



## Strategi Retensi Nasabah Perbankan Lokal Berbasis Machine Learning: Analisis Perbandingan Algoritma Klasifikasi dan Teknik Resampling

Ravensca Matatula<sup>1</sup>, Marchello Gefan Salenus<sup>2</sup>, Marvelous Marvin Riyoly<sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, Fakultas Ilmu Komputer, Universitas Kristen Indonesia Maluku, Ambon, Indonesia

Email: <sup>1</sup>ravenscamatatula@gmail.com, <sup>2</sup>sale.marchello@gmail.com, <sup>3</sup>marvelmarvin1692@gmail.com

### Informasi Artikel

Diterima : 17-09-2025

Disetujui : 11-10-2025

Diterbitkan : 15-11-2025

### ABSTRACT

*Customer retention has become an increasingly important strategic challenge for local banking institutions amid intensifying competition and the acceleration of digital transformation, making an understanding of customer loyalty patterns essential for designing effective and data-driven retention strategies. This study aims to analyze and compare the performance of machine learning algorithms in predicting customer loyalty in a local banking context, as well as to evaluate the impact of class imbalance handling techniques on model performance. Three classification algorithms—Decision Tree, Random Forest, and Logistic Regression—are employed in this study, with methodological stages including data preprocessing, model development, and performance evaluation. To address class imbalance in customer data, three approaches are applied, namely class weight adjustment, up sampling, and down sampling. Model performance is evaluated using the F1-Score and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The results indicate that the Random Forest algorithm combined with the up sampling technique demonstrates the most consistent performance compared to the other algorithms tested, particularly in handling the minority class. The model achieves an F1-Score of 60% and an AUC-ROC value of 84%, indicating a good balance between precision and recall as well as adequate class discrimination capability. These findings suggest that ensemble-based machine learning models, supported by appropriate class imbalance handling techniques, can serve as effective decision-support tools for customer retention strategies in the context of local banking.*

**Keyword:** Loyalty Prediction, Machine Learning, Decision Tree, Random Forest, Logistic Regression.

## ABSTRAK

Retensi nasabah merupakan tantangan strategis yang semakin penting bagi perbankan lokal seiring dengan meningkatnya persaingan dan percepatan transformasi digital, sehingga pemahaman terhadap pola loyalitas nasabah menjadi kebutuhan utama dalam merancang strategi retensi yang efektif dan berbasis data. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma machine learning dalam memprediksi loyalitas nasabah perbankan lokal, sekaligus mengevaluasi pengaruh teknik penanganan ketidakseimbangan kelas terhadap performa model. Tiga algoritma klasifikasi, yaitu Decision Tree, Random Forest, dan Logistic Regression, digunakan dalam penelitian ini dengan tahapan metodologi yang meliputi preprocessing data, pemodelan, dan evaluasi kinerja. Untuk mengatasi permasalahan ketidakseimbangan kelas pada data nasabah, diterapkan tiga pendekatan, yakni class weight adjustment, up sampling, dan down sampling. Evaluasi kinerja model dilakukan menggunakan metrik F1-Score dan Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Hasil penelitian menunjukkan bahwa algoritma Random Forest dengan teknik Up Sampling memberikan kinerja yang paling konsisten dibandingkan algoritma lain yang diuji, khususnya dalam menangani kelas minoritas. Model ini mencapai nilai F1-Score sebesar 60% dan AUC-ROC sebesar 84%, yang menunjukkan keseimbangan yang baik antara precision dan recall serta kemampuan diskriminasi kelas yang memadai. Temuan ini mengindikasikan bahwa penerapan model machine learning berbasis ensemble dengan penanganan ketidakseimbangan kelas yang tepat dapat berperan sebagai alat bantu pendukung pengambilan keputusan dalam strategi retensi nasabah pada konteks perbankan lokal.

