

Komparasi Algoritma Prediksi Churn Pelanggan Berbasis Machine Learning Pada Aplikasi Billing Admin

Nur Afrianto^{1*}, Aprilisa Arum Sari², Margaretha Evi Yuliana³

¹Teknik Informatika/Fakultas Ilmu
Komputer

Universitas Duta Bangsa

^{1*}220103161@udb.mhs.ac.id

²Teknik Informatika/Fakultas Ilmu
Komputer

Universitas Duta Bangsa

²aprilisa_arumsari@udb.ac.id

³Teknik Informatika/Fakultas Ilmu
Komputer

Universitas Duta Bangsa

³ margaretha@udb.ac.id

Abstrak— Pada perusahaan yang menerapkan model bisnis berbasis langganan, perpindahan pelanggan ke kompetitor menjadi permasalahan penting karena berdampak langsung terhadap stabilitas pendapatan dan keberlangsungan hubungan pelanggan. Penelitian ini bertujuan untuk membandingkan performa algoritma LightGBM dan Decision Tree dalam memprediksi churn pelanggan serta mengimplementasikan model terbaik ke dalam aplikasi Billing Admin. Pelaksanaan penelitian mengacu pada metodologi CRISP-DM yang terdiri atas tahap business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Dataset customer churn yang digunakan terdiri dari 2.800 data pelanggan dan dievaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score. Jika dibandingkan dengan Decision Tree, algoritma LightGBM terbukti menghasilkan performa yang lebih superior. Hal ini ditunjukkan oleh perolehan Accuracy 0,6274, Recall 0,6819, dan F1-Score 0,6770. Model LightGBM kemudian berhasil diintegrasikan ke dalam aplikasi Billing Admin berbasis CodeIgniter 4 menggunakan Flask API untuk melakukan prediksi churn pelanggan secara otomatis. Dengan demikian, model yang dihasilkan mampu membantu perusahaan dalam mengidentifikasi pelanggan yang berpotensi churn dan mendukung pengambilan keputusan dalam strategi retensi pelanggan.

Kata kunci— Billing admin, Churn Pelanggan, Decision Tree, LightGBM, Machine Learning.

Abstract— In subscription-based businesses, customer churn has become a critical issue as it directly affects revenue stability and long-term customer relationships. This study aims to compare the performance of the LightGBM and Decision Tree algorithms in predicting customer churn and to implement the best-performing model into a Billing Admin application. This study employed the CRISP-DM framework, encompassing six sequential stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. A churn dataset comprising 2,800 customer records was employed and assessed using the metrics of Accuracy, Precision, Recall, and F1-Score. Findings from the experiment show that LightGBM surpassed Decision Tree in terms of performance, yielding an Accuracy score of 0.6274, a Recall value of 0.6819, and an F1-Score of 0.6770. Furthermore, the LightGBM model was successfully integrated into a CodeIgniter 4-based Billing Admin application through a Flask API to perform automated customer churn prediction. The implemented model enables companies to identify customers with a high risk of churn and supports decision-making processes in customer retention strategies.

Keywords— Billing admin, Churn Pelanggan, Decision Tree, LightGBM, Machine Learning.

I. PENDAHULUAN

Dalam mengelola sebuah bisnis, mempertahankan loyalitas pelanggan agar terus menggunakan layanan yang diberikan merupakan tantangan yang penting bagi perusahaan. Persaingan bisnis yang semakin meningkat menyebabkan pelanggan memiliki banyak pilihan layanan sehingga memungkinkan pelanggan berpindah ke penyedia layanan lain. Kondisi ketika pelanggan berhenti menggunakan layanan perusahaan disebut customer churn[1].

Customer churn dapat menyebabkan kerugian finansial karena mempertahankan pelanggan yang telah ada dinilai lebih efisien dibandingkan melakukan akuisisi pelanggan baru yang memerlukan biaya lebih besar. [2]. Oleh karena itu, berbagai perusahaan mulai menerapkan strategi retensi pelanggan berbasis analisis data historis untuk mengidentifikasi pelanggan yang berpotensi berhenti menggunakan layanan[3].

