

**KLASIFIKASI KARAKTERISTIK CUSTOMER CHURN DI
TELEKOMUNIKASI INDUSTRI UNTUK MENINGKATKAN RETENSI
PELANGGAN MENGGUNAKAN ALGORITMA DECISION TREE ID3**

SKRIPSI

Program Studi Sistem Informasi

Jenjang Sarjana



Oleh

Stevan Desena Damanik

NIM 09031282025093

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2023

LEMBAR PENGESAHAN

SKRIPSI

**KLASIFIKASI KARAKTERISTIK CUSTOMER CHURN DI TELEKOMUNIKASI
INDUSTRI UNTUK MENINGKATKAN RETENSI PELANGGAN
MENGUNAKAN ALGORITMA DECISION TREE ID3**

Sebagai salah satu syarat untuk penyelesaian
studi di Program Studi Sistem Informasi SI

Oleh

Stevan Desena Damanik (09031282025093)

Palembang, 7 Januari 2024

**Mengetahui,
Ketua Jurusan Sistem Informasi**

Pembimbing



Endang Lestari Ruskan, M.T.
NIP 197811172006042001



Ir. Muhammed Ihsan Jambak, M.SC., M.M.
NIP 196804052013081201

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Stevan Desena Damanik
NIM : 09031282025093
Program Studi : Sistem Informasi
Judul Skripsi : Klasifikasi Karakteristik Customer Churn di
Telekomunikasi Industri Untuk
Meningkatkan Retensi Pelanggan
Menggunakan Algoritma Decision Tree ID3

Hasil pengecekan *software authenticate/Turnitin*: 13%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya sendiri dan bukan hasil plagiat. Apabila ditemukan unsur plagiat dalam laporan ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya dengan ketentuan yang berlaku.

Demikianlah pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Palembang, 7 Januari 2024



Stevan Desena Damanik

NIM. 09031282025093

HALAMAN PERSETUJUAN

Telah diterima untuk dipublikasikan pada jurnal KLIK : Kajian Ilmiah
Informatika dan Komputer (Sinta 4) pada :

Hari : Selasa

Tanggal : 4 Juli 2023

Nama : Stevan Desena Damanik

NIM : 09031282025093

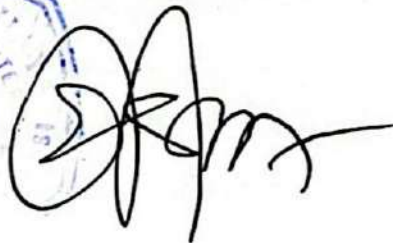
Judul Jurnal : Klasifikasi Customer Churn pada Telekomunikasi
Industri Untuk Retensi Pelanggan Menggunakan
Algoritma C4.5

Pembimbing : Ir. Muhammad Ihsan Jambak, M.SC., M.M.



Mengetahui,

Ketua Jurusan Sistem Informasi



Endang Lestari Ruskan, M.T.
NIP 197811172006042001

HALAMAN PERSEMBAHAN

“Karena perlindungan hikmat adalah seperti perlindungan uang. Dan beruntunglah yang mengetahui bahwa hikma memelihara hidup pemilik-pemiliknya” (Pengkhotbah 7:12)

Motto :

“Berkembanglah, mekarkan daun mu hingga semua melihat keindahan dan merasakan kehadiran mu. Belajarlah hingga kamu mendapati dirimu menjadi manfaat bagi orang lain”

Karya ini dipersembahkan kepada:

Diri saya sendiri, terima kasih untuk diriku yang telah berjuang sejauh ini. Akan ada banyak hal menarik yang ada di depan, teruslah mencoba hal baru dan jadilah yang terbaik.

Sondang Anita Dewi Simanjuntak, seorang Ibu yang luar biasa. Ibu yang mengajarkan dan membimbing dalam banyak hal. Ibu yang selalu sabar dalam membimbing anaknya dan selalu memberikan support yang luar biasa besar tanpa pamrih.

Amri Berman Damanik seorang Ayah yang begitu tegas. Ayah yang selalu memperhatikan anaknya dalam kecuekannya. Ayah yang selalu memberikan support kepada anaknya dan selalu membimbing hingga menjadi evan seperti sekarang.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan penyusunan skripsi yang berjudul **“Klasifikasi Karakteristik Customer Churn di Telekomunikasi Industri Untuk Meningkatkan Retensi Pelanggan Menggunakan Algoritma Decision Tree ID3”**. Skripsi ini penulisajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.

