

KLASIFIKASI CUSTOMER CHURN PADA INDUSTRI BANK UNTUK RETENSI PELANGGAN MENGGUNAKAN DECISION TREE

Shafa Edfinanda Maghfira¹, Lyonora Octavia², Rafhi Arizkia Nuriza³, Rahayu Rustianawati⁴, Dhea Eka Putri⁵, Aidah Tazkiyah Fasya⁶, Rendy Firman Maulana⁷, Andreas Adi Trinoto⁸

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer
Universitas Indraprasta PGRI

Jalan Raya Tengah No 80, Kelurahan Gedong, Pasar Rebo, Jakarta Timur
shafaedfinanda06@gmail.com¹, lyonoranora638@gmail.com², rafhiarizkia16@gmail.com³,
rrahayuayy@gmail.com⁴, dheacka.putri06@gmail.com⁵, aidahtazfas@gmail.com⁶,
rendyfirman788@gmail.com⁷, a.trinoto@gmail.com⁸

Abstrak

Dalam industri perbankan, mempertahankan pelanggan merupakan tantangan utama yang berdampak langsung pada profitabilitas dan kelangsungan bisnis. Kehilangan pelanggan, atau yang dikenal sebagai customer churn, dapat menyebabkan kerugian besar jika tidak ditangani secara efektif. Artikel ini berfokus pada pengembangan model klasifikasi customer churn dengan memanfaatkan algoritma Decision Tree sebagai metode utama. Algoritma ini dipilih karena keunggulannya dalam menangani data berukuran besar, kemudahan interpretasi hasil, dan tingkat akurasi yang andal. Penelitian ini menggunakan dataset yang digunakan diperoleh dari Kaggle dengan 10.000 data dan 18 atribut. Model yang dikembangkan menghasilkan akurasi sebesar 99% melalui evaluasi *confusion matrix*. Penelitian ini memberikan wawasan penting bagi bank untuk merancang strategi retensi pelanggan yang lebih efektif.

Kata Kunci : *Customer Churn, Decision Tree, Industri Perbankan, CRISP-DM*

Abstract

In the banking industry, customer retention is a major challenge that has a direct impact on profitability and business continuity. Losing customers, also known as customer churn, can cause huge losses if not handled effectively. This article focuses on the development of a customer churn classification model by utilizing the Decision Tree algorithm as the main method. This algorithm was chosen because of its advantages in handling large data, ease of interpretation of results, and reliable accuracy levels. This study uses a dataset obtained from Kaggle with 10,000 data and 18 attributes. The developed model produces an accuracy of 99% through confusion matrix evaluation. This study provides important insights for banks to design more effective customer retention strategies.

Keywords: *Customer Churn, Decision Tree, Banking Industry, CRISP-DM*

PENDAHULUAN

Setiap perusahaan memiliki strategi unik dalam memberikan layanan berkualitas dengan harga yang terjangkau. Langkah ini bertujuan untuk menarik lebih banyak pelanggan, sehingga pendapatan perusahaan dapat meningkat secara maksimal (Herawati et al., 2016). Perusahaan harus berusaha memahami kebutuhan dan keinginan pelanggan untuk membangun hubungan yang harmonis. Jika kebutuhan dan keinginan pelanggan terpenuhi, mereka akan merasa puas, sehingga loyalitas terhadap perusahaan dapat terwujud. Dengan demikian, tercipta hubungan yang kuat dan saling menguntungkan antara pelanggan dan perusahaan (Yudiana et al., 2023). Perkembangan teknologi telah membuat persaingan dalam industri perbankan semakin sengit (Anam et al., 2024). Guliyef dan Yerdelen berpendapat bahwa Bank churn, atau fenomena di mana pelanggan beralih ke bank lain, merupakan salah satu isi krusial dalam industri perbankan. Perpindahan pelanggan ini dapat berdampak pada pendapatan dan reputasi bank, sebab biaya untuk menarik pelanggan baru biasanya cukup tinggi. Dalam sektor ini, analisis terhadap perilaku pelanggan serta prediksi kemungkinan perpindahan mereka berdasarkan pola tersebut menjadi topik penelitian yang sangat relevan. Hasil dari analisis tersebut berpengaruh besar terhadap kebijakan bank, karena dapat

membantu dalam merancang strategi baru untuk mempertahankan pelanggan atau menyempurnakan strategi yang ada (Azmi, 2024).

