

Pemanfaatan Data Mining untuk Klasifikasi Customer Churn Menggunakan Algoritma Random Forest dalam Mendukung Strategi Bisnis

Desty Mutiara Putri^{1*}, Resvi Despina Mulyani², Firda Mawarni³

^{1,2,3}Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Pamulang, Tangerang Selatan, Indonesia

¹destymutiaraa12345@gmail.com, ²resvimulyani@gmail.com, ³firdamawarni07@gmail.com
(Email Correspondence: destymutiaraa12345@gmail.com)

Received: April 4, 2026 | Revised: April 8, 2026 | Accepted: April 30, 2026

Abstrak

Peningkatan persaingan di industri telekomunikasi mendorong perusahaan untuk mempertahankan pelanggan karena customer churn dapat menyebabkan penurunan pendapatan dan meningkatnya biaya akuisisi pelanggan baru. Namun, identifikasi pelanggan yang berpotensi churn masih menjadi tantangan, terutama pada data dengan distribusi kelas yang tidak seimbang. Penelitian ini bertujuan membangun model prediksi customer churn menggunakan algoritma Random Forest serta mengidentifikasi faktor-faktor yang memengaruhi keputusan pelanggan menghentikan layanan. Penelitian menggunakan Telco Customer Churn Dataset yang terdiri atas 7.043 data pelanggan dengan atribut tenure, TechSupport, Contract, dan MonthlyCharges sebagai variabel prediktor serta Churn sebagai variabel target. Tahapan penelitian meliputi prapemrosesan data, pembentukan model, dan evaluasi menggunakan confusion matrix serta ROC-AUC. Hasil pengujian menunjukkan model memperoleh accuracy 79,21%, precision 66,27%, recall 44,12%, dan ROC-AUC 83,94%. Analisis feature importance menunjukkan bahwa Contract merupakan faktor paling berpengaruh terhadap churn, diikuti tenure, MonthlyCharges, dan TechSupport. Temuan ini menunjukkan bahwa Random Forest efektif digunakan untuk memprediksi customer churn dan mendukung strategi retensi pelanggan berbasis data.

Kata kunci: Customer Churn, Random Forest, Data Mining, Klasifikasi, Telekomunikasi.

Abstract

Increasing competition in the telecommunications industry has intensified the need for customer retention, as customer churn can reduce revenue and increase customer acquisition costs. However, identifying customers with a high risk of churn remains challenging, particularly in datasets with imbalanced class distributions. This study aims to develop a customer churn prediction model using the Random Forest algorithm and identify factors influencing customer attrition. The study employs the Telco Customer Churn Dataset consisting of 7,043 customer records, with tenure, TechSupport, Contract, and MonthlyCharges as predictor variables and Churn as the target variable. The research process includes data preprocessing, model development, and evaluation using a confusion matrix and ROC-AUC metrics. Results show that the model achieved an accuracy of 79.21%, precision of 66.27%, recall of 44.12%, and ROC-AUC of 83.94%. Feature importance analysis reveals that Contract is the most influential factor, followed by tenure, MonthlyCharges, and TechSupport. These findings indicate that Random Forest provides reliable performance for customer churn prediction and supports data-driven customer retention strategies.

Keywords: Customer Churn, Random Forest, Data Mining, Classification, Telecommunications.

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong transformasi signifikan pada berbagai sektor industri, khususnya layanan berbasis digital dan sistem berlangganan. Perusahaan saat ini semakin memanfaatkan teknologi informasi untuk meningkatkan efisiensi operasional, memahami perilaku pelanggan, serta mendukung pengambilan keputusan berbasis data. Dalam industri telekomunikasi, data pelanggan menjadi aset penting karena dapat digunakan untuk menganalisis pola penggunaan layanan serta tingkat loyalitas pelanggan melalui pendekatan data mining.