Pada prosesnya, penyusunan skripsi ini tentu tidak dapat diselesaikan dengan baik tanpa adanya bantuan, arahan, dukungan, dan petunjuk dari semua pihak yang terlibat. Untuk itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Diri sendiri yang telah berjuang dan memberikan yang terbaik.
2. Kedua orang tua, seorang saudari saya, serta seluruh keluarga yang selalu memberikan dukungan dalam bentuk apapun.
3. Ibu Endang Lestari Ruskan, M.T. selaku Ketua Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Ir. Muhammad Ihsan Jambak, M.SC., M.M. selaku Dosen Pembimbing Tugas Akhir yang sudah bersedia menyediakan waktunya untuk memberikan bimbingan, bantuan, dan dukungan dalam menyelesaikan tugas akhir ini.
5. Mba Rifka, selaku Admin Program Studi Sistem Informasi Bilingual yang telah menjadi sumber informasi dan membantu seluruh proses

pemberkasan selama mengerjakan Tugas Akhir.

6. Seluruh Dosen dan Tenaga Pengajar yang telah membantu, membimbing, dan membagi ilmunya kepada penulis selama menuntut ilmu di Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Seluruh staf Fakultas Ilmu Komputer Universitas Sriwijaya yang sudah sangat profesional dalam membantu berbagai kegiatan mahasiswa.
8. Teman – teman seperjuangan dari Sistem Informasi angkatan 2020, terkhusus teman-teman kelas SIBIL B20 yang selalu memberikan dukungan dan motivasi terhadap penulis.

Penulis terbuka dalam menerima kritik dan saran pada tugas akhir ini, karena penulis menyadari bahwa masih terdapat banyak kekurangan dan ruang untuk dilakukan perbaikan dalam tugas akhir ini. Akhir kata, semoga skripsi ini membawa kebermanfaatn bagi kita semua.

Palembang, 7 Januari 2024

Penulis,



Stevan Desena Damanik

**KLASIFIKASI KARAKTERISTIK CUSTOMER CHURN DI
TELEKOMUNIKASI INDUSTRI UNTUK MENINGKATKAN RETENSI
PELANGGAN MENGGUNAKAN ALGORITMA DECISION TREE ID3**

Oleh

Stevan Desena Damanik 09031282025093

ABSTRAK

Ketidaktahuan perusahaan telekomunikasi mengenai alasan dan ciri-ciri customer churn membuat perusahaan telekomunikasi mengalami kerugian yang besar. Hal tersebut menjadikan customer churn sebagai masalah besar bagi perusahaan telekomunikasi. Pada Penelitian ini akan menggunakan Rapid Miner dan algoritma ID3 untuk melakukan proses data mining. Tujuan dari penelitian ini yaitu mengetahui apa saja karakteristik customer churn sehingga perusahaan dapat membuat kebijakan yang dapat mempertahankan customer serta meningkatkan retensi pelanggan. Penelitian ini didasarkan pada CRISP-DM. Data yang digunakan diambil dari kaggle.com dengan 21 atribut dan 7034 baris data dan akan dilakukan data preparation. Dari hasil penelitian diketahui terdapat 5 atribut yang memiliki pengaruh cukup besar terhadap customer churn yaitu contract, InternetService, TotalChares, tenure, PaperlessBilling MultipleLines, StreamingMovies. Dan dari hasil penelitian ini memiliki tingkat akurasi sebesar 79.53%

Kata Kunci: Data Mining; Karakteristik Customer Churn; Algoritma ID3; Perusahaan Telekomunikasi; Rapid Miner

**CLASSIFICATION OF CUSTOMER CHURN CHARACTERISTICS IN
INDUSTRIAL TELECOMMUNICATIONS TO INCREASE CUSTOMER
RETENTION USING THE DECISION TREE ID3 ALGORITGHM**

By

Stevan Desena Damanik 09031282025093

ABSTRACT

Ignorance of telecommunications companies regarding the reasons and characteristics of customer churn causes telecommunications companies to suffer huge losses. This makes customer churn a big problem for telecommunications companies. This study uses data mining with classification techniques as a solution to analyze customer churn characteristics. This research will use Rapid Miner and the ID3 algorithm to carry out the data mining process. . The purpose of this research is to find out what are the characteristics of customer churn so that companies can make policies that can retain customers and increase customer retention. This research is based on CRISP-DM. Data taken from kaggle.com with 21 attributes and 7034 rows of data and data preparation will be carried out. From the research results it is known that there are 5 attributes that have a considerable influence on customer churn, namely contracts, InternetService, TotalChares, tenure, PaperlessBilling, MultipleLines, StreamingMovies. And from the results of this study has an accuracy rate of 79.53%.