Perkembangan machine learning memberikan solusi dalam memprediksi pelanggan yang berpotensi churn melalui pemanfaatan data historis pelanggan. Melalui machine learning, sistem dapat mengidentifikasi pola perilaku pelanggan sekaligus menyusun model prediksi yang nantinya dimanfaatkan untuk menunjang proses pengambilan keputusan di lingkungan bisnis. [4][5].

Algoritma machine learning telah banyak digunakan dalam berbagai penelitian untuk mengidentifikasi potensi customer churn. Hasil kajian[6] menunjukkan bahwa penggunaan metode ensemble pada prediksi churn pelanggan lebih efektif dibandingkan algoritma klasifikasi tunggal berdasarkan evaluasi terhadap beberapa model machine learning. Penelitian oleh[7] menunjukkan bahwa LightGBM mampu menghasilkan performa klasifikasi yang baik dalam mengidentifikasi pelanggan yang berpotensi churn. Selain itu, penelitian[8] menunjukkan bahwa LightGBM dapat memberikan interpretasi prediksi churn yang lebih baik melalui analisis fitur penting.

Selain LightGBM, algoritma Decision Tree juga banyak digunakan pada permasalahan klasifikasi karena memiliki struktur yang mudah dipahami dan mampu menghasilkan performa yang kompetitif. Penelitian[2] menunjukkan bahwa Decision Tree mampu mencapai tingkat akurasi yang tinggi pada kasus prediksi customer churn.

Meskipun berbagai penelitian telah berhasil membangun model prediksi churn dengan performa yang baik, sebagian besar penelitian masih berfokus pada evaluasi performa dan komparasi algoritma klasifikasi tanpa mengintegrasikan model tersebut ke dalam sistem operasional perusahaan[6]. Selain itu, studi komparasi antara LightGBM dan Decision Tree yang diintegrasikan langsung ke dalam aplikasi Billing Admin masih relatif terbatas[9][10].

II. METODOLOGI PENELITIAN

Jenis dan Sumber Data

Penelitian ini mengadopsi metode kuantitatif dengan memanfaatkan data customer churn untuk membangun dan menguji kinerja model machine learning. Data yang digunakan mencakup data pelanggan, data layanan, data transaksi, serta status churn pelanggan. Sumber data penelitian terdiri dari:

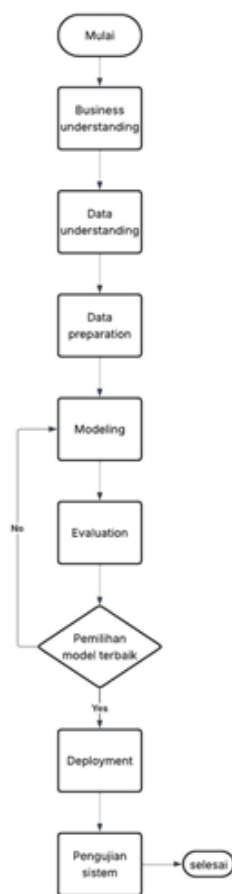
1. Data sekunder yang digunakan berupa dataset publik Customer Subscription Churn and Usage Patterns yang diperoleh melalui platform Kaggle dan dipublikasikan oleh Jay Joshi. Dataset tersebut terdiri atas 2.800 data pelanggan dengan 10 atribut yang merepresentasikan karakteristik pelanggan, penggunaan layanan, transaksi, dan status churn. Dataset ini digunakan sebagai data pelatihan dan pengujian model machine learning.
2. Data dummy digunakan pada tahap implementasi sistem untuk melakukan simulasi prediksi churn pada aplikasi Billing Admin berbasis CodeIgniter 4.

Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini dilaksanakan dengan cara :

1. Studi literatur dari jurnal, artikel ilmiah, dan penelitian sebelumnya yang berkaitan dengan customer churn, machine learning, LightGBM, dan Decision Tree.
2. Pengumpulan dataset public customer churn yang digunakan sebagai data pelatihan dan pengujian model machine learning.
3. Pembuatan data dummy yang digunakan untuk simulasi implementasi modul prediksi churn pada aplikasi Billing Admin.

