
**PREDIKSI *CUSTOMER CHURN* PADA LAYANAN INDIHOME
MENGUNAKAN ALGORITMA *DECISION TREE*
(STUDI KASUS PT. TELKOM AKSES)**

Andree Rizky Yuliansyah Siregar¹, Muhammad Iqbal²

Universitas Pembangunan Panca Budi, Medan, Indonesia

email: ¹siregarandree07@gmail.com, ²muhammadiqbal@dosen.pancabudi.ac.id

Abstract: This study aims to predict Customer Churn for IndiHome services at PT. Telkom Akses using the Decision Tree algorithm. The research analyzes several variables that influence customers' decisions to discontinue subscriptions, including complaints, service packages, and subscription duration. The analysis results show that the Complaint attribute has the highest Gain value (0.6898), making it the main factor distinguishing customers likely to churn from those who remain. This suggests a strong correlation between complaint frequency and the customers' decision to continue or terminate the service. Additionally, the Package attribute also has a significant impact, with a Gain of 0.2785, indicating that the selected package speed influences the likelihood of churn. Conversely, the Subscription Duration attribute has the lowest Gain (0.0411), indicating that this variable provides minimal information in predicting churn. Based on these findings, PT. Telkom Akses is recommended to consider improving complaint management and optimizing service package offerings to reduce Customer Churn rates.

Keyword: Customer Churn; Decision Tree; Prediction

Abstrak: Penelitian ini bertujuan untuk memprediksi Customer Churn pada layanan IndiHome di PT. Telkom Akses menggunakan algoritma Decision Tree. Penelitian ini dilakukan dengan menganalisis sejumlah variabel yang mempengaruhi keputusan pelanggan untuk berhenti berlangganan, antara lain: keluhan, paket layanan, dan lama berlangganan. Hasil analisis menunjukkan bahwa atribut Keluhan memiliki nilai Gain tertinggi (0.6898), menjadikannya faktor utama dalam membedakan pelanggan yang cenderung berhenti berlangganan (churn) dan yang tetap (tidak churn). Hal ini menunjukkan adanya korelasi yang kuat antara frekuensi keluhan dengan keputusan pelanggan untuk terus menggunakan layanan. Selain itu, atribut Paket juga memiliki pengaruh signifikan dengan Gain 0.2785, yang mengindikasikan bahwa kecepatan paket layanan memengaruhi kemungkinan pelanggan untuk churn. Sebaliknya, atribut Lama Berlangganan memiliki Gain terendah (0.0411), yang menunjukkan bahwa variabel ini kurang informatif dalam memprediksi churn. Dengan hasil ini, PT. Telkom Akses disarankan untuk mempertimbangkan pengelolaan keluhan pelanggan serta pemilihan paket layanan yang sesuai guna mengurangi tingkat Customer Churn.

Kata kunci: Customer Churn; Decision Tree; Prediksi

PENDAHULUAN

Dalam dunia usaha, perusahaan harus berupaya untuk mampu memahami kebutuhan dan keinginan pelanggan dengan tujuan agar terjalin hubungan baik antara kedua belah pihak. Apabila konsumen terpenuhi kebutuhan dan keinginannya, maka konsumen akan merasa puas dan akan membangun loyalitas pelanggan pada perusahaan sehingga terjalin ikatan yang erat dan saling menguntungkan antara kedua belah pihak (Yudiana et al., 2023). Pada dasarnya kepuasan pelanggan secara keseluruhan tidak akan tercapai sekalipun dalam kurun waktu sementara, karena kepuasan pada manusia bersifat kompleks dan tidak dapat diukur dan berbeda setiap individu. Upaya perbaikan dan pembaharuan selalu dilakukan dengan berbagai cara dan strategi. Mempertahankan pelanggan sangat berkaitan dengan kepuasan pelanggan (Purwadinata & Ridolof, 2020).