**Kata Kunci:** Prediksi Loyalitas, *Machine Learning*, *Decision Tree*, *Random Forest*, *Logistic Regression*.

## 1. PENDAHULUAN

Industri perbankan saat ini menghadapi tingkat persaingan yang semakin tinggi seiring dengan percepatan transformasi digital. Perkembangan layanan perbankan digital, seperti mobile banking dan internet banking, mendorong perubahan perilaku nasabah yang semakin dinamis serta meningkatkan risiko perpindahan nasabah (churn). Dalam konteks ini, retensi dan loyalitas nasabah menjadi isu strategis yang secara langsung memengaruhi keberlanjutan kinerja bisnis perbankan, khususnya bagi bank-bank lokal yang memiliki keterbatasan sumber daya dibandingkan dengan institusi perbankan nasional berskala besar (Azmi, 2024).

Pemanfaatan machine learning telah banyak digunakan sebagai pendekatan prediktif untuk mengidentifikasi potensi churn dan pola loyalitas nasabah berdasarkan data historis dan karakteristik individu. Berbagai algoritma klasifikasi seperti Decision Tree, Random Forest,

dan Logistic Regression terbukti efektif dalam memodelkan perilaku nasabah dan mendukung pengambilan keputusan berbasis data (Syawaludin & Hidayat, 2024; Admanegara & Handayani, 2024). Beberapa penelitian terdahulu juga menunjukkan keberhasilan pendekatan machine learning dalam berbagai konteks klasifikasi, seperti prediksi perilaku pelanggan, pengelolaan persediaan, dan klasifikasi berbasis data historis, yang memperkuat relevansi pendekatan yang digunakan dalam penelitian ini (Nasrullah, 2021; Harsiti et al., 2022; Marlina et al., 2023; Suci Amaliah et al., 2022). Namun, sebagian besar penelitian terdahulu masih mengandalkan dataset publik atau data simulasi dengan karakteristik yang relatif homogen, sehingga kurang merepresentasikan kondisi operasional dan perilaku nasabah pada konteks perbankan lokal.

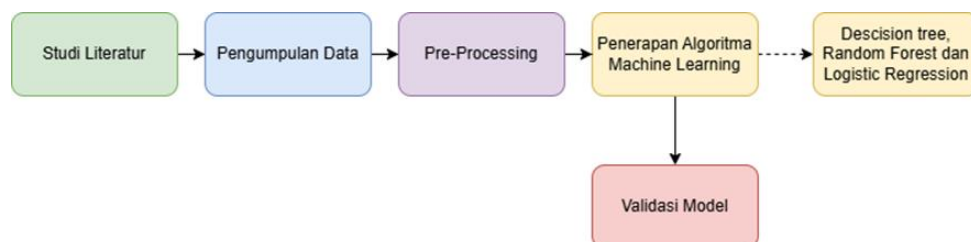
Karakteristik nasabah perbankan lokal, seperti di Kota Ambon, memiliki keunikan tersendiri yang dipengaruhi oleh faktor geografis, tingkat literasi digital, serta pola transaksi yang berbeda dibandingkan dengan data publik berskala nasional atau global. Oleh karena itu, penggunaan dataset yang melibatkan data perbankan lokal menjadi penting untuk menghasilkan model prediktif yang lebih kontekstual dan relevan secara praktis. Selain itu, penggabungan data lokal dengan dataset publik dilakukan untuk meningkatkan keberagaman sampel tanpa menghilangkan karakteristik utama nasabah lokal.

Pendekatan prediktif berbasis machine learning telah banyak digunakan dalam penelitian churn dan loyalitas nasabah, khususnya pada sektor jasa keuangan dan telekomunikasi. Studi-studi sebelumnya menunjukkan bahwa model klasifikasi dan ensemble mampu menangkap pola perilaku pelanggan secara lebih efektif dibandingkan pendekatan statistik konvensional, terutama pada data dengan karakteristik tidak seimbang (Bahnsen et al., 2015; Verbeke et al., 2012; Amin et al., 2019; Idris et al., 2019).

Dalam perspektif bisnis digital, model prediksi loyalitas nasabah tidak hanya berfungsi sebagai alat analisis teknis, tetapi juga sebagai dasar dalam pengembangan strategi retensi berbasis data, seperti personalisasi layanan digital, penguatan sistem *Customer Relationship Management* (CRM), serta perancangan intervensi preventif untuk menekan tingkat churn. Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Decision Tree, Random Forest, dan Logistic Regression, serta mengevaluasi pengaruh teknik penanganan ketidakseimbangan kelas terhadap performa model dalam mendukung strategi retensi nasabah pada perbankan lokal.