Memahami perilaku pelanggan merupakan langkah penting untuk membantu penerapan strategi yang tepat. Salah satu caranya dengan mengelompokkan pelanggan berdasarkan kategori variabel yang relevan, guna meningkatkan pendapatan perusahaan. Dalam hal ini, penggunaan teknik data mining menjadi salah satu solusi efektif untuk menyelesaikan masalah segmentasi konsumen (Adiana et al., 2018). Kemajuan teknologi di sektor perbankan kini mempermudah pelanggan untuk membuka rekening baru dan memindahkan aset mereka tanpa harus datang langsung ke kantor bank, cukup dilakukan dari rumah. Hal ini membuat bank semakin memperhatikan pentingnya loyalitas pelanggan. Namun, sebelum mengembangkan strategi yang efektif untuk mempertahankan pelanggan, bank perlu mengidentifikasi terlebih dahulu pelanggan yang berpotensi berhenti menggunakan layanan mereka (Husein & Harahap, 2021).

CRISP-DM (Cross Industry Standard Process for Data Mining) adalah suatu standar proses yang dirancang khusus untuk pengolahan data mining. Dalam pendekatan ini, data yang tersedia akan melewati setiap fase yang telah terstruktur dengan baik serta terdefinisi secara jelas dan efisien, sehingga memungkinkan proses analisis data yang sistematis dan dapat diandalkan (Hasanah et al., 2021). Data mining, yang merupakan salah satu alat analisis dalam analytical CRM, telah terbukti bermanfaat bagi perusahaan dalam mengidentifikasi pelanggan yang kemungkinan akan berhenti, memperkirakan waktu terjadinya churn, serta mengevaluasi tingkat akurasi prediksi tersebut (Suhartono, 2010). *Decision tree* adalah metode yang efektif dalam mengklasifikasikan *customer churn* di industri perbankan, yang berperan penting dalam strategi retensi pelanggan. Algoritma ini mampu mengidentifikasi pola perilaku nasabah yang cenderung berhenti menggunakan layanan bank, sehingga memungkinkan bank untuk mengambil tindakan preventif yang tepat. Konsep *Decision Tree* melibatkan penyajian pernyataan bersyarat di setiap langkah yang bercabang, yang membantu dalam pengambilan keputusan berdasarkan analisis data set itu sendiri (Syah et al., 2024). Ada beberapa macam algoritma Decision Tree yaitu: CART, C4.5, ID3 dan ada banyak algoritma lainnya. Maka dari itu pada penelitian ini akan menggunakan algoritma CART (*Classification And Regression Tree*).

Penelitian ini bertujuan untuk mengidentifikasi karakteristik customer churn, sehingga perusahaan di industri Bank dapat merumuskan kebijakan yang tepat dalam menangani masalah tersebut. Dengan keputusan yang efektif untuk mengatasi customer churn, diharapkan perusahaan dapat meningkatkan retensi pelanggan sekaligus mengurangi kerugian yang ditimbulkan akibat berhentinya pelanggan.

PENELITIAN RELEVAN

Penelitian oleh (Damanik & Jambak, 2023) yang berjudul *Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi Pelanggan Menggunakan Algoritma c.45* dari hasil penelitian ini memiliki tingkat akurasi sebesar 79.53%.

Penelitian oleh (Hasanah et al., 2021) yang berjudul *Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir* dari hasil pengujian algoritma ini memiliki kinerja yang cukup baik dengan akurasi sebesar 89,4%.

Penelitian relevan pertama berkaitan erat dengan klasifikasi yang dipakai pada subjek penelitian, yaitu customer churn, sedang penelitian relevan kedua berkaitan dengan model dan metode yang digunakan pada penelitian ini. Hasil akurasi keduanya memiliki tingkat akurasi yang tinggi yang membuat penelitian ini akan semakin bernilai bagi pelanggan bank.