Salah satu teknik data mining yang umum digunakan adalah klasifikasi, yaitu proses pengelompokan data ke dalam kategori tertentu berdasarkan karakteristiknya. Dalam konteks bisnis, teknik ini banyak diterapkan untuk memprediksi customer churn, yaitu kondisi ketika pelanggan berhenti menggunakan layanan. Customer churn menjadi isu penting karena dapat berdampak pada penurunan pendapatan perusahaan serta peningkatan biaya akuisisi pelanggan baru, sehingga perusahaan perlu memahami faktor-faktor yang memengaruhinya [1].

Berbagai penelitian terdahulu menunjukkan bahwa algoritma Random Forest merupakan salah satu metode klasifikasi yang efektif untuk memprediksi customer churn. Kemampuan algoritma ini dalam menggabungkan banyak decision tree melalui pendekatan ensemble learning memungkinkan peningkatan stabilitas prediksi sekaligus mengurangi risiko overfitting [2], [3]. Selain itu, beberapa studi mengidentifikasi bahwa karakteristik pelanggan, seperti lama berlangganan, biaya layanan bulanan, dan jenis kontrak, memiliki keterkaitan yang kuat terhadap kecenderungan pelanggan untuk menghentikan penggunaan layanan [4]. Meskipun demikian, permasalahan ketidakseimbangan kelas (*imbalanced dataset*) masih menjadi tantangan dalam proses prediksi churn karena jumlah pelanggan yang tidak melakukan churn umumnya jauh lebih besar dibandingkan pelanggan yang melakukan churn. Kondisi tersebut dapat menyebabkan model cenderung lebih baik dalam mengenali kelas mayoritas dibandingkan kelas minoritas sehingga menurunkan kemampuan deteksi pelanggan yang berpotensi churn [5], [6].

Kondisi ketidakseimbangan data tersebut dapat menyebabkan model cenderung bias terhadap kelas mayoritas sehingga menurunkan kemampuan deteksi pada pelanggan churn, terutama dalam metrik recall. Selain itu, belum semua penelitian secara optimal memanfaatkan analisis feature importance untuk mengidentifikasi variabel yang paling berpengaruh terhadap churn, sehingga interpretasi hasil model masih terbatas dalam mendukung pengambilan keputusan bisnis.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun model prediksi customer churn menggunakan algoritma Random Forest berdasarkan karakteristik pelanggan layanan telekomunikasi, meliputi tenure, monthly charges, contract, dan tech support. Selain itu, penelitian ini juga menganalisis faktor-faktor yang paling berpengaruh terhadap churn melalui feature importance. Hasil penelitian diharapkan dapat membantu perusahaan dalam meningkatkan strategi customer retention berbasis data serta mengurangi tingkat customer churn secara lebih efektif.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk membangun model prediksi Customer Churn pada layanan telekomunikasi. Proses penelitian menerapkan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri atas tahapan business understanding, data understanding, data preparation, modeling, dan evaluation. Pendekatan tersebut dipilih karena mampu memberikan alur kerja yang terstruktur dalam proses pengolahan data, pembangunan model, hingga evaluasi hasil klasifikasi. Fokus utama penelitian diarahkan pada identifikasi pelanggan yang berpotensi berhenti menggunakan layanan berdasarkan karakteristik penggunaan layanan telekomunikasi. Kajian terdahulu menunjukkan bahwa penerapan teknik data mining dapat membantu perusahaan dalam mengenali pola perilaku pelanggan serta mendukung strategi retensi pelanggan secara lebih efektif [7]. Penelitian sebelumnya menunjukkan bahwa algoritma Random Forest mampu memberikan performa klasifikasi yang stabil pada data pelanggan yang memiliki karakteristik beragam dan kompleks. Kemampuan tersebut menjadikan Random Forest banyak digunakan dalam berbagai kasus prediksi karena mampu menangani variasi data serta mengurangi risiko kesalahan klasifikasi [8].