## 2. METODE

Metode dalam penelitian ini berfokus terhadap beberapa tahapan secara umum yang akan diuraikan dalam gambar berikut.



Gambar 1. Tahapan Penelitian

## 2.1. Studi Literatur

Pengumpulan informasi mengenai identifikasi dan Literatur terkait metode klasifikasi yang dapat memberikan akurasi terbaik terhadap hasil

## 2.2. Pengumpulan Data

Dataset berjumlah 10.000 baris, dikumpulkan dari beberapa bank di Kota Ambon dan dilengkapi dengan sebagian data open-source (kaggle) untuk menambah data sampel. Fitur yang digunakan: RowNumber, CustomerId, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary. Target: status Exited (churn).

## 2.3. Pre-processing

Tahapan pre-processing dilakukan untuk memastikan kualitas dan konsistensi data sebelum proses pemodelan. Proses ini meliputi penanganan missing value dengan pendekatan imputasi yang sesuai, transformasi data kategorikal menggunakan teknik encoding, serta normalisasi fitur numerik untuk mengurangi bias skala antar variabel.

Selain itu, seleksi fitur (feature selection) dilakukan untuk menghilangkan atribut yang tidak memiliki kontribusi signifikan terhadap proses klasifikasi. Fitur-fitur identitas seperti RowNumber, CustomerId, dan Surname dihapus karena tidak memiliki hubungan kausal dengan loyalitas nasabah dan berpotensi menimbulkan noise pada model. Evaluasi signifikansi fitur dilakukan berdasarkan analisis korelasi dan kontribusi fitur terhadap performa model pada tahap awal pelatihan. Pendekatan ini bertujuan untuk meningkatkan efisiensi komputasi dan mengurangi risiko overfitting, sekaligus mempertahankan fitur-fitur yang relevan secara bisnis dan prediktif.

## 2.4. Penerapan Algoritma dan Hyperparameter Tuning

Pada tahap pemodelan, dataset yang telah melalui proses pre-processing dibagi ke dalam data pelatihan (*training*) dan data pengujian (*testing*) dengan proporsi yang ditetapkan untuk memastikan evaluasi model yang objektif. Penelitian ini menerapkan tiga algoritma klasifikasi, yaitu *Decision Tree*, *Random Forest*, dan *Logistic Regression*, yang dipilih karena kemampuannya dalam menangani data tabular serta interpretabilitas hasil yang relevan untuk konteks perbankan.

Proses hyperparameter tuning dilakukan untuk memperoleh konfigurasi model yang optimal. Pada algoritma Random Forest, parameter *n\_estimators* dievaluasi melalui serangkaian eksperimen dengan variasi jumlah pohon untuk mengamati pengaruhnya terhadap nilai F1-Score dan AUC-ROC. Pendekatan pencarian parameter dilakukan secara eksperimental terkontrol (manual tuning) dengan mempertimbangkan keseimbangan antara performa model dan efisiensi komputasi, hingga diperoleh konfigurasi terbaik pada *n\_estimators* = 90. Pendekatan ini dipilih karena memberikan performa yang stabil dan konsisten pada data yang digunakan.

## 2.5. Validasi Model dan Etika Data

Validasi model dilakukan untuk mengukur kinerja prediksi pada kondisi data yang tidak seimbang menggunakan tiga teknik, yaitu class weight adjustment, up sampling, dan down sampling. Evaluasi kinerja model dilakukan menggunakan metrik F1-Score dan *Area Under the Receiver Operating Characteristic Curve* (AUC-ROC), yang dianggap lebih representatif dibandingkan akurasi dalam kasus ketidakseimbangan kelas.

Penelitian ini menggunakan data perbankan lokal yang telah melalui proses anonimisasi untuk melindungi privasi nasabah. Seluruh informasi identitas pribadi dihilangkan atau disamarkan sebelum proses analisis, sehingga data yang digunakan tidak dapat mengidentifikasi individu secara langsung. Dengan demikian, penelitian ini memenuhi prinsip etika penelitian data dan menjaga kerahasiaan informasi nasabah.