METODE PENELITIAN

Metode Penelitian menggunakan metode CRISP DM sebagai alur penelitian, adapun langkah-langkahnya sebagai berikut :

1. Analisa Masalah

Bank perlu mengetahui faktor apa saja yang berkontribusi terhadap churn, seperti usia nasabah, jumlah produk bank yang digunakan, kepuasan layanan pelanggan, ataupun keluhan pelanggan.

2. Pengumpulan Data

Dalam alur CRISP-DM ada yang namanya *Data understanding*, disini berisi tentang cara mengambil data, memahami data, dan pemahaman atribut data.

Dalam penelitian di perlukan dataset yang dapat di gunakan untuk di analisa, dalam hal ini peneliti mengambil dataset dari situs kaggle.com <https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn/data> dataset ini terdiri dari 18 atribut yang terlihat pada tabel 1, dataset ini memiliki 10000 data.

Tabel 1. Informasi dataset

No	Atribut	Keterangan
1	RowNumber	Sesuai dengan nomor rekaman (baris) dan tidak berpengaruh pada output.
2	CustomerId	Berisi nilai acak dan tidak berpengaruh pada keputusan nasabah untuk meninggalkan bank.
3	Surname	Nama belakang nasabah
4	CreditScore	Dapat berdampak pada churn nasabah, karena nasabah dengan skor kredit yang lebih tinggi cenderung tidak meninggalkan bank.
5	Geography	Lokasi nasabah.
6	Gender	Jenis kelamin nasabah
7	Age	Usia Nasabah
8	Tenure	Masa nasabah menjadi pelanggan bank
9	Balance	Saldo nasabah
10	NumOfProducts	Jumlah produk bank yang di beli nasabah
11	HasCrCard	Kolom menunjukkan nasabah memiliki kredit atau tidak
12	IsActiveMember	Nasabah aktif atau tidak
13	EstimatedSalary	Estimasi gaji
14	Exited	Customer berhenti langganan atau tidak
15	Complain	Nasabah memiliki keluhan atau tidak
16	Satisfaction Score	Skore kepuasan nasabah atas tanggapan keluhan
17	Card Type	Tipe Kartu nasabah
18	Point Earned	Point yang di peroleh nasabah menggunakan kartu kredit

Tabel 2. Dataset Pelanggan Bank

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	...	Exited
1	15634602	Hargrave	619	France	Female	42	...	1
2	15647311	Hill	608	Spain	Female	41	...	0
3	15619304	Onio	502	France	Female	42	...	1
...	...	...	...	...	...	...	...	...

10000	15628319	Walker	792	France	Female	28	...	0
-------	----------	--------	-----	--------	--------	----	-----	---

3. Pemrosesan data

Langkah selanjutnya yaitu *Data processing* atau pemrosesan data, disini kita akan mencari atribut yang kosong, hilang, duplikat, ataupun tidak wajar, dan mengisi, mengganti dan menghapusnya bila ada atribut yang tidak wajar itu di temukan

Dari hasil pemrosesan data di dapat bahwa data yang dimiliki lengkap dan tidak ada yang tidak wajar, akan tetapi menghapus kolom yang tidak berkaitan dengan klasifikasi churn yaitu *RowNumber*, *CustomerId*, dan *Surname*. Sehingga kolom yang terpakai menjadi 15 kolom yang semula berjumlah 18 kolom setelah itu bisa ke tahap selanjutnya.

Tabel 3. Tabel Pemrosesan data

CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	...	Exited
619	France	Female	42	2	0.00	1	...	1
608	Spain	Female	41	1	83807.86	1	...	0
502	France	Female	42	8	159660.80	3	...	1
...	...	...	...	...	...	...	...	...
792	France	Female	28	4	130142.79	1	...	0

4. Transformasi data

Pada tahap tranformasi akan mengubah kolom” yang bertipe kategori ke tipe *numeric*, ada 3 kolom yang diubah yaitu kolom *geography*, *Gender*, dan *Card Type*.