Dataset yang digunakan dalam penelitian ini berasal dari Telco Customer Churn Dataset yang diperoleh melalui platform Kaggle. Dataset tersebut terdiri dari 7.043 data pelanggan dengan 21 atribut, yaitu *customerID*, *gender*, *SeniorCitizen*, *Partner*, *Dependents*, *tenure*, *PhoneService*, *MultipleLines*, *InternetService*, *OnlineSecurity*, *OnlineBackup*, *DeviceProtection*, *TechSupport*, *StreamingTV*, *StreamingMovies*, *Contract*, *PaperlessBilling*, *PaymentMethod*, *MonthlyCharges*, *TotalCharges*, dan *Churn*. Namun, penelitian ini hanya menggunakan empat atribut prediktor, yaitu *tenure*, *TechSupport*, *Contract*, dan *MonthlyCharges*, serta *Churn* sebagai variabel target. Pemilihan atribut tersebut dilakukan karena dianggap memiliki pengaruh terhadap keputusan pelanggan untuk tetap berlangganan atau menghentikan layanan telekomunikasi yang digunakan. Pada proses data understanding, dilakukan analisis terhadap struktur data, tipe atribut, distribusi kelas, serta hubungan antarvariabel untuk memahami pola data yang digunakan dalam penelitian. Penggunaan atribut pelanggan sebagai indikator prediksi churn menunjukkan bahwa karakteristik pelanggan memiliki keterkaitan terhadap keputusan pelanggan untuk berhenti menggunakan layanan. Selain itu, data historis pelanggan juga dapat dimanfaatkan untuk

mempelajari pola perilaku churn melalui pendekatan machine learning sehingga perusahaan dapat melakukan tindakan preventif terhadap potensi kehilangan pelanggan [9], [10].

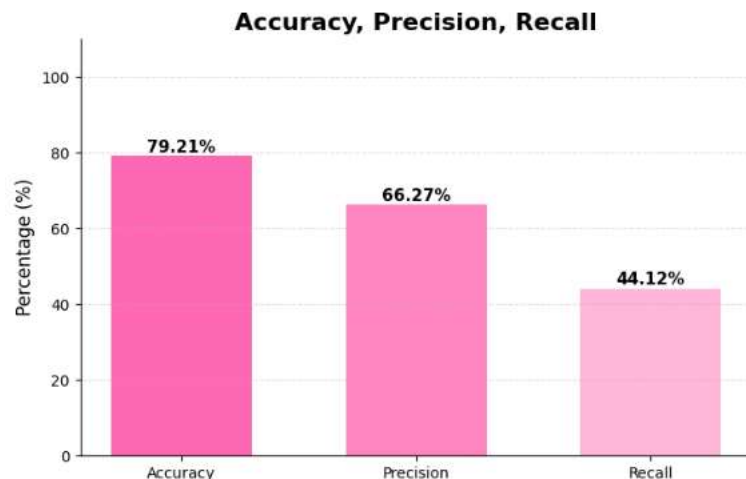
Sebelum proses pemodelan dilakukan, dataset terlebih dahulu melalui tahap data preparation untuk memastikan kualitas data yang digunakan dalam penelitian. Tahapan ini mencakup penanganan missing value, penghapusan data duplikat, pemeriksaan tipe data, serta deteksi dan penanganan outlier menggunakan metode *Interquartile Range* (IQR). Proses preprocessing menjadi bagian penting dalam penelitian machine learning karena kualitas data sangat memengaruhi performa model yang dihasilkan. Data yang belum melalui proses pembersihan secara optimal berpotensi menghasilkan model dengan performa klasifikasi yang kurang stabil serta meningkatkan kemungkinan munculnya bias pada hasil prediksi [11]. Setelah proses pembersihan selesai dilakukan, atribut kategorikal ditransformasikan menggunakan teknik one-hot encoding agar dapat diproses oleh algoritma machine learning berbasis numerik. Dataset kemudian dibagi menjadi data latih dan data uji menggunakan metode train-test split dengan rasio 80:20 untuk mengurangi risiko overfitting dan menguji kemampuan generalisasi model terhadap data baru.

