

Prediksi Customer Churn Berbasis Attention Gated Residual Deep Neural Network Dengan Optimasi Optuna Pada Tiga Dataset Transformasi Digital

Diajeng Tyas Purwa Hapsari¹, Gunawan², Nur Tulus Ujianto³, Gunawan Raharjo⁴, Didiek Trisatya⁵

^{*1,2,3,4,5}Sistem Informasi, Universitas Pancasakti Tegal, Tegal

e-mail: ^{*1}diajengtyaspurwa@upstegal.ac.id, ²gunawan.gayo@upstegal.ac.id, ³nurtulus@upstegal.ac.id, ⁴gunawanraharjo@upstegal.ac.id, ⁵didiektrisatya@upstegal.ac.id

Abstrak

Transformasi digital pada sektor jasa seperti telekomunikasi, perbankan, dan perdagangan elektronik menuntut sistem informasi yang mampu mengidentifikasi pelanggan berisiko churn secara akurat dan tepat waktu. Penelitian ini mengusulkan model deep learning Attention-Gated Residual Deep Neural Network (AGR-DNN) untuk prediksi customer churn, dengan hyperparameter yang dioptimasi menggunakan Optuna berbasis Tree-structured Parzen Estimator. Model diuji pada tiga dataset sekunder publik dari domain berbeda, yaitu Telco Customer Churn, Bank Customer Churn, dan E-Commerce Customer Churn, untuk memperkuat generalisasi temuan dan memperkaya kajian state of the art. Studi ablasi dilakukan dengan menghilangkan komponen attention gate, koneksi residual, teknik SMOTE, dan optimasi Optuna secara bergantian pada ketiga dataset. Hasil pengujian menunjukkan model AGR-DNN penuh mencapai rata-rata F1-score 0,717, AUC 0,900, dan recall 0,824, mengungguli baseline Random Forest dan Logistic Regression terutama pada recall yang krusial bagi strategi retensi pelanggan. SMOTE terbukti sebagai komponen dengan kontribusi konsisten terbesar terhadap kenaikan recall pada ketiga dataset (4,9-30,6 poin persentase), sementara optimasi Optuna meningkatkan F1-score hingga 3,7 poin persentase dibandingkan hyperparameter default. Penelitian ini berkontribusi pada pengembangan sistem informasi prediktif berbasis data yang mendukung transformasi digital pada organisasi jasa.

Kata Kunci: customer churn; transformasi digital; deep learning; attention mechanism; Optuna

Abstract

Digital transformation in service sectors such as telecommunications, banking, and e-commerce requires information systems capable of accurately and promptly identifying customers at risk of churning. This study proposes a deep learning model, the Attention-Gated Residual Deep Neural Network (AGR-DNN), for customer churn prediction, with hyperparameters optimized using Optuna based on the Tree-structured Parzen Estimator. The model was evaluated on three public secondary datasets from different domains: Telco Customer Churn, Bank Customer Churn, and E-Commerce Customer Churn, to strengthen the generalizability of the findings and enrich the state-of-the-art comparison. An ablation study removed the attention gate, residual connections, SMOTE oversampling, and Optuna tuning in turn across the three datasets. Results show the full AGR-DNN model achieves an average F1-score of 0.717, AUC of 0.900, and recall of 0.824, outperforming Random Forest and Logistic Regression baselines, particularly on recall, which is critical for customer retention strategy. SMOTE consistently contributed the largest recall improvement across all three datasets (4.9-30.6 percentage points), while Optuna-based optimization improved F1-score by up to 3.7 percentage points over default hyperparameters. This research contributes to predictive, data-driven information systems supporting digital transformation in service organizations.

Keywords: customer churn; digital transformation; deep learning; attention mechanism; Optuna

I. PENDAHULUAN

Percepatan transformasi digital di sektor jasa mencakup telekomunikasi, perbankan, dan perdagangan elektronik (e-commerce) telah mengubah pola perilaku pelanggan secara mendasar. Kemudahan berpindah penyedia layanan (*switching cost* yang rendah) menjadikan *customer churn* salah satu ancaman utama bagi keberlangsungan bisnis digital. Laporan McKinsey (2023) menyebutkan bahwa biaya mempertahankan satu pelanggan lama rata-rata lima kali lebih rendah dibandingkan biaya akuisisi pelanggan baru, sehingga kemampuan mengidentifikasi pelanggan berisiko churn secara dini bernilai strategis sangat tinggi.

Sistem Informasi (SI) berperan krusial sebagai infrastruktur pengambilan keputusan berbasis data pada organisasi yang bertransformasi digital. Integrasi kecerdasan buatan ke dalam SI khususnya *deep learning* membuka peluang baru dalam prediksi perilaku pelanggan dari data tabular multidimensi yang umumnya tersedia pada *Customer Relationship Management* (CRM). Namun, penerapan *deep learning* pada data tabular masih menghadapi tantangan signifikan: ketidakseimbangan kelas (*imbalanced class*), pemilihan hyperparameter yang kritis terhadap kinerja model, serta kesulitan menentukan arsitektur yang mampu menangkap interaksi fitur non-linear secara efisien.

Berbagai penelitian telah mengusulkan pendekatan *machine learning* klasik seperti Random Forest (Geiler et al., 2022) dan Gradient Boosting (Wagh et al., 2024) untuk prediksi churn, namun metode-metode tersebut tidak secara inheren mengeksplorasi hierarki fitur dan membutuhkan *feature engineering* manual yang intensif. Di sisi lain, *deep learning* berbasis jaringan saraf dalam seperti TabNet dan DeepFM (Borisov et al., 2024; Shwartz-Ziv & Armon, 2022) menunjukkan hasil menjanjikan, tetapi umumnya diuji pada dataset tunggal sehingga generalisasi lintas domain masih perlu diteliti lebih lanjut. Studi ablasi multi-dataset yang sistematis pun masih langka dalam literatur prediksi churn berbasis *deep learning*.

