

IMPLEMENTASI ALGORITMA DECISION TREE UNTUK PREDIKSI CUSTOMER CHURN PADA LAYANAN TELEKOMUNIKASI MENGGUNAKAN ALTAIR AI STUDIO (RAPIDMINER)

Hawa Awalya Ghuftron ^{a,1}, Kevin Handerson ^{a,2}, Yoseph Way^{a,3}, Dikdik Permana Wigandi ^{a,4}

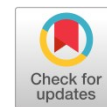
^a Universitas Bina Sarana Informatika, Jakarta, 10450, Indonesia

¹ hawaawalya@gmail.com; ² kevinhanderson3@gmail.com; ³ yosephyoseph508@gmail.com; ⁴ dikdik.dkn@bsi.ac.id

Diterima 26 Juni 2026; Direvisi 27 Oktober 2026; Disetujui 28 Oktober 2026

ABSTRAK

Customer churn merupakan permasalahan utama dalam industri telekomunikasi karena berdampak langsung pada penurunan pendapatan perusahaan, sehingga diperlukan model prediksi yang akurat sekaligus mudah ditafsirkan. Penelitian ini bertujuan menerapkan algoritma Decision Tree untuk memprediksi customer churn pada layanan telekomunikasi menggunakan Altair AI Studio (RapidMiner). Data yang digunakan adalah dataset Telco Customer Churn dari Kaggle yang terdiri atas 7.043 data pelanggan dan 21 atribut. Pengolahan data dilakukan melalui tahapan Retrieve, Set Role, Split Data dengan proporsi 70% data latih dan 30% data uji, Decision Tree dengan kriteria gain ratio, Apply Model, dan Performance. Hasil pengujian menunjukkan bahwa model mencapai akurasi sebesar 75,77%, dengan class precision 86,62% dan class recall 79,25% untuk kelas pelanggan bertahan, serta class recall 66,13% untuk kelas churn. Analisis pohon keputusan menunjukkan bahwa masa langganan (tenure), total biaya (TotalCharges), dan biaya bulanan (MonthlyCharges) merupakan atribut yang paling berpengaruh. Pelanggan dengan masa langganan panjang cenderung bertahan, sedangkan pelanggan dengan biaya rendah yang tidak menggunakan layanan hiburan tambahan lebih rawan melakukan churn. Penelitian ini berkontribusi dalam menyediakan aturan keputusan yang mudah dipahami sebagai dasar penyusunan strategi retensi pelanggan pada perusahaan telekomunikasi.

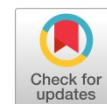


KATA KUNCI

Customer Churn,
Data Mining,
Decision Tree,
RapidMiner

ABSTRACT

Customer churn is a major problem in the telecommunications industry because it directly reduces company revenue, requiring an accurate and interpretable prediction model. This study aims to apply the Decision Tree algorithm to predict customer churn in telecommunication services using Altair AI Studio (RapidMiner). The data used is the Telco Customer Churn dataset from Kaggle, consisting of 7,043 customer records and 21 attributes. Data were processed through the Retrieve, Set Role, Split Data (70% training and 30% testing), Decision Tree (gain ratio criterion), Apply Model, and Performance stages. The test results show that the model achieves an accuracy of 75.77%, with a class precision of 86.62% and a class recall of 79.25% for the retained customer class, and a class recall of 66.13% for the churn class. The decision tree analysis shows that subscription length (tenure), total charges (TotalCharges), and monthly charges (MonthlyCharges) are the most influential attributes. Customers with long subscription periods tend to stay, while low-charge customers who do not use additional entertainment services are more likely to churn. This study contributes interpretable decision rules that can serve as a basis for developing customer retention strategies in telecommunications companies.