Proses modeling dilakukan menggunakan algoritma Random Forest Classifier yang diimplementasikan melalui library Scikit-learn pada bahasa pemrograman Python. Algoritma ini dipilih karena memiliki kemampuan klasifikasi yang baik dalam menangani data berukuran besar serta mampu meningkatkan stabilitas prediksi melalui pendekatan ensemble learning. Algoritma Random Forest bekerja dengan membangun sejumlah decision tree secara acak, kemudian menggabungkan hasil prediksi dari setiap tree menggunakan mekanisme majority voting sehingga mampu menghasilkan performa klasifikasi yang lebih konsisten dibandingkan metode single decision tree [12]. Sejumlah penelitian melaporkan bahwa Random Forest menunjukkan kinerja klasifikasi yang efektif pada kasus prediksi customer churn. Melalui mekanisme ensemble learning, algoritma ini mampu menghasilkan tingkat akurasi yang relatif tinggi serta performa yang lebih konsisten dibandingkan beberapa pendekatan klasifikasi konvensional [13]. Selain digunakan untuk proses klasifikasi, penelitian ini juga memanfaatkan analisis feature importance untuk mengidentifikasi atribut yang memiliki kontribusi terbesar terhadap perilaku churn pelanggan.

Tahap akhir penelitian dilakukan melalui proses evaluation menggunakan confusion matrix dan beberapa metrik evaluasi, yaitu accuracy, precision, recall, F1-score, dan ROC-AUC. Penggunaan beberapa metrik evaluasi bertujuan untuk memperoleh gambaran performa model secara lebih menyeluruh, terutama dalam mendeteksi pelanggan churn yang termasuk ke dalam kelas minoritas. Nilai recall menjadi perhatian penting karena berkaitan dengan kemampuan model dalam mengidentifikasi pelanggan yang berpotensi berhenti menggunakan layanan. Kondisi imbalanced dataset dapat memengaruhi performa model klasifikasi, khususnya pada kemampuan deteksi kelas minoritas. Oleh karena itu, hasil evaluasi model dalam penelitian ini digunakan untuk menganalisis efektivitas algoritma Random Forest dalam memprediksi customer churn serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap perilaku pelanggan berdasarkan karakteristik layanan telekomunikasi.

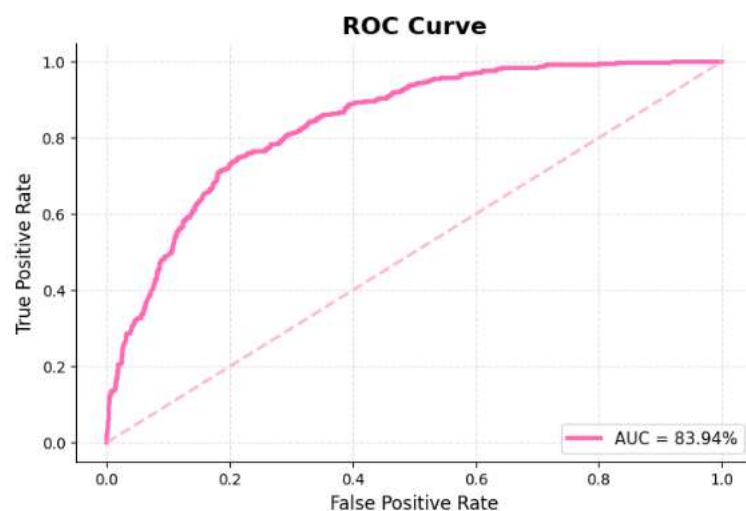
3. HASIL DAN PEMBAHASAN

Pada tahap eksperimen, dilakukan pembangunan model prediksi churn menggunakan algoritma Random Forest yang diimplementasikan dalam bentuk Random Forest Classifier. Dataset yang digunakan dibagi menjadi data training dan data testing dengan proporsi tertentu untuk memastikan evaluasi model yang objektif. Model dilatih menggunakan data training yang terdiri dari variabel fitur pelanggan seperti lama berlangganan, biaya bulanan, jenis kontrak, dan layanan dukungan teknis, serta variabel target yaitu status churn. Setelah proses pelatihan, model digunakan untuk melakukan prediksi terhadap data testing. Hasil prediksi kemudian dibandingkan dengan data aktual untuk mengukur kemampuan model dalam mengidentifikasi pelanggan yang berpotensi churn.