Mempertahankan pelanggan yang ada seringkali lebih menguntungkan dibanding menarik pelanggan baru dan juga membantu perusahaan menjual lebih banyak produk (Putriani et al., 2024). Masalah yang dihadapi oleh perusahaan adalah bagaimana mencegah fenomena *Customer Churn* ini, yang terjadi ketika pelanggan berhenti berlangganan layanan atau produk perusahaan (Agustina Indriyani, 2019). Fenomena ini menjadi masalah kritis karena dapat menyebabkan penurunan pendapatan dan mengganggu stabilitas bisnis perusahaan. Untuk mengatasi masalah ini, diperlukan metode prediksi yang efektif. Prediksi *Customer Churn* bertujuan untuk mengklasifikasikan data pelanggan sebelumnya menjadi dua kategori: pelanggan yang akan berhenti berlangganan dan pelanggan yang akan

terus berlangganan (Makmur & Sos, 2024). Prediksi tersebut memanfaatkan ilmu data mining peran klasifikasi yang merupakan menempatkan variabel atau objek ke dalam beberapa kategori relevan yang telah ditetapkan sebelumnya (Setiawan et al., 2023). Dalam proses eksekusi data mining, diperlukan sebuah algoritma yang dapat mengklasifikasikan apakah *Customer Churn* atau tidak *Churn*. Untuk itu dalam proses prediksi *Churn* algoritma yang digunakan adalah *Decision Tree*. Algoritma *Decision Tree* adalah salah satu algoritma yang sangat populer pada peran data mining klasifikasi strukturnya berbentuk pohon dimana setiap node merepresentasikan atribut yang diuji, dan setiap cabang dari node *Decision Tree* merepresentasikan hasil dari uji atribut tersebut (Kurniawan, 2022). Analisis ini memungkinkan perusahaan untuk mengadaptasi strategi pemasaran, layanan pelanggan, dan pengembangan produk sesuai dengan kebutuhan pelanggan, sehingga meningkatkan retensi, mengurangi *Churn*, dan loyalitas pelanggan (Sudirjo et al., 2023).

Dalam penelitian ini, penulis menggunakan model *Decision Tree*. Penelitian ini diharapkan dapat memberikan beberapa manfaat, antara lain (Etriyanti, 2021):

1. Membantu Mengidentifikasi Faktor-Faktor Penyebab *Churn*

Dengan menggunakan algoritma *Decision Tree*, penelitian ini dapat membantu mengidentifikasi faktor-faktor utama yang menyebabkan pelanggan *Churn* pada layanan Indihome. Informasi ini bermanfaat bagi PT. Telkom Akses untuk memahami alasan di balik keputusan pelanggan untuk berhenti berlangganan.

2. Meningkatkan Strategi Retensi Pelanggan

Hasil dari penelitian ini diharapkan mampu memberikan panduan bagi PT. Telkom Akses dalam merancang strategi retensi yang lebih efektif. Dengan mengetahui variabel-variabel yang berpengaruh terhadap *Churn*, perusahaan dapat mengambil tindakan preventif untuk mempertahankan pelanggan.

3. Optimalisasi Penggunaan Sumber Daya

Dengan prediksi *Churn* yang lebih akurat, PT. Telkom Akses dapat lebih efektif dalam mengalokasikan sumber daya dan anggaran pada pelanggan yang memiliki risiko *Churn* tinggi. Hal ini dapat membantu perusahaan dalam mengurangi biaya retensi yang tidak efisien.

4. Pengembangan Layanan Berbasis Data

Penelitian ini juga dapat mendukung pengembangan layanan yang lebih sesuai dengan kebutuhan dan preferensi pelanggan melalui pendekatan berbasis data. Dengan pemahaman yang lebih baik mengenai pola dan karakteristik pelanggan yang cenderung *Churn*, PT. Telkom Akses dapat melakukan peningkatan layanan secara lebih tepat.

5. Memberikan Kontribusi Terhadap Penelitian di Bidang *Customer Churn Prediction*

Selain manfaat praktis, penelitian ini juga diharapkan dapat memberikan kontribusi terhadap studi di bidang *Customer Churn Prediction*, khususnya dalam konteks industri telekomunikasi di Indonesia, sehingga dapat menjadi referensi bagi penelitian serupa di masa mendatang.

