof. H. M. Yamin No.2, Kesawan, Kec. Medan Baru., Kota Medan,  
Sumatera Utara 20236 | Telp. (061) 4151747

Medan, 24 Februari 2026

Kepada Yth.  
Bapak/Ibu Pimpinan  
Universitas Muhammadiyah Sumatera Utara  
Fakultas Ilmu Komputer dan Teknologi Informasi  
Jurusan Sistem Informasi  
di Medan

Perihal: Surat Keterangan Selesai Penelitian

Dengan ini menyatakan bahwa:

|                             |   |
|-----------------------------|---|
| Nama                        | Nadhilah Syafitri   |
| Nomor Induk Mahasiswa (NIM) | 2209010194  |
| Jurusan                     | Sistem Informasi  |
| Judul Penelitian            | Analisis Potensi Customer Churn Menggunakan Algoritma Decision Tree (C4.5) Pada Indibiz Telkom Regional I |

Telah menyelesaikan penelitian. Mahasiswa ditempatkan di Departemen Digital Marketing Operation, Telkom Regional 1 Sumatera.

Demikian surat keterangan ini kami sampaikan. Atas perhatian dan kerja samanya, kami ucapkan terima kasih.

Hormat kami,

Sry Hazzayani  
Manajer Regional Business Service  
Telkom Regional 1 Sumatera

## 5. Formulir Perbaikan Ujian Skripsi



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 174/SK/BAN-PT/Ak-Ppy/PT/19/2024

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

Website: [www.umsu.ac.id](http://www.umsu.ac.id)

Email: [info@umsu.ac.id](mailto:info@umsu.ac.id)

[fumsumedan](#)

[umsumedan](#)

[umsumedan](#)

[umsumedan](#)

**FORMULIR PERBAIKAN UJIAN SKRIPSI**

Pada hari ini, Senin, 02 Maret 2026 telah dilaksanakan Ujian Skripsi bagi mahasiswa Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Sbb:

Nama Mahasiswa : NADHILAH SYAFITRI  
 NPM : 2209010194  
 Program Studi : SISTEM INFORMASI  
 Judul Proposal : ANALISIS POTENSI CUSTOMER CHURN MENGGUNAKAN ALGORITMA DECISION TREE C4.5 PADA INDIBIZ TELKOM REGIONAL I

Materi/Point yang Diperbaiki :

|  |  | Paraf |
|--|--|-------|
| Rizaldy Khair, S.Kom., M.Kom<br>Pembimbing                     |  |       |
| Dr. Firahmi Rizky, S.Kom., M.kom<br>Penguji I                  | <ol style="list-style-type: none"> <li>1. Pisahkan activity diagram antara admin dan manager</li> <li>2. Tambahkan login digambar use case diagram</li> <li>3. Tambahkan sequence dan class diagram</li> <li>4. Pindahkan perhitungan manual ke bab 3</li> </ol> |       |
| Mahardika Abdi Prawira<br>Tanjung, S.Kom., M.Kom<br>Penguji II |  |       |

Berita acara ini ditandatangani setelah skripsi diperbaiki sesuai petunjuk/arahan dari Pembimbing dan Penguji/Pembahas.



## 6. Berita Acara Bimbingan Skripsi



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019  
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003  
<http://www.umhu.ac.id> [info@umhu.ac.id](mailto:info@umhu.ac.id) [fas@umhu.ac.id](mailto:fas@umhu.ac.id) [umsumedan](https://www.facebook.com/umsumedan) [umsumedan](https://www.instagram.com/umsumedan) [umsumedan](https://www.youtube.com/umsumedan) [umsumedan](https://www.linkedin.com/umsumedan)

### Berita Acara Pembimbingan Skripsi

Nama Mahasiswa : Nadhilah Syafitri  
 NPM : 2209010194  
 Program Studi : Sistem Informasi  
 Judul Penelitian : Analisis Potensi Customer Churn Menggunakan Algoritma Decision Tree (C4.5) Pada Indibiz Telkom Regional I  
 Nama Dosen Pembimbing : Rizaldy Khair, S.Kom., M.Kom