Gambar 1. Hasil Evaluasi Model (*Accuracy*, *Precision*, dan *Recall*)

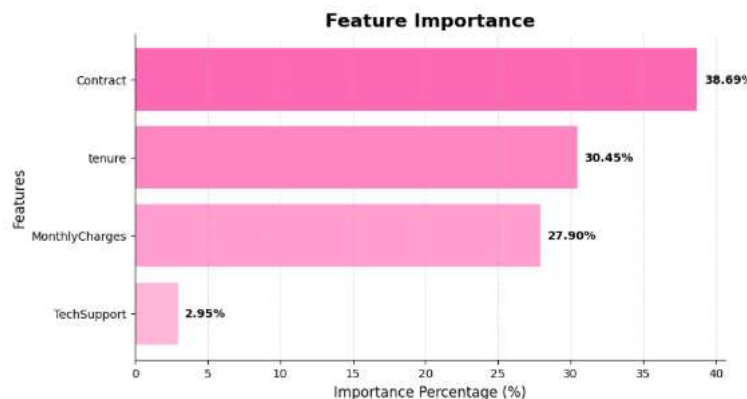
Berdasarkan hasil evaluasi model pada Gambar 1, diperoleh nilai accuracy sebesar 79,21%, precision sebesar 66,27%, dan recall sebesar 44,12%. Nilai accuracy menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam melakukan klasifikasi pelanggan churn dan non-churn secara keseluruhan. Selain itu, nilai precision yang cukup tinggi menunjukkan bahwa sebagian besar pelanggan yang diprediksi churn oleh model memang benar-benar termasuk ke dalam kategori churn. Namun demikian, nilai recall yang masih relatif rendah menunjukkan bahwa model belum mampu mendeteksi seluruh pelanggan churn secara optimal. Kondisi tersebut dapat dipengaruhi oleh distribusi data yang tidak seimbang (*imbalanced dataset*) antara kelas churn dan non-churn sehingga model cenderung lebih mudah mengenali kelas mayoritas dibandingkan kelas minoritas. Kondisi ini sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa ketidakseimbangan distribusi kelas dapat menurunkan kemampuan model dalam mengidentifikasi data pada kelas minoritas, sehingga performa deteksi terhadap kasus churn menjadi kurang optimal [14].



Gambar 2. Hasil Evaluasi Model Menggunakan ROC-AUC

Selain menggunakan metrik accuracy, precision, dan recall, penelitian ini juga melakukan evaluasi menggunakan ROC-AUC untuk mengukur kemampuan model dalam membedakan pelanggan churn dan non-churn. Berdasarkan Gambar 2, diperoleh nilai ROC-AUC sebesar 83,94% yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang cukup baik dalam memisahkan kedua kelas. Nilai tersebut mengindikasikan bahwa model mampu menghasilkan probabilitas prediksi yang relatif stabil dalam mengidentifikasi pelanggan yang berpotensi churn. Hasil evaluasi ini memperlihatkan bahwa algoritma

Random Forest mampu memberikan performa yang konsisten pada data pelanggan dengan karakteristik yang kompleks, sebagaimana juga ditunjukkan pada penelitian sebelumnya [15].