Tahapan Penelitian



Gambar 1. Flowchart Tahapan penelitian Menggunakan CRISP-DM

Tahapan penelitian mengikuti kerangka CRISP-DM yang digunakan secara luas dalam proyek data mining. Kerangka tersebut mencakup pemahaman masalah bisnis, eksplorasi data, persiapan data, pembangunan model, evaluasi, serta implementasi model ke lingkungan operasional.

1. Business understanding

Menentukan tujuan penelitian dan kebutuhan prediksi churn pelanggan.

2. Data understanding

Mengumpulkan serta memahami karakteristik dataset customer churn.

3. Data Preparation

Tahapan preprocessing yang dilakukan mencakup transformasi label churn, eliminasi atribut yang tidak relevan, penerapan One Hot Encoding,

normalisasi data dengan StandardScaler, dan pemisahan dataset menjadi data latih serta data uji.

4. Modeling

Model yang dilatih adalah LightGBM dan Decision Tree.

5. Evaluation

Membandingkan performa model menggunakan Accuracy, Precision, Recall, F1-score.

6. Deployment

Tahap ini dilakukan untuk mengintegrasikan dan menerapkan model *machine learning* ke dalam sistem operasional, yaitu pada aplikasi Billing Admin berbasis CodeIgniter 4.

7. Pengujian Sistem

Melakukan simulasi prediksi churn menggunakan data dummy pada sistem yang telah terintegrasi.

Algoritma Penelitian

1. LightGBM

Algoritma ini menggunakan teknik pemisahan leaf-wise yang memungkinkannya menangani data skala besar dengan memori yang lebih hemat serta waktu komputasi yang lebih cepat dibandingkan metode boosting lainnya[11]. Pada penelitian ini, LightGBM dipilih karena mampu membangun model klasifikasi melalui mekanisme boosting yang efisien sehingga cocok digunakan pada dataset customer churn dengan jumlah data yang cukup besar.

2. Decision Tree

Pohon keputusan atau Decision Tree adalah salah satu pendekatan dalam klasifikasi yang mengadopsi model grafis berupa percabangan, di mana setiap simpulnya mewakili suatu kondisi atau aturan yang mudah dipahami dan sering

digunakan pada permasalahan klasifikasi data[2].

Evaluasi Model

Pengukuran performa model klasifikasi dilakukan dengan memanfaatkan confusion matrix serta sejumlah metrik, yakni Accuracy, Precision, Recall, dan F1-Score [12], dengan rincian sebagai berikut:

1. Accuracy

Metrik ini digunakan untuk mengukur rasio prediksi benar (positif dan negatif) keseluruhan data yang diuji.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

(1)

2. Precision

Metrik ini mengukur tingkat kecocokan antara hasil prediksi model dan nilai sebenarnya pada data pengujian.

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

(2)

3. Recall

Metrik ini menilai seberapa baik model mampu menemukan kembali informasi yang benar – benar ada dari seluruh data positif.

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

(3)

4. F1-Score

Nilai F1-Score diperoleh dari kombinasi harmonik presisi dan recall, dengan tujuan memberikan porsi bobot yang adil dan seimbang di antara keduanya.

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

(4)

III. HASIL DAN PEMBAHASAN

Preprocessing Data

Pada tahap preprocessing dilakukan beberapa proses untuk mempersiapkan dataset sebelum digunakan dalam pelatihan model machine learning. Dataset customer churn yang digunakan merupakan dataset publik Customer Subscription Churn and Usage Patterns yang diperoleh dari platform Kaggle. Dataset tersebut terdiri atas 2.800 data pelanggan dengan 10 atribut. Dataset tersebut memuat informasi pelanggan seperti biaya bulanan, lama berlangganan, penggunaan layanan, dan status churn pelanggan.

Tabel 1. Deskripsi Fitur Dataset

Fitur	Keterangan
User_id	Identitas pelanggan
Signup_date	Tanggal pendaftaran
Monthly_fee	Biaya bulanan pelanggan
Avg_weekly_usage_hours	Rata – rata penggunaan mingguan
Support_tickets	Jumlah tiket bantuan
Payment_failures	Jumlah kegagalan pembayaran
Tenure_months	Lama berlangganan
Last_login_days_ago	Hari sejak login terakhir
Plan_type	Jenis paket layanan
churn	Status churn pelanggan

Tahap preprocessing diawali dengan transformasi label churn dari bentuk kategorikal menjadi numerik, yaitu “yes” menjadi 1 dan “no” menjadi 0 agar dapat diproses oleh algoritma machine learning. Selanjutnya dilakukan konversi atribut signup_date ke format datetime. Kolom user_id dan signup_date kemudian dihapus karena tidak digunakan dalam proses pelatihan model.