Keywords: *Data Mining; Characteristics of Customer Churn; ID3 Algorithm;
Telecommunication Companies; Rapid Miner*

DAFTAR ISI

LEMBAR PENGESAHAN	ii
HALAMAN PERNYATAAN.....	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
ABSTRAK	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan.....	3
1.4 Manfaat.....	4
1.5 Batasan Masalah.....	4
BAB II KAJIAN PUSTAKA	5
2.1 Metode Klasifikasi	5
2.2 Algoritma ID3	6
2.3 Training dan Validation Set	7
2.4 Confusion Matrix	7
2.5 Penelitian Terdahulu.....	9
BAB III METODE PENELITIAN	12
3.1 Objek Penelitian	12

3.2	Kerangka Penelitian	14
3.3	Analisa Masalah	14
3.4	Data Preparation	15
3.5	Membagi Data Training dan Data Testing	17
3.6	Penerapan Algoritma ID3.....	17
BAB IV HASIL DAN PEMBAHASAN		20
4.1	Confusion Matrix	20
4.2	Pohon Keputusan.....	22
BAB V KESIMPULAN DAN SARAN		26
5.1	Kesimpulan.....	26
5.2	Saran.....	27
DAFTAR PUSTAKA		28
L A M P I R A N.....		30

DAFTAR GAMBAR

Gambar 2.1 Proses Klasifikasi	6
Gambar 3.1 Kerangka Kerja Penelitian.....	14
Gambar 3.2 Proses Modeling	18
Gambar 3.3 Parameter Decision Tree.....	19
Gambar 4.1 Prediksi Churn	20
Gambar 4.2 Pohon Keputusan	23

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	9
Tabel 3.1 Informasi Dataset	12
Tabel 3.2 Dataset Pelanggan Telekomunikasi	13
Tabel 3.3 Dataset Setelah Preparation.....	16
Tabel 4.1 Confusion Matrix	21

DAFTAR LAMPIRAN

Lampiran A. Surat Kesiediaan Membimbing.....	A-1
Lampiran B. Form Pengajuan Topik Skripsi	B-1
Lampiran C. Surat Keputusan Tugas Akhir (SK TA).....	C-1
Lampiran D. Bukti Proses Submit Jurnal.....	D-1
Lampiran E. Letter Of Acceptance	E-1
Lampiran F. Hasil Pengecekan Similiarity	F-1
Lampiran G. Surat Keterangan Pengecekan Similiarity	G-1
Lampiran H. Form Perubahan Judul Skripsi.....	H-1
Lampiran I. Kartu Konsultasi.....	I-1
Lampiran J. Log Book	J-1

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Kepentingan telekomunikasi telah mendorong munculnya berbagai perusahaan, yang pada gilirannya menciptakan persaingan yang sengit dalam industri ini (Irfiani, 2014). Untuk meraih keunggulan dalam persaingan ini, perusahaan-perusahaan melakukan inovasi guna mempertahankan pelanggan mereka dan menarik para pelanggan baru (Irmada et al., 2019). Persaingan perebutan pelanggan merupakan masalah yang tidak dapat dihindari oleh perusahaan telekomunikasi (Alwis et al., 2018). Persaingan ini membawa dampak kerugian bagi beberapa pihak yang kehilangan pelanggan (*churn*), sementara pihak lainnya memperoleh keuntungan dari perolehan pelanggan baru (Syarif & Nugraha, 2023).

Masalah *customer churn* di industri ini sangat perlu diperhatikan karena biaya kerugian yang ditimbulkan sangat besar untuk perusahaan di industri telekomunikasi (Govindaraju et al., n.d.). Ketidakpuasan pelanggan, kekecewaan terhadap pelayanan serta layanan lain yang lebih menawarkan keuntungan yang lebih merupakan alasan pelanggan berhenti berlayanan. Untuk mengatasinya, perusahaan-perusahaan telekomunikasi perlu untuk menemukan dan memahami alasan dan karakteristik dari pelanggan yang akan berhenti berlangganan, sehingga perusahaan dapat meningkatkan kualitas pelayanan kepada pelanggan dan dapat meningkatkan kepuasan pelanggan (Wicaksono, 2021).