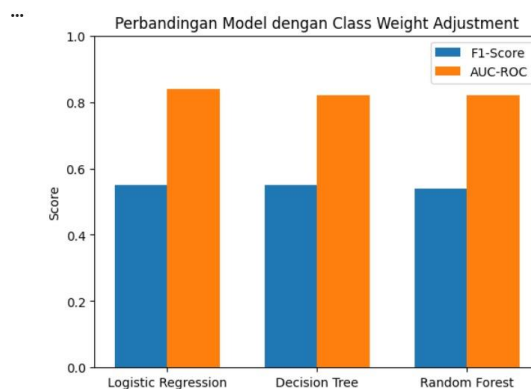
Permasalahan ketidakseimbangan kelas merupakan tantangan umum dalam pemodelan churn nasabah, di mana jumlah pelanggan loyal biasanya jauh lebih besar dibandingkan pelanggan yang berhenti menggunakan layanan. Kondisi ini dapat menyebabkan bias model terhadap kelas mayoritas jika tidak ditangani secara tepat (He & Garcia, 2009). Oleh karena itu, penelitian ini menerapkan teknik penanganan data tidak seimbang untuk meningkatkan sensitivitas model terhadap kelas minoritas.

Evaluasi performa model menggunakan metrik F1-Score dan AUC-ROC dipilih karena metrik tersebut lebih representatif dibandingkan akurasi dalam konteks klasifikasi dengan distribusi kelas yang tidak seimbang (Fawcett, 2006).

## 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil evaluasi kinerja model prediksi loyalitas nasabah menggunakan tiga algoritma klasifikasi, yaitu Logistic Regression, Decision Tree, dan Random Forest, dengan menerapkan tiga teknik penanganan ketidakseimbangan kelas: class weight adjustment, up sampling, dan down sampling. Pembahasan tidak hanya difokuskan pada perbandingan nilai metrik evaluasi, tetapi juga pada analisis penyebab perbedaan performa antar teknik serta implikasinya terhadap strategi retensi nasabah pada perbankan lokal.

### 3.1 Analisis Class weight adjustment

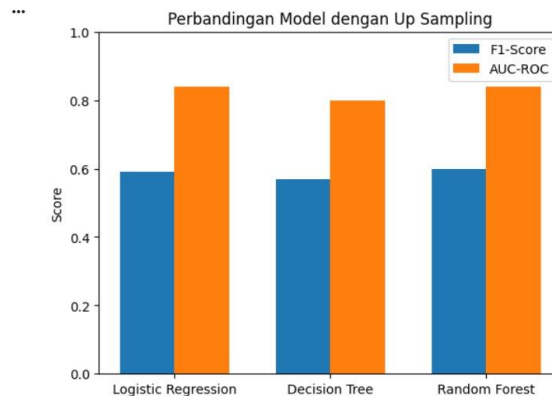


Gambar 2. Perbandingan Model dengan Class Weight Adjustment

Pada tahap ini, kinerja algoritma Logistic Regression, Decision Tree, dan Random Forest dievaluasi dengan menerapkan teknik class weight adjustment sebagai upaya menangani ketidakseimbangan kelas pada data nasabah. Pendekatan ini digunakan agar model dapat memberikan perhatian yang lebih proporsional terhadap kelas minoritas, tanpa perlu mengubah distribusi data secara langsung.

Hasil pengujian menunjukkan bahwa penerapan teknik class weight adjustment menghasilkan performa yang relatif stabil pada ketiga algoritma, namun belum mampu meningkatkan nilai F1-Score secara signifikan. Hal ini mengindikasikan bahwa penyesuaian bobot kelas saja belum cukup untuk menangkap kompleksitas distribusi data nasabah yang tidak seimbang. Pada konteks data perbankan lokal, karakteristik nasabah yang berpotensi churn cenderung memiliki pola yang lebih beragam, sehingga pendekatan berbasis pembobotan kelas masih menyisakan keterbatasan dalam mengenali minoritas kelas secara optimal.