Data kategorikal pada atribut *plan_type* ditransformasikan menggunakan Teknik One hot encoding sehingga data dapat diproses oleh algoritma klasifikasi. Penggunaan One Hot Encoding bertujuan mengubah data kategorikal menjadi representasi numerik tanpa memberikan hubungan ordinal antar kategori. Teknik ini banyak digunakan pada proses persiapan data untuk meningkatkan kompatibilitas data dengan algoritma machine learning[13]. Setelah proses encoding selesai, dilakukan standardisasi data menggunakan *StandardScaler* untuk menyamakan rentang nilai antar fitur numerik agar proses pelatihan model menjadi lebih optimal. Standardisasi data menggunakan *StandardScaler* dilakukan untuk mengurangi perbedaan skala antar fitur sehingga proses pelatihan model menjadi lebih stabil dan optimal. Pendekatan ini telah banyak digunakan untuk meningkatkan performa model klasifikasi pada berbagai kasus prediksi berbasis machine learning[14].

Selanjutnya, dilakukan analisis distribusi kelas terhadap label customer churn untuk mengidentifikasi kemungkinan terjadinya *class imbalance*. Hasil analisis menunjukkan bahwa dataset terdiri atas 1.605 data churn (57,3%) dan 1.195 data tidak churn (42,7%), sehingga distribusi kedua kelas masih tergolong relatif seimbang. Oleh karena itu, penelitian ini tidak menerapkan teknik penanganan *class imbalance* seperti SMOTE (Synthetic Minority Over-sampling Technique) maupun *class weighting* karena distribusi kelas pada dataset masih dianggap memadai untuk proses pelatihan model.

Tahap pembagian dataset dilakukan dengan menetapkan 70% data sebagai data pelatihan dan 30% sebagai data pengujian menggunakan metode train-test split. Data training dikhususkan untuk keperluan pembentukan model, adapun data testing difungsikan sebagai alat uji untuk menilai kemampuan model yang telah dilatih.

Hasil Training Model

Pada tahap modeling, algoritma Decision tree dan LightGBM digunakan untuk melakukan klasifikasi customer churn berdasarkan dataset yang telah melalui proses preprocessing. Dataset dipisahkan menjadi data pelatihan sebanyak 70% serta data pengujian sebanyak 30% menggunakan metode train-test split.

Model dilatih menggunakan data training, sementara evaluasi kemampuan model dalam memprediksi churn pelanggan dilakukan dengan data testing. Untuk menilai performa model, digunakan empat metrik klasifikasi, yaitu Accuracy, Precision, Recall, dan F1-Score.

Tabel 2. Hasil Uji Performa Model

Algoritma	Accuracy	Precision	Recall	F1-score
Decision tree	0,6190	0,6809	0,6299	0,6544
LightGBM	0,6274	0,6271	0,6819	0,6770

Berdasarkan hasil pengujian, algoritma LightGBM menghasilkan performa yang lebih baik dibandingkan Decision Tree. Meskipun demikian, nilai akurasi yang diperoleh masih berada pada kisaran 62%, sehingga menunjukkan bahwa prediksi customer churn pada dataset yang digunakan masih merupakan permasalahan yang kompleks. Rendahnya performa model dapat dipengaruhi oleh beberapa faktor, antara lain jumlah fitur prediktif yang terbatas, karakteristik data pelanggan yang belum sepenuhnya merepresentasikan perilaku churn, serta belum dilakukannya optimasi hiperparameter secara mendalam. Selain itu, penelitian ini hanya menggunakan fitur yang tersedia pada dataset publik sehingga kemungkinan masih terdapat faktor-faktor lain yang memengaruhi customer churn tetapi belum tercakup di dalam data.