KEYWORD

Customer Churn,
Data Mining,
Decision Tree,
RapidMiner



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

1. Pendahuluan

Dalam era kompetisi industri telekomunikasi yang semakin ketat, mempertahankan pelanggan lama menjadi strategi yang jauh lebih efisien dibandingkan mencari pelanggan baru. Biaya untuk mengakuisisi pelanggan baru dinilai dapat mencapai lima kali lipat lebih besar daripada biaya menjaga loyalitas pelanggan yang sudah ada [9]. Fenomena berpindahnya pelanggan ke kompetitor lain, atau yang dikenal sebagai *customer churn*, berdampak langsung pada penurunan pendapatan dan stabilitas finansial perusahaan secara signifikan. Faktor yang mendorong pelanggan bertahan atau berpindah pun dapat berbeda antarpasar, misalnya kualitas layanan, kepuasan pelanggan, paket berlangganan, dan cakupan jaringan [11]. Oleh karena itu, kemampuan untuk memprediksi pelanggan yang rawan melakukan *churn* secara dini sangat diperlukan agar perusahaan dapat melakukan tindakan pencegahan yang tepat sasaran.

Data mining dengan pendekatan klasifikasi menawarkan solusi komputasional untuk menggali pola tersembunyi dari data historis pelanggan yang berjumlah besar [14]. Perkembangan *machine learning* dalam beberapa tahun terakhir juga telah mendorong pemanfaatan berbagai algoritma klasifikasi pada permasalahan bisnis nyata, termasuk analisis perilaku pelanggan [13]. Salah satu algoritma klasifikasi yang memiliki keunggulan dalam hal interpretabilitas adalah *Decision Tree*. Algoritma ini mengonversi data yang kompleks menjadi pohon keputusan berupa aturan logis *IF-THEN* yang mudah dipahami oleh pihak manajemen bisnis [2], [8].

Sejumlah penelitian terdahulu telah menerapkan berbagai algoritma untuk memprediksi *churn* pada industri telekomunikasi, mulai dari model berbasis pohon keputusan hingga *deep learning* [7], [12]. Penelitian pada dataset *Telco Customer Churn* menunjukkan bahwa algoritma berbasis pohon keputusan menghasilkan akurasi pada kisaran 74% hingga 80% [1], [3], sedangkan algoritma *ensemble* seperti *Random Forest* menghasilkan akurasi 77,15% [6]. Meskipun model yang lebih kompleks dapat memberikan akurasi sedikit lebih tinggi, model tersebut umumnya lebih sulit ditafsirkan oleh pengambil keputusan. Hal ini menjadikan *Decision Tree* tetap relevan sebagai model yang menyeimbangkan kinerja prediksi dan kemudahan interpretasi.

Penelitian ini menerapkan algoritma *Decision Tree* menggunakan perangkat lunak Altair AI Studio (RapidMiner) pada dataset pelanggan telekomunikasi. Tujuan penelitian ini adalah mengukur performa akurasi model dalam memprediksi perilaku pelanggan, serta mengekstrak karakteristik utama yang membuat pelanggan memilih untuk tetap loyal atau justru berpotensi melakukan *churn*.

2. Tinjauan Pustaka

2.1 Data Mining dan Klasifikasi

Data mining merupakan proses mengekstraksi atau menggali pengetahuan tersembunyi, pola, atau informasi yang bermanfaat dari sekumpulan data yang berukuran besar [4]. Salah satu tugas utama dalam *data mining* adalah klasifikasi, yaitu proses menempatkan objek data ke dalam salah satu dari beberapa kategori atau kelas yang telah ditentukan sebelumnya berdasarkan atribut yang dimiliki oleh data tersebut [14]. Dalam konteks *machine learning*, klasifikasi termasuk dalam kategori *supervised learning* karena model dilatih menggunakan data yang telah memiliki label kelas [13].

2.2 Algoritma Decision Tree

Algoritma *Decision Tree* (pohon keputusan) adalah salah satu metode klasifikasi yang paling populer karena strukturnya yang menyerupai pohon terbalik, mudah diimplementasikan, dan hasilnya mudah diinterpretasikan oleh manusia [15]. Struktur pohon keputusan tersusun atas tiga komponen

utama. Simpul akar (*root node*) merupakan atribut paling utama yang menjadi pemisah awal dari keseluruhan data. Simpul cabang (*internal node*) merupakan atribut yang memeriksa kondisi data berdasarkan kriteria tertentu, sedangkan simpul daun (*leaf node*) merupakan hasil keputusan akhir berupa kelas target, yang dalam penelitian ini menyatakan apakah pelanggan diprediksi melakukan *churn* atau tidak [2].