Berdasarkan hasil analisis feature importance, variabel yang paling berpengaruh terhadap churn adalah jenis kontrak (Contract) dengan nilai sebesar 38,69%, diikuti oleh lama berlangganan (tenure) sebesar 30,45%, biaya bulanan (MonthlyCharges) sebesar 27,90%, dan layanan dukungan teknis (TechSupport) sebesar 2,95%. Hasil ini menunjukkan bahwa jenis kontrak memiliki pengaruh paling signifikan terhadap keputusan pelanggan untuk berhenti menggunakan layanan. Pelanggan dengan kontrak jangka pendek, masa berlangganan yang relatif singkat, serta biaya bulanan yang tinggi cenderung memiliki potensi churn yang lebih besar. Sementara itu, layanan dukungan teknis memiliki pengaruh yang relatif kecil dibandingkan variabel lainnya. Temuan ini dapat menjadi dasar bagi perusahaan dalam memahami karakteristik pelanggan yang berisiko tinggi terhadap churn.

Hasil analisis tersebut dapat dimanfaatkan perusahaan sebagai dasar dalam menyusun strategi retensi pelanggan secara lebih terarah. Pelanggan dengan biaya bulanan yang tinggi dapat diberikan penawaran khusus, seperti potongan harga atau rekomendasi paket layanan yang lebih sesuai dengan kebutuhan pelanggan. Selain itu, perusahaan juga dapat menerapkan program loyalitas bagi pelanggan baru untuk meningkatkan durasi berlangganan serta mengurangi potensi churn pada pelanggan dengan masa penggunaan layanan yang masih singkat. Integrasi model prediksi churn ke dalam sistem informasi perusahaan juga dapat mendukung proses monitoring pelanggan secara lebih cepat melalui mekanisme early warning system. Dengan demikian, perusahaan dapat melakukan tindakan preventif secara lebih efektif sebelum pelanggan benar-benar berhenti menggunakan layanan.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma Random Forest Classifier mampu digunakan untuk melakukan klasifikasi customer churn pada layanan telekomunikasi dengan performa yang cukup baik. Model yang dibangun menghasilkan nilai accuracy sebesar 79,21%, precision sebesar 66,27%, recall sebesar 44,12%, dan ROC-AUC sebesar 83,94%. Hasil tersebut menunjukkan bahwa model mampu mengenali pola perilaku pelanggan berdasarkan karakteristik layanan yang digunakan. Selain itu, analisis feature importance menunjukkan bahwa variabel contract, tenure, monthly charges, dan tech support memiliki pengaruh terhadap potensi churn pelanggan, dengan jenis kontrak menjadi faktor yang paling dominan.

Implikasi penelitian ini menunjukkan bahwa pemanfaatan data mining dalam sistem informasi dapat membantu perusahaan dalam mendukung pengambilan keputusan berbasis data, khususnya pada proses identifikasi pelanggan yang berpotensi churn. Integrasi model prediksi ke dalam sistem informasi perusahaan memungkinkan proses pemantauan pelanggan dilakukan secara lebih cepat dan terstruktur sehingga perusahaan dapat menerapkan strategi retensi pelanggan secara lebih efektif. Selain itu, hasil penelitian ini juga menunjukkan bahwa pendekatan machine learning dapat dimanfaatkan sebagai bagian dari sistem pendukung keputusan untuk meningkatkan kualitas layanan dan loyalitas pelanggan.

Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan beragam agar model yang dihasilkan memiliki tingkat generalisasi yang lebih baik. Selain itu, penelitian selanjutnya dapat membandingkan algoritma Random Forest dengan metode lain seperti XGBoost atau Logistic Regression untuk mengetahui performa yang lebih optimal. Penambahan variabel lain yang relevan juga dapat

dilakukan untuk meningkatkan akurasi model. Selain itu, pengembangan sistem berbasis web atau aplikasi yang mengintegrasikan model prediksi churn dapat menjadi langkah lanjutan agar hasil penelitian lebih aplikatif. Dengan demikian, penelitian ini dapat dikembangkan lebih lanjut untuk memberikan manfaat yang lebih luas dalam dunia industri.