Hasil yang diperoleh dalam studi ini konsisten dengan temuan pada referensi [7] yang mengindikasikan efektivitas algoritma LightGBM dalam melakukan klasifikasi customer churn melalui penerapan mekanisme boosting berbasis decision tree. LightGBM

menghasilkan nilai recall yang lebih besar, yang menandakan kemampuannya yang lebih baik dalam menangkap sinyal pelanggan yang cenderung churn. Dalam konteks bisnis billing, kemampuan tersebut lebih penting dibandingkan sekadar meningkatkan akurasi karena perusahaan dapat melakukan tindakan retensi lebih dini terhadap pelanggan berisiko.

Hasil yang diperoleh dalam studi ini sejalan dengan temuan pada referensi [8] yang mengindikasikan kemampuan LightGBM mampu menghasilkan performa prediksi churn yang kompetitif serta dapat memberikan interpretasi terhadap faktor-faktor yang memengaruhi churn pelanggan melalui analisis fitur penting. Di samping itu, temuan dari studi ini selaras dengan riset [6] yang melakukan komparasi antar berbagai metode machine learning dalam kasus churn pelanggan. Penelitian tersebut menunjukkan bahwa metode berbasis ensemble dan boosting cenderung memberikan performa yang lebih baik dibandingkan algoritma klasifikasi tunggal seperti Decision Tree karena mampu menangani pola data yang lebih kompleks.

Berdasarkan temuan evaluasi, algoritma LightGBM ditentukan sebagai model yang memberikan performa terbaik untuk digunakan dalam implementasi sistem prediksi churn pada aplikasi Billing Admin.

Integrasi Sistem

Model terbaik hasil proses training, yaitu LightGBM, diintegrasikan ke dalam aplikasi Billing Admin berbasis CodeIgniter 4 menggunakan Flask API sebagai layanan machine learning. Pendekatan integrasi menggunakan Flask API memungkinkan pemisahan antara sistem aplikasi dan model machine learning sehingga proses pengembangan dan pemeliharaan sistem menjadi lebih fleksibel. Pendekatan serupa telah digunakan dalam pengembangan platform operasional berbasis machine learning menggunakan framework Flask[10]. Model yang telah dilatih sebelumnya disimpan dalam format .pkl dan digunakan oleh

Flask untuk melakukan proses prediksi churn pelanggan.

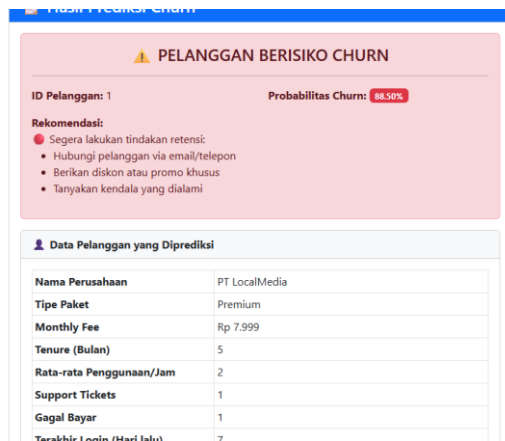
Proses integrasi sistem diawali ketika admin menginput data pelanggan dan data aktivitas pelanggan melalui form pada aplikasi Billing Admin. Data yang dimasukkan meliputi informasi pelanggan seperti tipe paket, biaya bulanan, lama berlangganan, rata-rata penggunaan layanan, jumlah tiket bantuan, jumlah gagal bayar, dan aktivitas login pelanggan.

Selanjutnya aplikasi CodeIgniter 4 mengirimkan data tersebut ke Flask API melalui endpoint prediksi menggunakan metode HTTP request. Flask API kemudian memproses data menggunakan model LightGBM dan menghasilkan output berupa status prediksi churn serta nilai probabilitas churn pelanggan.

Hasil prediksi kemudian ditampilkan kembali pada dashboard aplikasi Billing Admin dalam bentuk daftar pelanggan berisiko churn beserta persentase probabilitas churn. Sistem juga menyediakan halaman detail prediksi yang menampilkan informasi pelanggan, status prediksi, probabilitas churn, serta rekomendasi tindakan retensi pelanggan. Implementasi model machine learning ke dalam sistem operasional merupakan bagian penting dari tahapan deployment karena memungkinkan hasil prediksi dimanfaatkan secara langsung oleh pengguna sistem. Integrasi semacam ini menjadi salah satu faktor penting dalam keberhasilan penerapan machine learning pada lingkungan produksi[9].