Algoritma ini bekerja dengan mengukur tingkat kemurnian hasil pemisahan data menggunakan kriteria tertentu, seperti *Information Gain*, *Gain Ratio*, atau *Gini Index*, untuk menentukan atribut yang paling layak dijadikan simpul pemisah [14]. Proses pemisahan dilakukan secara rekursif hingga seluruh data pada suatu simpul berada dalam kelas yang sama atau kriteria penghentian terpenuhi [2].

2.3 Customer Churn

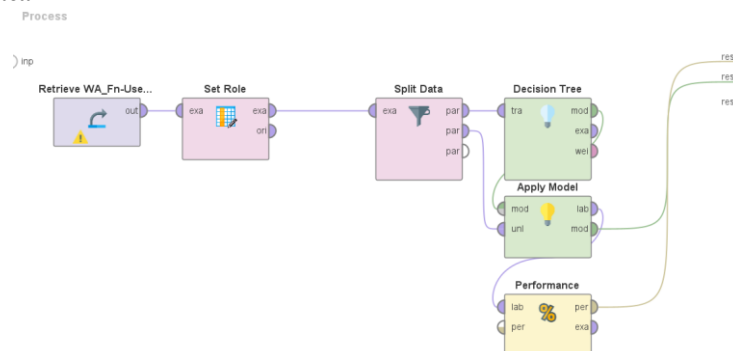
Customer churn adalah kondisi ketika pelanggan memutuskan untuk menghentikan hubungan kontrak atau langganan layanan dengan sebuah perusahaan [9]. Di industri telekomunikasi yang sangat kompetitif, deteksi *churn* secara dini sangat krusial agar pihak manajemen dapat segera memberikan intervensi berupa promo khusus atau perbaikan layanan sebelum pelanggan benar-benar berpindah ke kompetitor [11]. Prediksi *churn* juga merupakan permasalahan klasifikasi yang menantang karena data pelanggan umumnya memiliki banyak atribut dengan tipe yang beragam serta jumlah pelanggan *churn* yang lebih sedikit dibandingkan pelanggan yang bertahan [12].

2.4 Penelitian Terkait

Damanik dan Jambak [3] menerapkan algoritma C4.5 menggunakan RapidMiner pada data *customer churn* telekomunikasi dari Kaggle dan memperoleh akurasi sebesar 79,53%. Antoh dkk. [1] menggunakan dataset *Telco Customer Churn* dan menunjukkan bahwa akurasi C4.5 dasar sebesar 74,09% dapat ditingkatkan menjadi 80,05% melalui *tuning hyperparameter* dengan metode *Random Search*. Pada dataset yang sama, Junaidi dan Kaban [6] menerapkan *Random Forest* dan memperoleh akurasi 77,15% dengan nilai *precision* 55,65% dan *recall* 68,45% untuk kelas *churn*.

Putriani dkk. [10] menerapkan *Decision Tree* pada dataset *customer churn* berukuran 64.000 data dan memperoleh akurasi yang sangat tinggi, yaitu 99,7%, yang menunjukkan bahwa kinerja model sangat dipengaruhi oleh karakteristik dataset yang digunakan. Sementara itu, Khattak dkk. [7] membandingkan model *deep learning* gabungan BiLSTM-CNN dengan sejumlah algoritma klasik dan memperoleh akurasi 81% untuk model usulan, sedangkan *Decision Tree* memperoleh akurasi 78%. Temuan-temuan tersebut menunjukkan bahwa *Decision Tree* memiliki kinerja yang kompetitif, sekaligus memberikan keunggulan berupa aturan keputusan yang dapat ditelusuri secara langsung.

3. Pengolahan Data



Gambar 1. Alur Process Workflow Klasifikasi Decision Tree pada RapidMiner.

Metodologi yang digunakan dalam penelitian ini terdiri atas beberapa tahapan pengolahan data yang terstruktur. Proses pemodelan dilakukan menggunakan perangkat lunak Altair AI Studio (RapidMiner), dengan alur pengolahan data (*workflow*) sebagaimana ditunjukkan pada Gambar 1.