Tabel 4. Tabel Transformasi data

CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	...	Exited
619	0	0	42	2	0.00	1	...	1
608	2	0	41	1	83807.86	1	...	0
502	0	0	42	8	159660.80	3	...	1
...	...	...	...	...	...	...	...	...
792	0	0	28	4	130142.79	1	...	0

5. Modelling

Pada tahap *Modelling*, pemodelan menggunakan teknik klasifikasi untuk mendapatkan suatu pola yang berisi informasi yang dapat digunakan, teknik pemodelan yang digunakan adalah model klasifikasi *decision tree* dengan algoritma CART.

Langkah-langkah dalam algoritma CART (Hasanah et al., 2021)

- Menentukan akar root
- Menghitung gain information masing-masing calon cabang
- Memilih atribut dengan gain information yang mempunyai nilai terbesar sebagai splitting atribut yang dipilih sebagai calon cabang berikutnya.
- Ulangi langkah 2 dan 3 hingga terdapat leaf node

Gain information mengukur nilai impurity dari suatu partisi A, dengan perhitungan

$$Gini(D) = 1 - \sum_{i=1}^m P_i^2$$

Dimana :

Gini (D) = nilai impurity dari partisi,

M = jumlah indeks

P = sebuah frekuensi dari satu record dataset tuple D pada indeks ke i

Average Gini Impurity dapat dihitung dengan

$$GiniA(D) = \frac{|D1|}{|D|} Gini(D1) + \frac{|D2|}{|D|} Gini(D2)$$

Dimana :

D = tuple D

D1 = partisi pertama tuple D

D2 = partisi kedua tuple D

GiniA(D) = impurity dari partisi D pada atribut A

Gini(D1) = impurity dari partisi pertama tuple D

Gini(D2) = impurity dari partisi kedua tuple D

Penurunan tingkat impurity yang diperoleh terhadap atribut A, bisa dihitung dengan :

$$\Delta Gini(A) = Gini(D) - Gini_A(D)$$

Dimana :

$\Delta Gini(A)$ = tingkat impurity

6. *Evaluation.*

Pada tahap *Evaluation* dilakukan dengan menghitung metrik seperti akurasi, presisi, dan recall untuk memastikan kinerja model. Evaluasi menggunakan confusion matrix.

$$\text{Akurasi} = \frac{\text{True Positif} + \text{True Negatif}}{\text{True Positif} + \text{True Negatif} + \text{False Positif} + \text{False Negatif}}$$