No	ID Pelanggan	Nama Pelanggan	Status Risiko	Probabilitas Churn	Aksi
1	1	PT LokalMedia	Berisiko Tinggi	88.50%	Detail
2	2	PT SejatiEra Utama	Low	27.81%	Detail

Gambar 2. Dashboard daftar pelanggan berisiko churn



Gambar 3. Hasil Prediksi churn



Gambar 4. Detail Prediksi Churn Pelanggan

Implementasi sistem ini bertujuan untuk membantu admin dalam melakukan monitoring pelanggan yang berpotensi churn sehingga perusahaan dapat melakukan tindakan preventif lebih awal untuk mempertahankan pelanggan. Selain itu, penerapan model prediksi churn secara langsung pada aplikasi operasional mendukung konsep deployment machine learning yang skalabel dan dapat diterapkan pada lingkungan bisnis modern sebagaimana dijelaskan oleh[15].

IV. KESIMPULAN

Berdasarkan hasil penelitian, algoritma LightGBM dan Decision Tree berhasil diterapkan untuk membangun model prediksi customer churn menggunakan dataset publik yang terdiri atas 2.800 data pelanggan. Hasil evaluasi menunjukkan bahwa algoritma LightGBM memiliki performa yang lebih baik dibandingkan Decision Tree dengan nilai Accuracy sebesar 0,6274, Recall sebesar 0,6819, dan F1-Score sebesar 0,6770 sehingga dipilih sebagai model terbaik untuk diimplementasikan pada sistem. Meskipun demikian, performa yang diperoleh masih menunjukkan bahwa prediksi customer churn pada dataset yang digunakan merupakan permasalahan yang kompleks

sehingga masih terdapat peluang untuk meningkatkan kinerja model.

Model LightGBM berhasil diintegrasikan ke dalam aplikasi Billing Admin berbasis CodeIgniter 4 menggunakan Flask API sebagai layanan *machine learning*. Sistem mampu melakukan prediksi churn pelanggan, menampilkan probabilitas churn, serta menyajikan informasi pelanggan yang berpotensi churn sebagai pendukung pengambilan keputusan dalam strategi retensi pelanggan.

Penelitian ini memiliki keterbatasan, antara lain penggunaan dataset publik dengan jumlah fitur yang terbatas serta belum dilakukannya optimasi hiperparameter secara menyeluruh. Oleh karena itu, penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, menerapkan optimasi hiperparameter, serta mengeksplorasi teknik penanganan class imbalance apabila diperlukan. Selain itu, penelitian berikutnya dapat membandingkan algoritma lain, seperti XGBoost, CatBoost, atau Random Forest, untuk memperoleh performa prediksi yang lebih optimal.

UCAPAN TERIMA KASIH

Keberhasilan penyelesaian penelitian ini tidak lepas dari peran serta berbagai pihak yang telah berkontribusi. Pertama, penulis menyampaikan rasa hormat dan terima kasih kepada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta, yang telah memberikan dukungan fasilitas selama kegiatan riset berlangsung. Kedua, penghargaan setinggi-tingginya penulis sampaikan kepada dosen pembimbing yang selalu meluangkan waktunya untuk memberi arahan, koreksi mendalam, serta motivasi yang membangun. Atas bimbingan yang diberikan, penelitian ini dapat berjalan dengan baik dan mencapai hasil yang diharapkan.