| Tanggal Bimbingan | Hasil Evaluasi   | Paraf Dosen          |
|-------------------|--|----------------------|
| 16 Desember 2025  | ● Pengerjaan bab 1-3   | <i>Rizaldy Khair</i> |
| 26 Desember 2025  | ● Diskusi terkait kendala dalam penelitian<br>● Didarakan untuk mengganti objek penelitian | <i>Rizaldy Khair</i> |
| 30 Desember 2025  | ● Pergantian judul dan penetapan judul baru  | <i>Rizaldy Khair</i> |
| 25 Januari 2026   | ● Revisi diagram, penambahan aktor/user<br>● Rapikan struktur penulisan                    | <i>Rizaldy Khair</i> |
| 03 Februari 2026  | ● Diskusi mengenai perbaikan sempro yaitu penggunaan data dummy                            | <i>Rizaldy Khair</i> |
| 04 Februari 2026  | ● Lanjut bab 4 dan 5   | <i>Rizaldy Khair</i> |
| 25 Februari 2026  | ● Penyerahan file bab 4 dan 5 serta video sistem yang telah dibuat<br>● Revisi bab 4       | <i>Rizaldy Khair</i> |
| 26 Februari 2026  | ● Revisi bab 4<br>● ACC sidang   | <i>Rizaldy Khair</i> |

Medan, 26 Februari 2026

Diketahui oleh :

Ketua Program Studi

(Mahardika Abdi Prawira  
Tanjung, S.Kom., M.Kom)

Disetujui oleh :

Dosen Pembimbing

(Rizaldy Khair, S.Kom., M.Kom)



## 7. Surat Penerimaan Naskah Publikasi Artikel Ilmiah (LOA)



Medan, 30 April 2026

No : 010/ LoA/JIRSI-Vol.5 No.2/04/2026

Hal : Surat Penerimaan Naskah Publikasi Artikel Ilmiah

**Kepada Yth:**

Bapak/ Ibu Penulis (Author)

**Nadhilah Syafitri; Rizaldy Khair**

Di

Tempat

Assalamu'alaikum Wr. Wb.

Dengan hormat

Tim Redaksi **Jurnal Ilmu Komputer dan Informatika (JIRSI) Unity Academy Research and Publication** telah menerima paper dengan judul:

**"ANALISIS POTENSI CUSTOMER CHURN MENGGUNAKAN ALGORITMA DECISION TREE (C4.5) PADA INDIBIZ TELKOM REGIONAL I"**

Berdasarkan hasil review, artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan di Jurnal "JIRSI" Volume 5, Nomor 2, Mei 2026. Kami akan mengirimkan softcopy edisi tersebut pada akhir bulan penerbitan ke email penulis. Artikel tersedia secara online di alamat <https://jurnal.unity-academy.sch.id/index.php/jirsi/index>

Demikian surat ini kami sampaikan, atas perhatian dan kerjasamanya kami ucapkan terimakasih.

Unity Academy Research and Publication  
Editor in Chief

**Muhammad Eka, ST., M.Kom**

## 8. Turnitin

### TURNITIN NADIL.docx

#### ORIGINALITY REPORT

|                                |                                |                            |                              |
|--------------------------------|--------------------------------|----------------------------|------------------------------|
| <b>28%</b><br>SIMILARITY INDEX | <b>23%</b><br>INTERNET SOURCES | <b>15%</b><br>PUBLICATIONS | <b>15%</b><br>STUDENT PAPERS |
|--------------------------------|--------------------------------|----------------------------|------------------------------|

#### PRIMARY SOURCES

|           |  |               |
|-----------|--|---------------|
| <b>1</b>  | <b>repository.umsu.ac.id</b><br>Internet Source  | <b>4%</b>     |
| <b>2</b>  | <b>www.coursehero.com</b><br>Internet Source   | <b>2%</b>     |
| <b>3</b>  | <b>Submitted to Universitas Pamulang</b><br>Student Paper  | <b>1%</b>     |
| <b>4</b>  | <b>Submitted to Universitas Putera Batam</b><br>Student Paper  | <b>1%</b>     |
| <b>5</b>  | <b>eprints.poltektegal.ac.id</b><br>Internet Source  | <b>1%</b>     |
| <b>6</b>  | <b>docplayer.info</b><br>Internet Source   | <b>&lt;1%</b> |
| <b>7</b>  | <b>repository.its.ac.id</b><br>Internet Source   | <b>&lt;1%</b> |
| <b>8</b>  | <b>Submitted to Universitas Islam Riau</b><br>Student Paper  | <b>&lt;1%</b> |
| <b>9</b>  | <b>journalcenter.org</b><br>Internet Source  | <b>&lt;1%</b> |
| <b>10</b> | <b>journal.ubm.ac.id</b><br>Internet Source  | <b>&lt;1%</b> |
| <b>11</b> | <b>Noperla Anjelisari Lumbanbatu, Yadi Limanta<br/>Maha, Sardo Pardingotan Sipayung.<br/>"Penerapan Metode Decision Tree untuk<br/>Memorediksi Dampak Kesehatan Akibat</b> | <b>&lt;1%</b> |