$$\text{Presisi} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Positif}}$$

$$\text{Recal} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Negatif}}$$

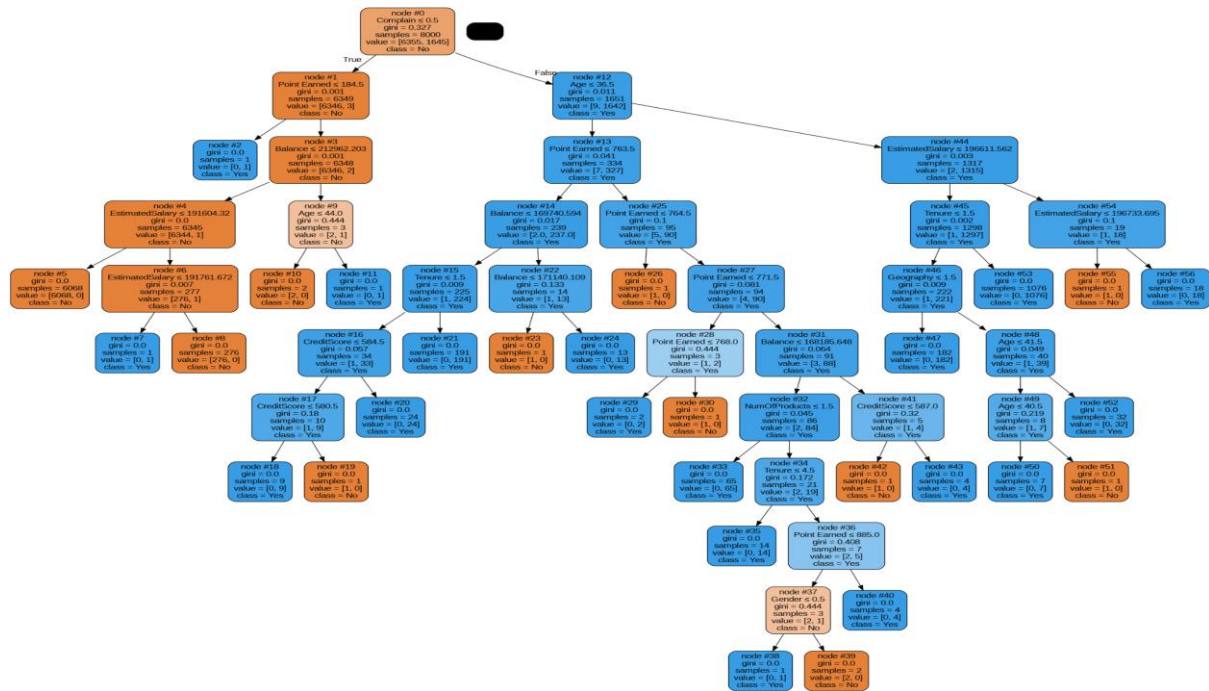
7. *Deployment*

Pada tahap *Deployment* merupakan hasil dari penelitian berupa jurnal, artikel dan dashboard.

HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan ini masuk ke dalam bagian Modeling dan Evaluation di mana disini akan menjabarkan hasil modelling menggunakan Decision Tree untuk mendapatkan hasil klasifikasi dan mendapatkan informasi terhadap kinerja algoritma.

Algoritma yang digunakan adalah algoritma *CART (Classification and Regresion)*, *CART* adalah algoritma yang populer karena mudah digunakan, algoritma ini bekerja dengan memberikan aturan-aturan pada suatu keputusan sehingga pengambilan keputusan menjadi lebih mudah dan efisien, susunan proses pohon keputusan terdiri dari *root node*, *internal node*, dan *leaf node*. berikut adalah hasil pengolahan algoritma *CART* :



Gambar 1. Pohon Keputusan Industri Bank

Berdasarkan hasil visualisasi pohon keputusan di atas dapat di pahami bahwa ada 9 atribut yang memiliki pengaruh terhadap berhentinya pelanggan bank, yaitu *Complain*, *PointEarned*, *Balance*, *Age*, *Tenure*, *KreditScore*, *NumOfProduct* dan *Geography*. Algoritma CART ini menghasilkan 39 *node* dengan tingkat kedalaman 17 level. Atribut *Complain* menjadi *root node* dengan kriteria *splitting* $Complain \leq 0.5$, dengan kata lain atribut *Complain* menjadi pemecah kondisi sebelum *splitting*. Berikutnya terdapat beberapa informasi di *root node* ini, seperti perolehan parameter gini impurity sebesar 0,327 dengan sampel (jumlah data) = 8000 untuk 6355 kelas 0 (no) dan 1645 kelas 1 (yes). *Root node* ini akan di *splitting* hingga menghasilkan internal node, dan terus di *splitting* hingga menghasilkan leaf node(keputusan) dengan gini impurity 0.

Dari pola klasifikasi pada algoritma CART di atas di temukan bahwa atribut *Complain* adalah hal yang perlu di perhatikan dari seorang customer. dimana 99% customer yang melakukan *complain*, mereka berhenti langganan, selain itu perlu juga di perhatikan *NumOfProduct*, *Age*, dan *tenure*. dimana customer yang hanya memiliki 1 produk perbankan, usia diatas 36.5 tahun, dan *Tenure* diatas 1 perlu perhatian khusus untuk retensi pelanggan.