### 3.2 Analisis Up Sampling

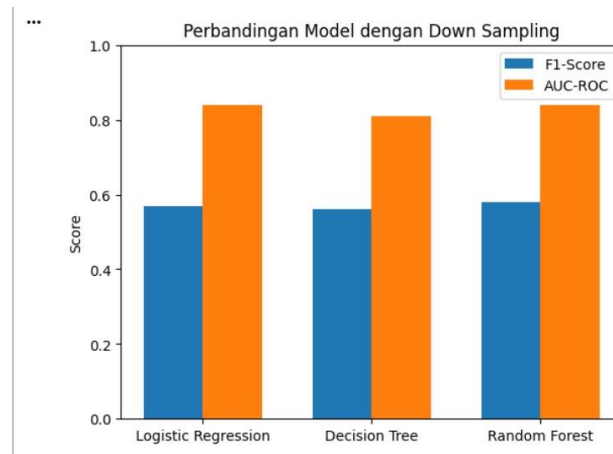


Gambar 3. Perbandingan Model dengan Up Sampling

Penerapan teknik up sampling menunjukkan peningkatan performa yang lebih konsisten dibandingkan dengan class weight adjustment dan down sampling, khususnya pada algoritma Random Forest. Peningkatan ini disebabkan oleh kemampuan up sampling dalam memperkaya representasi kelas minoritas pada data pelatihan, sehingga model dapat mempelajari pola perilaku nasabah yang berpotensi churn dengan lebih baik. Dalam konteks perbankan lokal, di mana jumlah nasabah loyal umumnya lebih dominan dibandingkan nasabah yang berhenti menggunakan layanan, pendekatan ini membantu mengurangi bias model terhadap kelas mayoritas dan meningkatkan sensitivitas prediksi terhadap nasabah berisiko churn.

Temuan penelitian ini sejalan dengan hasil penelitian sebelumnya yang menunjukkan bahwa algoritma Random Forest memiliki keunggulan dalam menangani kompleksitas dan non-linearitas data perilaku nasabah, khususnya pada konteks prediksi churn dan loyalitas pelanggan (Breiman, 2001; Guliyev & Tatoğlu, 2021; Ehsani, 2022). Keunggulan ini diperkuat ketika dikombinasikan dengan teknik penanganan data tidak seimbang yang mampu meningkatkan representasi kelas minoritas.

### 3.3 Analisis Down Sampling



Gambar 4. Perbandingan Moden dengan Down Sampling

Selain up sampling, penelitian ini juga mengevaluasi kinerja model dengan menerapkan teknik down sampling. Pendekatan ini dilakukan dengan mengurangi jumlah data pada kelas mayoritas sehingga distribusi antar kelas menjadi lebih seimbang.

Sementara itu, teknik down sampling cenderung menghasilkan performa yang lebih rendah dibandingkan up sampling. Pengurangan jumlah data pada kelas mayoritas menyebabkan hilangnya sebagian informasi penting yang merepresentasikan variasi perilaku nasabah loyal. Dampak ini menjadi lebih signifikan pada dataset berukuran terbatas, sehingga kemampuan model dalam melakukan generalisasi menjadi berkurang.