METODE

Desain Penelitian

Penelitian ini menggunakan desain metode kuantitatif dengan pendekatan analisis prediktif. Desain ini cocok untuk penelitian yang mengolah data numerik atau kategori dengan tujuan untuk mengidentifikasi pola dan memprediksi perilaku pelanggan, khususnya kecenderungan untuk *churn* (berhenti berlangganan). Dengan menggunakan Decision Tree, penelitian ini bertujuan untuk memahami faktor-faktor yang mempengaruhi *churn* dan membuat model prediksi berdasarkan data historis pelanggan (Pugu et al., 2024).

Sumber dan Teknik Pengumpulan Data

Data diperoleh dari PT. Telkom Akses yang berisi informasi pelanggan Indihome. Data ini terdiri dari berbagai variabel seperti ID Pelanggan, Lama Berlangganan, Paket Layanan, Keluhan, dan Status *Churn*. Data yang digunakan adalah data sekunder, yang berfokus pada data historis pelanggan dengan beberapa variabel penentu. Data diambil langsung dari sistem atau basis data PT. Telkom Akses atau melalui divisi terkait yang mengelola data pelanggan (Jauhar, 2024).

Dataset

Dataset tersebut berisi data pelanggan di PT Telkom Akses yang memiliki 7 variabel, yaitu:

1. Pelanggan ID: ID unik yang diberikan kepada setiap pelanggan untuk mengidentifikasi mereka dalam sistem.
2. Nama Pelanggan: Nama dari pelanggan yang berlangganan layanan Indihome.
3. Lama Berlangganan: Durasi atau waktu pelanggan telah berlangganan

- layanan, yang diukur dalam satuan bulan atau tahun.
4. Paket: Jenis paket layanan yang dipilih oleh pelanggan, seperti paket internet dasar, paket internet + TV, atau paket lainnya yang tersedia.
 5. Harga: Biaya atau tarif bulanan yang dibayarkan pelanggan untuk paket layanan yang mereka pilih.
 6. Keluhan: Jumlah atau frekuensi keluhan yang diajukan oleh pelanggan terkait layanan yang diterima.
 7. Status Churn: Status apakah pelanggan masih aktif atau telah berhenti berlangganan layanan (churn/non-churn), yang menjadi variabel target dalam penelitian ini.

Variabel-variabel ini diharapkan dapat memberikan informasi yang cukup untuk memprediksi kemungkinan pelanggan akan churn, dengan memanfaatkan algoritma *Decision Tree* untuk menemukan pola-pola yang signifikan dari karakteristik dan perilaku pelanggan tersebut (Putra et al., 2024).

Metode Pengolahan Data

Metode pengolahan data terdiri dari beberapa langkah utama:

Pra-pemrosesan Data: Pada tahap ini, data dibersihkan untuk memastikan tidak ada nilai yang hilang (*missing values*) atau data yang tidak konsisten. Variabel kategori seperti paket layanan dapat di-*encoding* menjadi nilai numerik untuk mempermudah analisis.

Pembagian Data: Data dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian ini dilakukan untuk menguji akurasi model yang dikembangkan (Irfansyah et al., 2021).
Pengolahan dengan *Decision Tree*: Algoritma *Decision Tree* diaplikasikan pada data latih untuk membangun model prediktif. Selanjutnya, data uji digunakan untuk menilai performa model (Depari et al., 2022).