REFERENSI

- [1] Y. Yudiana, A. Yulia Agustina, and D. Nur Khofifah, "Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan," *Indones. J. Islam. Econ. Bus.*, vol. 8, no. 1, pp. 1–20, 2023, doi: <https://doi.org/10.30631/ijoeib.v8i1.1710>.
- [2] D. Putriani, A. P. A. Prayogi, A. I. Shofyana, A. Ristyawan, and E. Daniati, "Prediksi Customer Churn Menggunakan Algoritma Decision Tree," *Inotek*, vol. 8, pp. 85–94, 2024, doi: <https://doi.org/10.29407/inotek.v8i1.4914>.
- [3] A. Wahid and A. Muliawan, "Strategi Retensi Pelanggan Berbasis Historis: Optimalisasi," vol. 11, no. 2, pp. 133–141, 2025, doi: <https://doi.org/10.55679/semantik.v11i2.237>.
- [4] H. Kerner, J. Campbell, and M. Strickland, "Introduction to machine learning," *Mach. Learn. Planet. Sci.*, vol. 01, pp. 1–24, 2022, doi: [10.1016/B978-0-12-818721-0.00007-0](https://doi.org/10.1016/B978-0-12-818721-0.00007-0).
- [5] E. Alpaydin, "Introduction to Machine Learning, Fourth Edition," *MIT Press*, vol. 4, pp. 1–4, 2020, [Online]. Available: https://books.google.com.br/books?hl=pt-BR&lr=&id=uZnSDwAAQBAJ&oi=fnd&pg=PR7&ots=xOuVtwKntU&sig=3BWSHbEyRmi77S3Zuk4SLXh1MeM&redir_esc=y#v=onepage&q&f=false
- [6] G. Sam, P. Asuquo, and B. Stephen, "Customer Churn Prediction using Machine Learning Models," vol. 26, no. 2, pp. 181–193, 2024, doi: [10.9734/JERR/2024/v26i21081](https://doi.org/10.9734/JERR/2024/v26i21081).
- [7] I. Z. A. Illah, W. S. Jauharis Sapu, and A. T. Damaliana, "Implementasi Metode Klasifikasi LightGBM dan Analisis Survival dalam Memprediksi Pelanggan Churn," *J. Komtika (Komputasi dan Inform.)*, vol. 8, no. 1, pp. 43–53, 2024, doi: [10.31603/komtika.v8i1.11194](https://doi.org/10.31603/komtika.v8i1.11194).
- [8] R. Sinaga and S. Widiyanto, "Understanding telecommunication customer churn : Insights from LightGBM predictive modelling and SHAP feature interpretation Understanding Telecommunication Customer Churn : Insights from LightGBM Predictive Modelling and SHAP Feature Interpretation," 2023, doi: [10.7454/amj.v15i1.1229](https://doi.org/10.7454/amj.v15i1.1229).
- [9] J. A. Pruneski *et al.*, "The development and deployment of machine learning models," *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 30, no. 12, pp. 3917–3923, 2022, doi: [10.1007/s00167-022-07155-4](https://doi.org/10.1007/s00167-022-07155-4).
- [10] M. S. Bonney, M. De Angelis, M. D. Borgo, S. Beregi, N. Jamia, and D. J. Wagg, "Development of a digital twin operational platform using Python Flask," pp. 1–14, 2022, doi: [10.1017/dce.2022.1](https://doi.org/10.1017/dce.2022.1).
- [11] L. P. Nugroho, R. E. Saputro, and F. S. Utomo, "Performance Comparison Of Xgboost Lightgbm And Lstm For E- Commerce Repeat Buyer Prediction," *J. Tek. Inform.*, vol. 7, no. 1, pp. 730–742, 2026, doi: <https://doi.org/10.52436/1.jutif.2026.7.1.5746>.
- [12] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min.*, pp. 1–23, 2023, doi: [10.1186/s13040-023-00322-4](https://doi.org/10.1186/s13040-023-00322-4).
- [13] P. Ekaterina and K. Alexey, "Encoding categorical data: Is there yet anything 'hotter' than one-hot encoding?," 2023, doi: <https://doi.org/10.48550/arXiv.2312.16930>.
- [14] F. Aldi, F. Hadi, N. A. Rahmi, S. Defit, I. Technology, and D. Program, "STANDARDSCALER ' S POTENTIAL IN ENHANCING BREAST CANCER," *J. Appl. Eng. Technol. Sci.*, vol. 5, no. 1, pp. 401–413, 2023.
- [15] P. Srikanth and S. Raju, "Scalable Deployment of Machine Learning Models on Kubernetes Clusters: A DevOps Perspective," vol. 7, no. 6, pp. 11640–11648, 2024, doi: [10.15662/IJRAI.2024.0706003](https://doi.org/10.15662/IJRAI.2024.0706003).