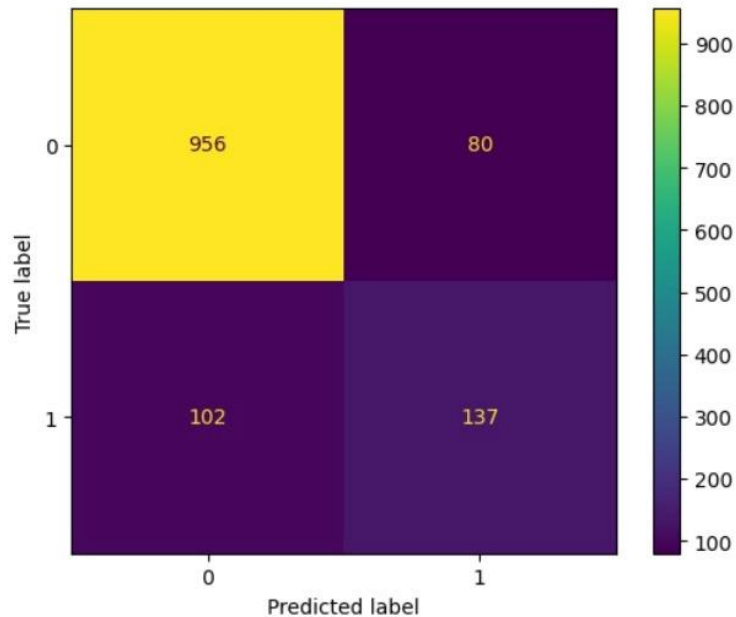
Sintesis hasil

Table 1. Hasil Evaluasi Kinerja Algoritma

<i>Parameter</i>	<i>Logistic Regression</i>		<i>Decision Tree</i>		<i>Random Forest</i>	
	<i>F1 Score</i>	<i>AUC-ROC</i>	<i>F1 Score</i>	<i>AUC-ROC</i>	<i>F1 Score</i>	<i>AUC-ROC</i>
<i>Class Weight Adjustment</i>	55%	84%	55%	82%	54%	82%
<i>Up Sampling</i>	59%	84%	57%	80%	60%	84%
<i>Down sampling</i>	57%	84%	56%	81%	58%	84%

Secara keseluruhan, hasil perbandingan yang disajikan pada Tabel 1 menunjukkan bahwa kombinasi algoritma Random Forest dengan teknik up sampling memberikan performa terbaik berdasarkan nilai F1-Score dan AUC-ROC. Hal ini menguatkan temuan bahwa model ensemble yang dikombinasikan dengan teknik penanganan ketidakseimbangan kelas yang tepat lebih efektif dalam menangani kompleksitas data perilaku nasabah.

### 3.4 Evaluasi Model Terbaik Menggunakan Confusion Matrix



Gambar 5. Confusion Matrix Model Random Forest dengan Teknik Up Sampling

Pada tahap model testing, dilakukan evaluasi lebih lanjut terhadap model terbaik yang diperoleh dalam penelitian ini, yaitu Random Forest dengan teknik Up Sampling dan parameter  $n\_estimators = 90$ . Evaluasi dilakukan menggunakan confusion matrix untuk mengetahui secara rinci kemampuan model dalam mengklasifikasikan nasabah loyal (tidak churn) dan nasabah yang berpotensi churn.

Berdasarkan hasil confusion matrix yang ditunjukkan pada Gambar 5, model Random Forest berhasil mengklasifikasikan 956 data nasabah tidak churn secara benar (True Negative). Nilai ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali nasabah loyal, sehingga risiko kesalahan dalam mempertahankan nasabah yang sebenarnya loyal dapat diminimalkan.

Namun demikian, terdapat 80 data nasabah tidak churn yang salah diklasifikasikan sebagai churn (False Positive). Kesalahan ini mengindikasikan bahwa model cenderung memberikan peringatan churn pada sebagian kecil nasabah yang sebenarnya masih loyal. Dalam konteks bisnis perbankan, kondisi ini masih dapat ditoleransi karena bank dapat memberikan perhatian tambahan atau penawaran retensi tanpa risiko kehilangan nasabah.



Pada sisi kelas churn, model berhasil mengidentifikasi 137 nasabah churn secara benar (True Positive). Hal ini menunjukkan bahwa model mampu mendeteksi nasabah yang berpotensi meninggalkan layanan bank dengan cukup baik. Akan tetapi, masih terdapat 102 nasabah churn yang tidak terdeteksi dan diklasifikasikan sebagai tidak churn (False Negative). Kesalahan ini menjadi perhatian penting karena nasabah yang berpotensi churn tidak mendapatkan perlakuan retensi yang diperlukan.

Secara keseluruhan, hasil confusion matrix ini menunjukkan bahwa model Random Forest dengan teknik Up Sampling memiliki keseimbangan yang baik antara kemampuan mendeteksi churn dan mempertahankan akurasi pada kelas mayoritas. Hal ini selaras dengan nilai F1-Score sebesar 60%, yang menandakan bahwa model mampu menjaga keseimbangan antara precision dan recall, serta nilai AUC-ROC sebesar 84%, yang menunjukkan kemampuan diskriminasi kelas yang sangat baik.

Dengan demikian, hasil model testing ini memperkuat kesimpulan bahwa Random Forest dengan teknik Up Sampling merupakan model yang paling optimal dalam penelitian ini dan layak digunakan sebagai alat bantu prediksi loyalitas nasabah. Model ini dapat membantu pihak perbankan dalam mengidentifikasi nasabah yang berisiko churn secara lebih dini sehingga strategi retensi dapat dilakukan secara lebih efektif dan tepat sasaran.