HASIL DAN PEMBAHASAN

Analisis Data

Data primer dari perusahaan dapat dilihat pada Tabel 1 sebagai berikut:

Tabel 1. Data Primer

Nama Pelanggan	Lama Berlangganan (Bulan)	Paket	Harga (Rp)	Keluhan	Status
Rizky Mellean	12	Internet 50Mbps	290000	2	Churn
Irfan Lutfan	24	Internet 30Mbps	265000	3	Churn
Ogi Reynaldi	36	Internet 100Mbps	350000	1	Tidak Churn
Oki Siregar	12	Internet 50Mbps	290000	2	Churn
...
Bernad Ardiansyah	36	Internet 30Mbps	265000	4	Churn

Menentukan Entropy Tahap 1

Setelah menentukan data awal, langkah selanjutnya yaitu melakukan perhitungan dengan mencari nilai entropy dari masing-masing data, menggunakan rumus entropy, dapat dilihat pada persamaan 1.

$$Entropy (S) = \sum_{i=1}^k -p_i * \log_2 p_i$$

Selanjutnya data pada Tabel 1 akan di lakukan perhitungan Entropy terdapat pada Tabel 2, yang dapat dilihat pada Tabel.

Tabel 2. Perhitungan Node 1

No de	Jenis	Jumlah Pengguna (S)	Churn (S ₁)	Tidak Churn (S ₂)	Entropy
1	30	30	17	13	0.9871

Entropy (Total)

$$= \left(- \left(\frac{13}{30} \right) * \log_2 \left(\frac{13}{30} \right) \right) + \left(- \left(\frac{17}{30} \right) * \log_2 \left(\frac{17}{30} \right) \right) = 0.9871$$

Selanjutnya data pada Tabel 1 dan Tabel 2 akan di lakukan perhitungan Entropy dan Gain.

Entropy (12 Bulan)

$$= \left(- \left(\frac{3}{9} \right) * \log_2 \left(\frac{3}{9} \right) \right) + \left(- \left(\frac{6}{9} \right) * \log_2 \left(\frac{6}{9} \right) \right) = 0.9183$$

Entropy (24 Bulan)

$$= \left(- \left(\frac{6}{10} \right) * \log_2 \left(\frac{6}{10} \right) \right) + \left(- \left(\frac{4}{10} \right) * \log_2 \left(\frac{4}{10} \right) \right) = 0.971$$

Entropy (36 Bulan)

$$= \left(- \left(\frac{4}{11} \right) * \log_2 \left(\frac{4}{11} \right) \right) + \left(- \left(\frac{7}{11} \right) * \log_2 \left(\frac{7}{11} \right) \right) = 0.946$$

Entropy (30Mbps)

$$= \left(- \left(\frac{5}{13} \right) * \log_2 \left(\frac{5}{13} \right) \right) + \left(- \left(\frac{8}{13} \right) * \log_2 \left(\frac{8}{13} \right) \right) = 0.961$$

Entropy (50Mbps)

$$= \left(- \left(\frac{6}{11} \right) * \log_2 \left(\frac{6}{11} \right) \right) + \left(- \left(\frac{5}{11} \right) * \log_2 \left(\frac{5}{11} \right) \right) = 0.994$$

Entropy (100Mbps)

$$= \left(- \left(\frac{2}{6} \right) * \log_2 \left(\frac{2}{6} \right) \right) + \left(- \left(\frac{4}{6} \right) * \log_2 \left(\frac{4}{6} \right) \right) = 0.918$$

$$Entropy (1) = \left(- \left(\frac{8}{8} \right) * \log_2 \left(\frac{8}{8} \right) \right) + \left(- \left(\frac{0}{8} \right) * \log_2 \left(\frac{0}{8} \right) \right) = 0$$

$$Entropy (2) = \left(- \left(\frac{5}{9} \right) * \log_2 \left(\frac{5}{9} \right) \right) + \left(- \left(\frac{2}{9} \right) * \log_2 \left(\frac{2}{9} \right) \right) = 0.947$$

$$Entropy (3) = \left(- \left(\frac{0}{7} \right) * \log_2 \left(\frac{0}{7} \right) \right) + \left(- \left(\frac{7}{7} \right) * \log_2 \left(\frac{7}{7} \right) \right) = 0$$

$$Entropy (4) = \left(- \left(\frac{0}{8} \right) * \log_2 \left(\frac{0}{8} \right) \right) + \left(- \left(\frac{8}{8} \right) * \log_2 \left(\frac{8}{8} \right) \right) = 0$$