Evaluation

Setelah pola klasifikasi di dapatkan pada algoritma CART selanjutnya tahap evaluasi menggunakan confusion matrix untuk mencari akurasi, presisi dan recal dari model yang sudah di buat. Berikut hasil evaluasi dari model decision tree.

```
Accuracy: 1.00
```

Confusion Matrix:

```
[[1604  3]
 [ 3 390]]
```

Classification Report:

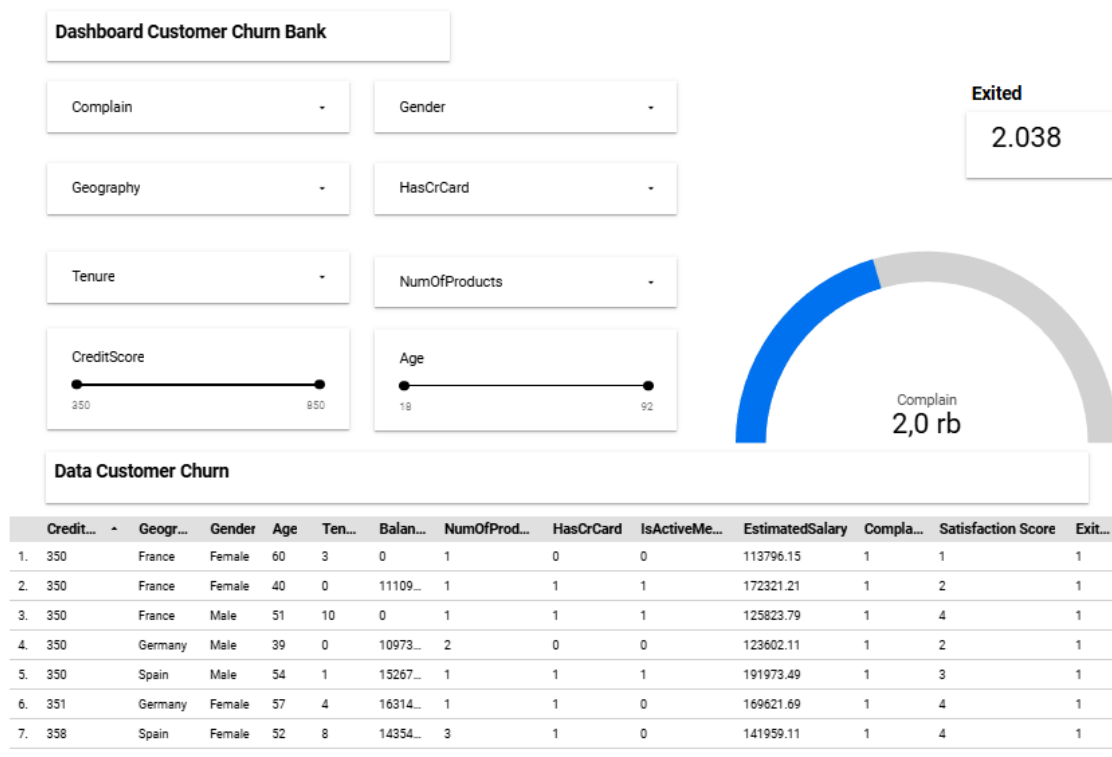
	precision	recall	f1-score	support
0	1.00	1.00	1.00	1607
1	0.99	0.99	0.99	393
accuracy			1.00	2000
macro avg	1.00	1.00	1.00	2000
weighted avg	1.00	1.00	1.00	2000

Gambar 2. Hasil Evaluasi pada decision tree

Berdasarkan gambar, dapat diketahui bahwa nilai akurasi adalah 100 % dengan jumlah prediksi benar 1604 data pada tidak berhenti langganan, 3 data memprediksi salah pada tidak berhenti langganan, 390 data memprediksi benar pada berhenti langganan, dan 3 data memprediksi salah pada berhenti langganan, sehingga precision, dan recall 99%.

Deployment

Setelah tahap evaluasi dimana menilai secara detail sebuah model maka dilakukan sebuah pengimplementasian dari keseluruhan model yang telah dibangun, dalam kasus ini adalah membuat dashboar visualisasi menggunakan Looker Studio, Looker studio adalah alat gratis yang dapat mengubah data menjadi dashboard dan laporan yang informatif dan mudah dibaca. Berikut adalah hasil dari dashboar berikut.