Tahap pertama adalah pengambilan data menggunakan operator *Retrieve WA_Fn-UseC*, yang berfungsi untuk memanggil dan memasukkan dataset historis *Telco Customer Churn* ke dalam lembar kerja. Dataset ini berisi informasi karakteristik dari 7.043 data pelanggan yang siap diolah lebih lanjut. Data tersebut kemudian dialirkan menuju operator *Set Role* yang digunakan untuk menentukan fungsi atau peran setiap atribut data [5]. Pada tahap ini, atribut *Churn* ditetapkan sebagai *label*, yaitu variabel target yang akan diprediksi, sedangkan atribut lainnya berperan sebagai *regular attribute* atau variabel fitur.

Data yang telah memiliki label selanjutnya dipisahkan menggunakan operator *Split Data*. Pemisahan ini bertujuan agar model diuji menggunakan data yang tidak dilibatkan dalam proses pelatihan, sehingga kinerja model dapat diukur secara objektif dan gejala *overfitting* dapat diketahui [14]. Proporsi pembagian yang digunakan adalah 70% sebagai data latih (*training*), yang dialirkan dari *port* par pertama menuju *port* tra pada operator *Decision Tree* untuk melatih algoritma mengenali pola data, dan 30% sebagai data uji (*testing*), yang dialirkan dari *port* par kedua menuju *port* unl pada operator *Apply Model* untuk menguji performa model terhadap data baru.

Data latih kemudian diolah oleh operator *Decision Tree* untuk membentuk aturan-aturan keputusan berbasis pohon dengan parameter standar (*criterion: gain_ratio*). Model yang dihasilkan diteruskan melalui *port* mod ke operator *Apply Model*. Operator ini menerima dua masukan, yaitu objek model dari *Decision Tree* dan data uji sebesar 30% yang belum diketahui labelnya melalui *port* unl, lalu menerapkan aturan yang telah dipelajari model untuk menghasilkan prediksi *churn* pada data uji tersebut.

Tahap akhir adalah evaluasi performa, yaitu dengan mengalirkan data hasil prediksi melalui *port* lab ke operator *Performance (Classification)*. Operator ini membandingkan hasil prediksi model dengan label aktual pada data uji, kemudian menghasilkan tabel *confusion matrix* serta persentase akurasi model. Rangkaian tahapan tersebut, mulai dari pemanggilan dataset, penentuan label target, pemisahan data, pemodelan, hingga evaluasi, membentuk alur prediksi *customer churn* menggunakan metode *Decision Tree* di RapidMiner.

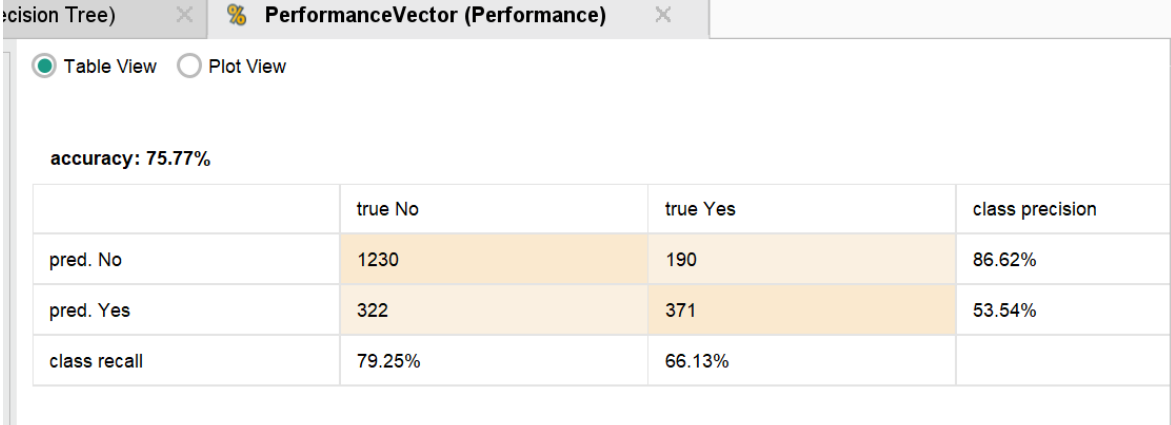
3.1 Dataset Directory

Penelitian ini menggunakan Telco Customer Churn Dataset yang diperoleh dari platform Kaggle. Dataset tersebut berisi data pelanggan perusahaan telekomunikasi yang digunakan untuk memprediksi kemungkinan pelanggan berhenti berlangganan (churn). Dataset terdiri dari 7.043 record pelanggan dan 21 atribut yang mencakup informasi demografi pelanggan, layanan yang digunakan, lama berlangganan (tenure), biaya bulanan (MonthlyCharges), total biaya (TotalCharges), serta status churn sebagai variabel target.