Permasalahan yang dihadapi oleh perusahaan telekomunikasi yaitu apa saja karakteristik dari pelanggan yang loyal atau churn (Latief et al., 2021). Permasalahan ini tentunya harus segera diatasi untuk meningkatkan loyalitas dan retensi pelanggan (WARDANI et al., 2018). Ketidakpahaman perusahaan mengenai karakteristik dari pelanggan merupakan penyebab customer churn. Penerapan analisis merupakan cara untuk menemukan karakteristik pelanggan berhenti berlayanan. Penerapan analisis ini menjadi krusial untuk meningkatkan retensi pelanggan pada perusahaan di industri telekomunikasi (Pratama et al., 2023). Ada berbagai metode untuk menemukan karakteristik pelanggan yang akan berhenti berlangganan, salah satunya yaitu data mining

Pemanfaatan data mining akan menghasilkan pengetahuan dengan nilai yang signifikan bagi perusahaan (Mardi, n.d.). Dalam konteks ini, data mining sering diartikan sebagai sebuah langkah yang dilakukan untuk menemukan pola tertentu dan menarik pada sebuah dataset (Sudarsono et al., 2021). pada data mining terdapat banyak algoritma yang digunakan untuk permodelan. Algoritma seperti rules-based learning, neural networks, dan decision tree telah terbukti mampu memprediksi customer churn dengan tingkat keakuratan yang memadai (Wicaksono, 2021). Pada Penelitian yang telah dilakukan menggunakan algoritma ID3 untuk mencari karakteristik customer churn diketahui bahwa algoritma ID3 sesuai untuk tujuan yang ingin dicapai yaitu mencari karakteristik dari customer churn. (Tri Utami et al., 2020).

Kerugian besar yang dapat dialami perusahaan telekomunikasi akibat *customer churn* memotivasi penulis untuk melakukan penelitian dengan tujuan menemukan faktor atau karakteristik dari customer yang akan berhenti

memakai layanan. Penelitian ini dapat menjadi acuan untuk tindakan preventif terhadap customer churn untuk meningkatkan loyalitas dan mempertahankan pelanggan. Untuk menemukan faktor tersebut Penulis memilih untuk menganalisis masalah dengan data mining dan menerapkan Algoritma ID3 guna mengeksplorasi karakteristik pelanggan yang berpotensi berhenti berlangganan,

1.2 Rumusan Masalah

Penulis merumuskan masalah penelitian dari penjelasan pada latar belakang yaitu “Apa saja karakteristik dari customer churn yang ada pada perusahaan telekomunikasi?”. Dari rumusan masalah tersebut maka penulis menguraikan rumusan masalah menjadi beberapa *research question* untuk membantu penelitian, yaitu :

1. Mengapa menggunakan teknik pembagian data otomatis?
2. Mengapa memilih algoritma ID3 dalam proses modeling?
3. Mengapa menggunakan confusion matrix sebagai teknik analisis?

1.3 Tujuan

1. Mengetahui mengapa menggunakan teknik pembagian data otomatis
2. Mengetahui mengapa menggunakan algoritma ID3
3. Mengetahui mengapa menggunakan confusion matrix sebagai teknik analisis

1.4 Manfaat

1. Dapat memahami keunggulan dalam menggunakan *split* data otomatis
2. Dapat memahami cara kerja serta kelebihan dan kelemahan algoritma ID3
3. Dapat memahami penggunaan confusion matrix sebagai teknik analisis

1.5 Batasan Masalah

Untuk memastikan penelitian yang dilakukan ini tetap sesuai dengan rumusan masalah yang ada, maka penulis membuat batasan masalah sebagai berikut :

1. Data yang dipakai penelitian ini diperoleh dari media online yaitu dari [kaggle.com](https://www.kaggle.com)
2. Penelitian ini hanya membahas tentang klasifikasi customer churn untuk mengetahui karakteristik customer churn.
3. Penelitian ini menggunakan hanya confusion matrix untuk membahas evaluasi dari modeling