DAFTAR PUSTAKA

- [1] Y. Gunawan, "Analisis faktor-faktor pengaruh customer churn pada sebuah perusahaan internet provider," *Jurnal Konseling dan Pendidikan*, vol. 14, no. 1, pp. 577–590, 2026, doi: 10.1002/9781118349212.ch6.
- [2] B. Gegeleso and O. Ebiesuwa, "Comparative Analysis of Random Forest and LSTM Models for Customer Churn Prediction Based on Customer Satisfaction and Retention," *Indonesian Journal of Data and Science*, vol. 6, no. 2, pp. 313–323, 2025, doi: 10.56705/ijodas.v6i2.244.
- [3] R. T. Prasetyo, Y. Ramdhani, V. H. C. Putra, and Pratiwi, "Algoritma Random Forest Sebagai Uplift Modeling pada Prediksi Retensi Pelanggan Layanan Telekomunikasi," *Jurnal Infotech*, vol. 7, no. 2, pp. 84–88, 2025.
- [4] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," *International Journal of Intelligent Networks*, vol. 4, no. April, pp. 145–154, 2023, doi: 10.1016/j.ijin.2023.05.005.
- [5] E. Mufida, D. Andriansyah, and H. Hertiana, "Customer Churn Prediction Pada Sektor Perbankan Dengan Model Logistic Regression dan Random Forest," *Computer Science (CO-SCIENCE)*, vol. 5, no. 1, pp. 58–66, 2025, doi: 10.31294/coscience.v5i1.7576.
- [6] S. Nurhaliza, A. Harismahyanti, M. A. Fathan, M. E. Rizal, Muh. Z. Yahya, and Asfar, "Penanganan Data Churn Tidak Seimbang Menggunakan Pembobotan pada Model Supervised," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 7, no. 1, pp. 1–9, 2026.
- [7] S. D. Damanik and M. I. Jambak, "Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi," *Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1303–1309, 2023, doi: 10.30865/klik.v3i6.829.
- [8] D. Kurniasari, L. Humairoso, Warsono, and Notiragayu, "Implementation of Random Forest Method for Customer Churn Classification," *JSI : Jurnal Sistem Informasi (E-Journal)*, vol. 17, no. 1, pp. 187–202, 2025.
- [9] R. C. Admanegara and W. Handayani, "Customer Churn Analysis Using Machine Learning to Improve Customer Retention on Vissie Net," *International Journal of Scientific Research and Management (IJSRM)*, vol. 12, no. 09, pp. 7379–7387, 2024, doi: 10.18535/ijrm/v12i09.em05.
- [10] A. Handijono and Z. Suhatman, "Retensi Nasabah Bertarget : Pendekatan Berbasis RFM dan AI dalam Perbankan," *SAINTECH: Jurnal Penelitian dan Pengkajian Sains dan Teknologi*, vol. 35, no. 3, pp. 48–61, 2025.
- [11] B. Mahendra, Martanto, D. Pratama, A. Faqih, and R. Kurniawan, "Pengaruh Kualitas Data Terhadap Performa Model Machine Learning Dalam Pendekatan Data-Centric AI," *Jurnal Sistem Informasi dan Teknologi*, vol. 5, no. 1, pp. 107–113, 2021.
- [12] M. W. Nugroho, "Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. 4, pp. 1562–1571, 2025, doi: 10.35870/jtik.v9i4.4236.
- [13] D. A. Kusuma, A. R. Dewi, and A. R. Wijaya, "Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 10, no. 2, pp. 186–194, 2025, doi: 10.14421/jiska.2025.10.2.186-194.
- [14] S. Mutmainah, "Penanganan Imbalance Data Pada Klasifikasi Kemungkinan Penyakit Stroke," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 1, no. 1, pp. 10–16, 2021, doi: 10.20885/snati.v1i1.2.
- [15] I. K. M. Bili, I. W. Sudiarta, M. Y. Wungabelen, N. K. A. Rosdiana, and P. Rafiana, "Analisis dan Prediksi Customer Churn pada Platform Streaming Berbasis Langganan Menggunakan Metode Random Forest," *Jurnal Bisnis Inovatif dan Digital*, vol. 3, no. 1, pp. 1–15, 2026.