### **3.5 Implikasi Manajerial dalam Konteks Bisnis Digital**

Dari perspektif manajerial, model prediksi loyalitas nasabah yang dihasilkan dalam penelitian ini dapat dimanfaatkan sebagai alat pendukung pengambilan keputusan dalam strategi retensi berbasis data. Bank dapat mengintegrasikan model ini ke dalam sistem Customer Relationship Management (CRM) untuk mengidentifikasi nasabah yang memiliki probabilitas tinggi untuk churn. Informasi tersebut memungkinkan manajemen bank untuk merancang intervensi preventif, seperti personalisasi penawaran produk digital, peningkatan kualitas layanan melalui kanal digital, serta kampanye pemasaran yang lebih terarah. Dengan demikian, pemanfaatan model prediktif tidak hanya meningkatkan efisiensi operasional, tetapi juga memperkuat hubungan jangka panjang antara bank dan nasabah.

Dalam konteks bisnis digital, pemanfaatan model prediksi churn memungkinkan organisasi untuk mengimplementasikan strategi retensi yang lebih proaktif dan terarah, seperti personalisasi layanan, optimalisasi kampanye pemasaran digital, serta penguatan sistem Customer Relationship Management (CRM). Pendekatan ini telah banyak direkomendasikan dalam penelitian sebelumnya sebagai bagian dari pengambilan keputusan berbasis data pada sektor jasa (Moro et al., 2015; Ren, 2024).

## **4. PENUTUP**

### **4.1 Kesimpulan**

Berdasarkan hasil pengujian dan evaluasi yang telah dilakukan, penelitian ini menunjukkan bahwa algoritma Random Forest dengan teknik Up Sampling memberikan kinerja yang paling konsisten dibandingkan Logistic Regression dan Decision Tree dalam memprediksi loyalitas nasabah pada konteks perbankan lokal. Penerapan teknik Up Sampling

terbukti membantu meningkatkan kemampuan model dalam mengenali kelas minoritas (nasabah churn) tanpa mengorbankan performa klasifikasi pada kelas mayoritas secara signifikan.

Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa model Random Forest mampu mengklasifikasikan sebagian besar nasabah loyal dengan baik, serta mendeteksi nasabah yang berpotensi churn secara cukup efektif, meskipun masih terdapat sejumlah kesalahan klasifikasi pada kelas churn yang perlu menjadi perhatian. Nilai F1-Score sebesar 60% menunjukkan keseimbangan yang moderat antara precision dan recall, sementara nilai AUC-ROC sebesar 84% mengindikasikan kemampuan diskriminasi kelas yang baik.

Dengan demikian, Random Forest dengan teknik Up Sampling dapat digunakan sebagai alat bantu pendukung pengambilan keputusan dalam strategi retensi nasabah perbankan lokal, khususnya untuk mengidentifikasi risiko churn secara lebih dini. Penelitian ini juga menegaskan pentingnya pemilihan teknik penanganan ketidakseimbangan kelas dalam pengembangan model prediksi loyalitas nasabah. Ke depan, penelitian lanjutan dapat dilakukan dengan menambahkan variabel perilaku yang lebih beragam serta mengeksplorasi pendekatan ensemble atau metode optimasi lainnya untuk meningkatkan performa model, khususnya pada kelas churn.

## 4.2 Saran

Penelitian selanjutnya disarankan untuk mengeksplorasi penggunaan dataset dengan skala dan karakteristik yang lebih beragam, serta melakukan pengujian terhadap variasi rasio kelas yang berbeda untuk meningkatkan generalisasi model. Selain itu, penerapan metode hyperparameter tuning yang lebih sistematis, seperti Grid Search atau Random Search, serta eksplorasi algoritma yang lebih kompleks atau berbasis deep learning dapat dipertimbangkan untuk meningkatkan performa prediksi. Integrasi model prediktif dengan sistem operasional perbankan digital, seperti CRM atau platform pemasaran digital, juga menjadi arah penelitian lanjutan yang relevan untuk memperkuat implikasi praktis penelitian ini.