Menentukan Gain Tahap 1

Setelah melakukan perhitungan entropy, langkah selanjutnya yaitu melakukan perhitungan nilai gain dari masing-masing atribut. Adapun rumus yang digunakan dapat dilihat dalam Persamaan 2.

$$Gain (A) = \sum_{i=1}^k \frac{|S_i|}{|S|} * Entropi (S_i)$$

Tabel 3. Perhitungan Entropy dan Gain

Kategori	Jenis	Jumlah Pengguna (S)	Churn (S ₁)	Tidak Churn (S ₂)	Entropy	Gain
Lama Berlangganan	12 Bulan	9	6	3	0.9183	0.0411
	24 Bulan	10	4	6	0.971	
	36 Bulan	11	7	4	0.946	
Paket	30Mbps	13	8	5	0.961	0.2785
	50Mbps	11	5	6	0.994	
	100Mbps	6	4	2	0.918	
Keluhan	1	8	0	8	0	0.6898
	2	7	2	5	0.947	
	3	7	7	0	0	
	4	8	8	0	0	

Gain (Lama Berlangganan)

$$= 0.9871 - \left(\frac{9}{30} * 0.9183\right) + \left(\frac{10}{30} * 0.971\right) + \left(\frac{11}{30} * 0.946\right) = 0.0411$$

Gain (Paket) = 0.9871 - $\left(\frac{13}{30} * 0.961\right)$

$$+ \left(\frac{11}{30} * 0.994\right) + \left(\frac{16}{30} * 0.918\right) = 0.2785$$

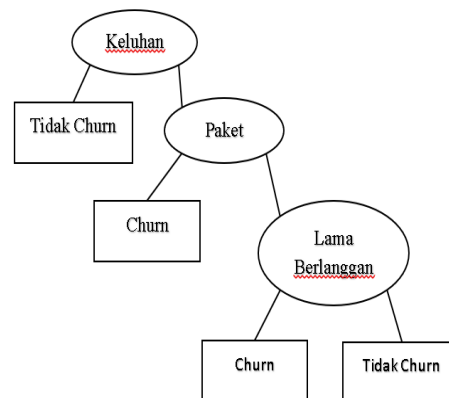
Gain (Keluhan) = 0.9871 - $\left(\frac{8}{30} * 0\right)$

$$+ \left(\frac{9}{30} * 0.947\right) + \left(\frac{7}{30} * 0\right) + \left(\frac{8}{30} * 0\right) = 0.6898$$

Setelah ditentukan data awal, langkah selanjutnya yaitu melakukan perhitungan dengan mencari nilai entropy dan gain dalam sebuah table agar lebih rinci dan jelas.

Keluhan memiliki Gain tertinggi (0.6898), yang berarti atribut Keluhan paling efektif dalam membedakan pengguna yang Churn (S₁) dan Tidak Churn (S₂). Paket memiliki Gain yang cukup besar (0.2785) dan merupakan atribut yang efektif setelah Keluhan. Lama Berlangganan memiliki Gain yang paling kecil (0.0411), sehingga ini adalah atribut yang kurang efektif untuk

digunakan di awal pohon keputusan. Hasil pohon keputusan sebagai berikut:



Gambar Hasil Pohon Keputusan

SIMPULAN

Berdasarkan hasil penelitian yang dilakukan dengan menggunakan algoritma *Decision Tree* untuk memprediksi *Customer Churn* pada layanan IndiHome di PT. Telkom Akses, dapat disimpulkan bahwa atribut Keluhan memiliki Gain tertinggi (0.6898), menjadikannya sebagai faktor yang paling signifikan dalam membedakan pelanggan yang Churn (S₁) dan yang Tidak Churn (S₂). Hal ini menunjukkan bahwa frekuensi atau jenis keluhan yang diajukan oleh pelanggan memiliki