Variabel target dalam penelitian ini adalah atribut Churn yang memiliki dua kelas, yaitu "Yes" yang menunjukkan pelanggan berhenti berlangganan dan "No" yang menunjukkan pelanggan tetap menggunakan layanan. Dataset ini dipilih karena memiliki jumlah data yang cukup besar dan sesuai untuk penerapan metode klasifikasi menggunakan algoritma Decision Tree.

Sumber dataset: Kaggle – [Telco Customer Churn](#)

3.2 Evaluasi Performa Model (*Confusion Matrix*)



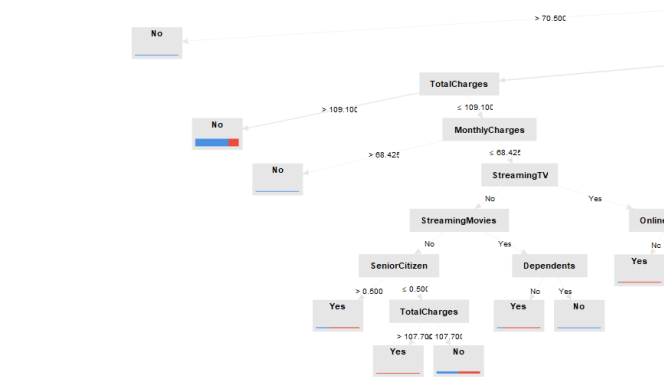
	true No	true Yes	class precision
pred. No	1230	190	86.62%
pred. Yes	322	371	53.54%
class recall	79.25%	66.13%	

Gambar 2. Confusion Matrix Pengujian Model Decision Tree.

Berdasarkan Gambar 2, model *Decision Tree* mencapai tingkat akurasi keseluruhan sebesar 75,77%. Dari total 2.113 data uji, sebanyak 1.230 pelanggan yang bertahan dan 371 pelanggan yang melakukan *churn* berhasil diprediksi dengan benar. Nilai *class precision* untuk prediksi pelanggan bertahan (*pred. No*) mencapai 86,62%, yang menunjukkan bahwa sebagian besar pelanggan yang diprediksi bertahan memang benar-benar tetap berlangganan. Sementara itu, nilai *class precision* untuk prediksi pelanggan *churn* (*pred. Yes*) sebesar 53,54%, karena masih terdapat 322 pelanggan bertahan yang diprediksi sebagai *churn*.

Dari sisi sensitivitas, nilai *class recall* untuk pelanggan bertahan sebesar 79,25%, yang berarti model mampu mengenali sebagian besar pelanggan setia. Adapun nilai *class recall* untuk pelanggan *churn* sebesar 66,13%, yang menunjukkan bahwa sekitar dua pertiga pelanggan yang benar-benar berhenti berlangganan berhasil dideteksi oleh model. Secara keseluruhan, hasil pengujian ini menunjukkan bahwa model yang dibangun menggunakan RapidMiner dapat dimanfaatkan sebagai instrumen bantu deteksi dini risiko kehilangan pelanggan di perusahaan.

3.3 Analisis Pohon Keputusan (*Tree Description*)

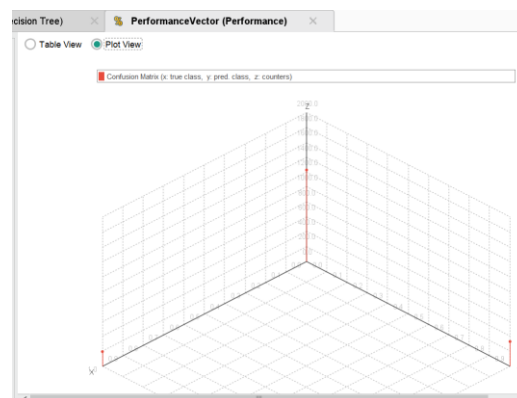


Gambar 3. Potongan Cabang Visualisasi Decision Tree pada RapidMiner.