## DAFTAR PUSTAKA

- Admanegara, R. C., & Handayani, W. (2024). *Customer Churn Analysis Using Machine Learning to Improve Customer Retention on Vissie Net*. 12(09), 7379–7387. <https://doi.org/10.18535/ijssrm/v12i09.em05>
- 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>.
- Guliyev, H., & Yerdelen Tatoğlu, F. (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning models. *Journal of Applied Microeconometrics*, 1(2), 85–99. <https://doi.org/10.53753/jame.1.2.03>

Strategi Retensi Nasabah Perbankan Lokal Berbasis Machine Learning: Analisis Perbandingan Algoritma Klasifikasi dan Teknik Resampling

- Harsiti, Muttaqin, Z., & Srihartini, E. (2022). Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet. *JSiI (Jurnal Sistem Informasi)*, 9(1), 12–16. <https://doi.org/10.30656/jsii.v9i1.4426>
- Marlina Haiza, Elmayati, Zulus Antoni, & Wijaya Harma Oktafia Lingga. (2023). Penerapan Algoritma Random Forest Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo. *Penerapan Kecerdasan Buatan*, 4(2), 138–143.
- Nasrullah, A. H. (2021). Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris. *Jurnal Ilmiah Ilmu Komputer*, 7(2), 45–51. <https://doi.org/10.35329/jiik.v7i2.203>
- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konjiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127. <https://doi.org/10.35580/variansiunm31>.
- Syawaludin, M. A., & Hidayat, R. (2024). *Prediksi Churn Pelanggan*
- Fawcett, T. (2006). *An Introduction to ROC Analysis*. *Pattern Recognition Letters*, 27(8). <https://doi.org/10.1016/j.patrec.2005.10.010>
- Breiman, L. (2001). *Random Forests*. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Moro, S., Cortez, P. & Rita, P. (2015). *A Data-driven Approach to Predicting Bank Telemarketing Success*. *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2014.03.001>
- Chawla, N. et al. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. *JAIR*, 16, 321–357. <https://doi.org/10.1613/jair.953>.
- He, H. & Garcia, E. (2009). *Learning from Imbalanced Data*. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2008.239>
- Guliyev, H., Tatoğlu, F. Y. (2021). *Customer churn analysis in banking sector: Evidence from explainable machine learning models*. *JAME E-ISSN:2791-7401 Volume: 1, Issue 2, 2021*
- Bahnsen, A.C., Aouada, D., Ottersten. B. (2015). *Anovel cost-sensitive framework for customer churn predictive modeling*. *DecisionAnalytics* (2015) 2:5 a SpringerOpen Journal. DOI10.1186/s40165-015-0014-6
- Ren, H. (2024). *Machine Learning-Based Prediction of Customer Churn Risk in E-commerce*. *Proceedings of the 3rd International Conference on Financial Technology and Business Analysis* DOI: 10.54254/2754-1169/153/2024.19473.

Strategi Retensi Nasabah Perbankan Lokal Berbasis Machine Learning: Analisis Perbandingan Algoritma Klasifikasi dan Teknik Resampling

- Ehsani, F. (2022). Customer churn prediction from Internet banking transactions data using an ensemble metaclassifier algorithm. Research Square preprint. <https://doi.org/10.21203/rs.3.rs-1630808/v1>
- Moro, S., Cortez, P. & Rita, P. (2015). *A Data-driven Approach to Predicting Bank Telemarketing Success*. Decision Support Systems. <https://doi.org/10.1016/j.dss.2014.03.001>.
- Verbeke, W., Martens, D., & Baesens, B. (2012). *Social network analysis for customer churn prediction*. *Applied Soft Computing* 14:431–446 <https://doi.org/10.1016/j.asoc.2013.09.017>
- Idris, A., Khan, A., & Lee, Y. S. (2019). *Intelligent churn prediction model using boosted trees*. IEEE Conference on Computer Applications (ICCA). DOI:10.1109/ICCA51723.2023.10181933
- Amin, A., Al-Obeidat, F., Shah, B., et al. (2019). *Customer churn prediction in telecom using data certainty*. <https://doi.org/10.1016/j.jbusres.2018.03.003>.