korelasi yang sangat kuat dengan keputusan mereka untuk bertahan atau berhenti berlangganan layanan. Atribut Paket juga memberikan kontribusi yang cukup signifikan dengan Gain 0.2785, yang menunjukkan bahwa kecepatan paket yang dipilih oleh pelanggan (seperti 30Mbps, 50Mbps, atau 100Mbps) berpengaruh terhadap kemungkinan mereka untuk churn. Oleh karena itu, penawaran paket yang lebih sesuai dengan kebutuhan pelanggan dapat menjadi salah satu strategi untuk mengurangi churn. Sebaliknya, atribut Lama Berlangganan memiliki Gain yang paling kecil (0.0411), yang mengindikasikan bahwa lama berlangganan pelanggan tidak memberikan banyak informasi untuk memprediksi churn, setidaknya pada tahap awal pembentukan pohon keputusan. Ini berarti bahwa meskipun durasi pelanggan menggunakan layanan berpotensi relevan, faktor lain (seperti keluhan dan paket) lebih berpengaruh dalam menentukan perilaku churn.

DAFTAR PUSTAKA

- Agustina Indriyani, H. (2019). *Algoritma Decision Tree (ID3) dan K-Nearest Neighbour (K-NN) dalam Prediksi Costumer Churn Berdasarkan Segmentasi Pelanggan (Studi Kasus Radity Craft Tasikmalaya)*. Universitas Siliwangi.
- Depari, D. H., Widiastiwi, Y., & Santoni, M. M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Informatik: Jurnal Ilmu Komputer*, 18(3), 239. <https://doi.org/10.52958/iftk.v18i3.4694>
- Etriyanti, E. (2021). Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, 3(1), 6–14.
- Irfansyah, D., Mustikasari, M., & Suroso, A. (2021). Arsitektur Convolutional Neural Network (CNN) Alexnet Untuk Klasifikasi Hama Pada Citra Daun Tanaman Kopi. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(2), 87–92.
- Jauhar, L. M. (2024). *Penerapan Persuasion Melalui Costumer Purchase Decision dalam Program Pemasaran Linnect*. Universitas Islam Indonesia.
- Kurniawan, D. (2022). *Pengenalan machine learning dengan python*. Elex Media Komputindo.
- Makmur, M., & Sos, S. (2024). BAB 4 PEMILIHAN LOKASI RITEL. *Manajemen Bisnis Ritel*, 39.
- Pugu, M. R., Riyanto, S., & Haryadi, R. N. (2024). *Metodologi Penelitian; Konsep, Strategi, dan Aplikasi*. PT. Sonpedia Publishing Indonesia.
- Purwadinata, S., & Ridolof, W. B. (2020). *Pengantar Ilmu Ekonomi: Kajian Teoritis dan Praktis Mengatasi Masalah Pokok Perekonomian*.
- Putra, R. F., Mukhlis, I. R., Datya, A. I., Pipin, S. J., Reba, F., Al-Husaini, M., Mandowen, S. A., Zain, N. N. L. E., & Judijanto, L. (2024). *Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi*. PT. Sonpedia Publishing Indonesia.
- Putriani, D., Prayogi, A. P. A., Shofyana, A. I., Ristyawan, A., & Daniati, E. (2024). Prediksi Customer Churn Menggunakan Algoritma Decision Tree. *Prosiding SEMNAS INOTEK*

(Seminar Nasional Inovasi Teknologi), 8(1), 85–94.

Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., Widiputra, H., Meilani, B. D., Ibrahim, R. N., & Azdy, R. A. (2023). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.

Sudirjo, F., Purwati, T., Widyastuti, W., Budiman, Y. U., & Manuhutu, M. (2023). Analisis Dampak Strategi Pemasaran Digital dalam Meningkatkan Loyalitas Pelanggan: Perspektif Industri E-commerce.

Jurnal Pendidikan Tambusai, 7(2), 7524–7532.

Yudiana, Y., Agustina, A. Y., & Khofifah, N. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1–20.