Berdasarkan Gambar 3, proses klasifikasi dan penarikan aturan keputusan dilakukan secara bertahap melalui sejumlah variabel kunci. Pada percabangan atas, pelanggan disaring berdasarkan masa langganan (*tenure*). Pelanggan dengan nilai *tenure* lebih dari 70,500 langsung diklasifikasikan sebagai pelanggan yang cenderung bertahan (*No*). Pada kelompok pelanggan dengan *tenure* di bawah batas tersebut, model memeriksa atribut *TotalCharges*. Pelanggan dengan nilai *TotalCharges* lebih dari 109,100 diprediksi tidak melakukan *churn*, sedangkan pelanggan dengan nilai *TotalCharges* kurang dari atau sama dengan 109,100 dievaluasi lebih lanjut menggunakan atribut *MonthlyCharges* dengan ambang batas 68,425.

Bagi pelanggan dengan biaya bulanan rendah ($MonthlyCharges \leq 68,425$), keputusan selanjutnya dipengaruhi oleh layanan tambahan yang digunakan, yaitu *StreamingTV* dan *StreamingMovies*. Pelanggan yang tidak berlangganan layanan hiburan tersebut diteruskan ke pemeriksaan variabel demografi. Pada cabang terdalam, status lansia (*SeniorCitizen*) dan status memiliki tanggungan keluarga (*Dependents*) menjadi penentu akhir. Pelanggan non-lansia dengan total pengeluaran tertentu diprediksi bertahan (*No*), sedangkan pelanggan tanpa tanggungan (*Dependents = No*) pada percabangan layanan hiburan tertentu memiliki kecenderungan kuat untuk berhenti berlangganan (*Yes*).

3.4 Visualisasi Spasial Performa Model (*Plot View*)



Gambar 4. Visualisasi Grafik 3D Confusion Matrix pada RapidMiner.

Gambar 4 memetakan persebaran data hasil prediksi ke dalam tiga sumbu koordinat. Sumbu X merepresentasikan kelas data aktual (*true class*), sumbu Y merepresentasikan kelas hasil prediksi model (*predicted class*), dan sumbu Z yang bersifat vertikal menunjukkan jumlah frekuensi data (*counters*). Sebaran data terpusat pada sudut-sudut koordinat karena variabel target bersifat biner. Nilai frekuensi tertinggi ditunjukkan oleh batang berwarna merah yang menjulang paling tinggi pada sumbu Z, yang merepresentasikan 1.230 data pelanggan yang berhasil diklasifikasikan secara tepat pada kategori *true No*. Visualisasi ini memperjelas hasil *confusion matrix* sebelumnya dalam bentuk ruang tiga dimensi.

4. Kesimpulan

Berdasarkan hasil eksperimen dan pengujian yang telah dilakukan, algoritma *Decision Tree* pada perangkat lunak RapidMiner terbukti efektif dalam mengklasifikasikan *customer churn* pada layanan telekomunikasi dengan tingkat akurasi sebesar 75,77%, *class precision* sebesar 86,62% dan *class recall* sebesar 79,25% untuk kelas pelanggan bertahan. Faktor utama yang memengaruhi terjadinya *churn* adalah masa langganan (*tenure*) serta akumulasi biaya (*TotalCharges* dan *MonthlyCharges*). Pelanggan dengan masa langganan panjang menunjukkan tingkat kesetiaan tertinggi, sedangkan pelanggan dengan biaya total rendah yang tidak menggunakan layanan hiburan tambahan (*StreamingTV* dan *StreamingMovies*) merupakan segmen yang paling rawan berhenti berlangganan. Visualisasi tiga dimensi juga menunjukkan bahwa volume prediksi terbesar terkonsentrasi pada kategori *true No* sebanyak 1.230 data. Berdasarkan temuan tersebut, manajemen perusahaan disarankan untuk memfokuskan program retensi pada pelanggan dengan masa langganan pendek melalui insentif pada total tagihan, serta menawarkan paket *bundling* layanan hiburan untuk meningkatkan keterikatan pelanggan.