Gambar 3. Dashboard LookerStudio

SIMPULAN

Penelitian ini membahas tentang klasifikasi *customer churn* pada industri perbankan menggunakan algoritma *Decision Tree*. Tujuannya adalah untuk membantu bank dalam memahami karakteristik pelanggan yang berpotensi berhenti (*churn*) sehingga dapat merumuskan kebijakan retensi pelanggan yang lebih efektif. Penelitian dilakukan dengan menggunakan metode CRISP-DM sebagai kerangka kerja.

Dataset yang digunakan terdiri dari 10.000 data dengan 18 atribut, yang diolah dan dimodelkan menggunakan algoritma CART (*Classification and Regression Tree*). Model yang dihasilkan menunjukkan tingkat akurasi tinggi sebesar 99%. Beberapa karakteristik utama yang memengaruhi *churn* meliputi keluhan pelanggan (*Complain*), saldo, umur, lama menjadi pelanggan (*Tenure*), dan jumlah produk yang dimiliki pelanggan (*NumOfProducts*). Oleh karena itu, perusahaan diharapkan memberikan perhatian khusus kepada pelanggan yang mengajukan keluhan (*Complain*) untuk mengurangi tingkat *churn*.

Mengingat akurasi tinggi yang diperoleh dari algoritma *Decision Tree*, peneliti menyarankan penelitian selanjutnya untuk menggunakan algoritma lain guna melakukan perbandingan dan

mendapatkan hasil yang lebih optimal. Dengan demikian, pelanggan yang berpotensi berhenti dapat diidentifikasi dengan lebih baik.

DAFTAR PUSTAKA

- Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering. *Jurnal Terapan Teknologi Informasi*, 2(1), 23–32. <https://doi.org/10.21460/jutei.2018.21.76>
- Anam, K., Rinaldi, A. R., & Fathurrohman, F. (2024). Komparasi Algoritma Machine Learning Dalam Klasifikasi Loyalitas Nasabah Bank Berbasis Particle Swarm Optimization. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(4), 8212–8218. <https://doi.org/10.36040/jati.v8i4.10941>
- Azmi, A. F. (2024). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix. *Komputa: Jurnal Ilmiah Komputer dan Informatika*, 13(1), 111–119. <https://ojs.unikom.ac.id/index.php/komputa/article/view/12639>
- Damanik, S. D., & Jambak, M. I. (2023). Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi Pelanggan Menggunakan Algoritma C4.5. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 3(6), 1303–1309. <https://doi.org/10.30865/klik.v3i6.829>
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108. <https://doi.org/10.30871/jaic.v5i2.3200>
- Herawati, M., Wibowo, I. L., & Mukhlash, I. (2016). Prediksi Customer Churn Menggunakan Algoritma Fuzzy Iterative Dichotomiser 3. *Limits: Journal of Mathematics and Its Applications*, 13(1), 23. <https://doi.org/10.12962/j1829605x.v13i1.1913>
- Husein, A. M., & Harahap, M. (2021). Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning. *Data Sciences Indonesia (DSI)*, 1(1), 8–13. <https://doi.org/10.47709/dsi.v1i1.1169>
- Suhartono, E. (2010). *PREDIKSI PERILAKU LOYALITAS PELANGGAN DENGAN TEHNIK DATA MINING DECISION TREE (Studi Kasus pada Pelanggan Kartu Telkomsel Halo di Semarang)*. 6(1), 146–177.
- Syah, K., Palluvi, R., Syaada, N., & Intan, B. (2024). *Komparasi Metode Decision Tree dan K-Nearest Neighbor (KNN) dalam Memprediksi Costumer Churn Pada Perusahaan Telekomunikasi*. 3(1), 39–45.
- Yudiana, Y., Yulia Agustina, A., & Nur Khofifah, dan. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 01–20. <http://e-journal.lp2m.uinjambi.ac.id/ojp/index.php/ijoieb>