DAFTAR PUSTAKA

- Alwis, P. K. D. N. M., Kumara, B. T. G. S., & Hapuarachchi, H. A. C. S. (2018). *Customer Churn Analysis and Prediction in Telecommunication for Decision Making*.
- Anita, Wicaksono, A., & Padilah, T. N. (2021). Pengaruh Jumlah Record Dataset Terhadap Algoritma Klasifikasi Berdasarkan Data Customer Churn. *Jurnal Ilmiah Informatika*, 6(1), 1–10. <https://doi.org/10.35316/jimi.v6i1.1223>
- Govindaraju, R., Simatupang, T., Samadhi, T. A., Keahlian, K., Industri, S., Tekno-Ekonomi, &, Industri, T., Bandung, I. T., & Manufaktur, S. (n.d.). *PERANCANGAN SISTEM PREDIKSI CHURN PELANGGAN PT. TELEKOMUNIKASI SELULER DENGAN MEMANFAATKAN PROSES DATA MINING*.
- Hozairi, Anwari, & Alim, S. (2021). IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES ORANGE DATA MINING IMPLEMENTATION FOR STUDENT GRADUATION CLASSIFICATION USING K-NEAREST NEIGHBOR, DECISION TREE AND NAIVE BAYES MODELS. In *Jurnal Ilmiah NERO* (Vol. 6, Issue 2).
- Irfiani, E. (2014). PENENTUAN LOYALITAS PELANGGAN PADA DISTRIBUTOR PULSA ELEKTRONIK MENGGUNAKAN ALGORITMA C4.5 DAN NAIVE BAYES. *JURNAL SISTEM INFORMASI STMIK ANTAR BANGSA*.
- Irmanda, H. N., Astriratma, R., & Afrizal, S. (2019). PERBANDINGAN METODE JARINGAN SYARAF TIRUAN DAN POHON KEPUTUSAN UNTUK PREDIKSI CHURN. *JSI : Jurnal Sistem Informasi (E-Journal)*, 11(2). <http://ejournal.unsri.ac.id/index.php/jsi/index>
- Kristiawan, K., & Widjaja, A. (2021). Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1). <https://doi.org/10.28932/jutisi.v7i1.3182>
- Latief, I., Subekti, A., & Gata, W. (2021). PREDIKSI TINGKAT PELANGGAN

CHURN PADA PERUSAHAAN TELEKOMUNIKASI DENGAN
ALGORITMA ADABOOST. *Jurnal Informatika*, 21.

- Mardi, Y. (n.d.). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- Marlina, D., & Bakri, M. (2021). PENERAPAN DATA MINING UNTUK MEMREDIKSI TRANSAKSI NASABAH DENGAN ALGORITMA C4.5. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2.
- Miryam Clementine, & Arum. (2022). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan ID3. *Jurnal Processor*, 17(1), 9–18. <https://doi.org/10.33998/processor.2022.17.1.1170>
- Pratama, R., Herdiansyah, M. I., Syamsuar, D., & Syazili, A. (2023). Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12(1), 96–104. <https://doi.org/10.32736/sisfokom.v12i1.1507>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER. *JBASE - Journal of Business and Audit Information Systems*, 4(1). <https://doi.org/10.30813/jbase.v4i1.2729>
- Syarif, M., & Nugraha, W. (2023). MWMOTE DALAM MENGATASI KETIDAKSEIMBANGAN KELAS PADA PREDIKSI CHURN MENGGUNAKAN KLASIFIKASI C4.5. *Jurnal Mahasiswa Teknik Informatika*, 7.
- Tri Utami, Y., Asiah Shofiana, D., Heningtyas, Y., Ilmu Komputer FMIPA Universitas Lampung Jl Soemantri Brodjonegoro No, J., & Lampung, B. (2020). Penerapan Algoritma C4.5 Untuk Prediksi Churn Rate Pengguna Jasa Telekomunikasi. In *Jurnal Komputasi* (Vol. 8, Issue 2).
- WARDANI, N., DANTES, G., & INDRAWAN, G. (2018). PREDIKSI CUSTOMER CHURN DENGAN ALGORITMA DECISION TREE C4.5 BERDASARKAN SEGEMENTASI PELANGGAN PADA PERUSAHAAN RETAIL. *RESISTOR*.
- Wicaksono, A. (2021). Uji Performa Teknik Klasifikasi untuk Memprediksi Customer Churn. *Bianglala Informatika*, 9(1).