4.1 Saran

Berdasarkan keterbatasan dan hasil yang diperoleh, terdapat beberapa saran untuk implementasi praktis maupun pengembangan penelitian selanjutnya. Dari sisi praktis, manajemen perusahaan telekomunikasi dapat mengintegrasikan model prediksi ini ke dalam dasbor *Customer Relationship Management* (CRM), sehingga tim pemasaran memperoleh peringatan dini secara otomatis ketika profil pelanggan mendekati ambang risiko *churn*, seperti masa langganan pendek dan biaya bulanan rendah tanpa layanan hiburan.

Dari sisi pengembangan penelitian, algoritma berbasis *ensemble learning* seperti *Random Forest* atau *Gradient Boosted Trees* perlu diterapkan sebagai pembanding untuk melihat apakah akurasi dapat ditingkatkan melampaui capaian 75,77%. Proses *hyperparameter tuning*, misalnya menggunakan operator *Optimize Parameters* di RapidMiner pada parameter *maximal depth* atau *minimal gain*, juga perlu dilakukan untuk memperoleh konfigurasi pohon yang optimal. Selain itu, penelitian selanjutnya dapat menambahkan atribut yang lebih beragam, seperti skor kepuasan pelanggan, jumlah keluhan ke *call center*, dan kestabilan jaringan internet di lokasi pelanggan, agar pola perilaku *churn* dapat dianalisis secara lebih komprehensif.

Daftar Pustaka

- [1] S. Antoh, R. Herteno, I. Budiman, D. Kartini, dan M. I. Mazdadi, "Prediksi churn pelanggan telekomunikasi dengan optimalisasi seleksi fitur dan tuning hyperparameter pada algoritma klasifikasi C4.5," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 1, hlm. 60–67, 2025, doi: 10.14710/vol15iss1pp60-67.
- [2] B. Charbuty dan A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, hlm. 20–28, 2021, doi: 10.38094/jastt20165.

- [3] S. D. Damanik dan M. I. Jambak, "Klasifikasi customer churn pada telekomunikasi industri untuk retensi pelanggan menggunakan algoritma C4.5," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, hlm. 1303–1309, 2023, doi: 10.30865/klik.v3i6.829.
- [4] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2012.
- [5] M. Hofmann dan R. Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. CRC Press, 2016.
- [6] M. Junaidi dan R. Kaban, "Implementasi algoritma random forest untuk prediksi customer churn pada perusahaan telekomunikasi," *Jurnal SINTA: Sistem Informasi dan Teknologi Komputer*, vol. 3, no. 3, hlm. 165–173, 2026, doi: 10.61124/sinta.v3i3.320.
- [7] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, dan A. Khan, "Customer churn prediction using composite deep learning technique," *Scientific Reports*, vol. 13, 2023, doi: 10.1038/s41598-023-44396-w.
- [8] D. T. Larose dan C. D. Larose, *Data Mining and Predictive Analytics*. Wiley, 2019.
- [9] E. Prasetyo dan B. Sutopo, "Analisis Retensi Pelanggan di Era Digital," *Jurnal Sistem Informasi*, vol. 10, no. 2, hlm. 45–55, 2023.
- [10] D. Putriani, A. P. A. Prayogi, A. I. Shofyana, A. Ristyawan, dan E. Daniati, "Prediksi customer churn menggunakan algoritma decision tree," *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 8, no. 1, hlm. 85–94, 2024, doi: 10.29407/inotek.v8i1.4914.
- [11] S. Saleh dan S. Saha, "Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university," *SN Applied Sciences*, vol. 5, no. 7, 2023, doi: 10.1007/s42452-023-05389-6.
- [12] J. K. Sana, M. Z. Abedin, M. S. Rahman, dan M. S. Rahman, "A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection," *PLoS ONE*, vol. 17, no. 12, Art. no. e0278095, 2022, doi: 10.1371/journal.pone.0278095.
- [13] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, Art. no. 160, 2021, doi: 10.1007/s42979-021-00592-x.
- [14] P. N. Tan, M. Steinbach, A. Karpatne, dan V. Kumar, *Introduction to Data Mining*, 2nd ed. Pearson, 2020.
- [15] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, hlm. 1–37, 2008.