Penelitian ini mengajukan tiga pertanyaan penelitian utama: (RQ1) *Apakah arsitektur AGR-DNN yang menggabungkan attention gate, residual connection, dan optimasi Optuna mampu*

mengungguli model baseline konvensional pada dataset churn dari domain berbeda? (RQ2) *Sejauh mana kontribusi masing-masing komponen attention gate, residual connection, SMOTE, dan optimasi Optuna terhadap kinerja model secara individual?* (RQ3) *Apakah kontribusi komponen tersebut konsisten di ketiga domain dataset (telekomunikasi, perbankan, dan e-commerce)?*

Untuk menjawab pertanyaan tersebut, penelitian ini memberikan kontribusi berikut: (1) Usulan arsitektur *Attention-Gated Residual Deep Neural Network* (AGR-DNN) yang mengintegrasikan *feature attention gate* (mekanisme *squeeze-and-excite* untuk data tabular), *residual connection* pada setiap blok *dense*, dan *entity embedding* untuk variabel kategoris; (2) Optimasi hyperparameter sistematis menggunakan Optuna dengan *Tree-structured Parzen Estimator* (TPE); (3) Studi ablasi terstruktur pada tiga dataset sekunder publik dari domain berbeda (Telco Customer Churn (IBM, 2023), Bank Customer Churn (Kaggle, 2023), dan E-Commerce Customer Churn (Verma, 2023)) yang memungkinkan analisis komponen secara lintas domain; (4) Pembahasan implikasi hasil terhadap pengembangan SI prediktif dalam konteks transformasi digital organisasi jasa.

II. METODE PENELITIAN

Dataset Penelitian

Penelitian ini menggunakan tiga dataset sekunder yang tersedia secara publik, yang mewakili tiga domain berbeda dalam konteks transformasi digital jasa. Penggunaan dataset publik memungkinkan reproduktibilitas penelitian dan perbandingan langsung dengan hasil dari literatur terdahulu. Karakteristik ketiga dataset dirangkum pada Tabel 2.

Tabel 2. Karakteristik Dataset Penelitian

Atribut	Telco Customer Churn	Bank Customer Churn	E-Commerce Churn
Sumber	IBM Watson Analytics (IBM, 2023)	Kaggle/UCI (Kaggle, 2023)	Kaggle (Verma, 2023)
Jumlah sampel	7.043	10.000	5.630

Fitur numerik	3	6	11
Fitur kategoris	16	4	7
Proporsi kelas positif	26,5%	20,4%	16,8%
Nilai hilang	Tidak ada	Tidak ada	1.856 (numerik)
Domain	Telekomunikasi	Perbankan	E-Commerce

Dataset *Telco Customer Churn* (IBM, 2023) dirilis oleh IBM Watson dan memuat informasi layanan berlangganan dari 7.043 pelanggan telekomunikasi. Dataset *Bank Customer Churn* (Kaggle, 2023) berisi catatan 10.000 nasabah bank di tiga negara (Prancis, Jerman, Spanyol) dengan berbagai atribut keuangan. Dataset *E-Commerce Customer Churn* (Verma, 2023) mencakup 5.630 pelanggan platform e-commerce dengan atribut perilaku belanja daring.

Pra-Pemrosesan Data

Setiap dataset menjalani tahapan pra-pemrosesan yang seragam: (1) Penanganan nilai hilang dataset e-commerce mengandung 1.856 nilai hilang pada kolom numerik yang diimputasi dengan nilai *median* kolom bersangkutan; (2) *Label Encoding* untuk variabel kategoris setiap nilai unik dipetakan ke bilangan bulat sebagai input *entity embedding*; (3) Standardisasi fitur numerik menggunakan *StandardScaler* (*zero mean, unit variance*) yang di-fit hanya pada data latih untuk mencegah kebocoran data (*data leakage*); (4) Pembagian dataset dengan rasio 70:15:15 secara stratifikasi (*train:val:test*) untuk mempertahankan proporsi kelas minoritas di setiap partisi; (5) SMOTE diterapkan hanya pada data latih menggunakan $k=5$ *nearest neighbors* untuk menghasilkan sampel sintetis kelas minoritas.

Arsitektur AGR-DNN

Arsitektur *Attention-Gated Residual Deep Neural Network* (AGR-DNN) dirancang dengan tiga komponen utama sebagaimana diilustrasikan pada Gambar 1.

Lapisan Entity Embedding.

Setiap variabel kategoris direpresentasikan sebagai vektor padat berdimensi $\min(d_{emb}, \max(2, \lfloor C/2 \rfloor + 1))$ di mana C adalah jumlah kategori unik dan

d_{emb} adalah dimensi *embedding* yang dapat dikonfigurasi. Metode ini banyak digunakan pada pemodelan tabular modern (Borisov et al., 2024; Gorishniy et al., 2022) dan terbukti menangkap relasi ordinal laten antar kategori lebih efektif dibandingkan *one-hot encoding*.

Feature Attention Gate.

Setelah konkatenasi *embedding* dan fitur numerik ternormalisasi, diterapkan mekanisme *squeeze-and-excite* tabular:

$$W_{attn} = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot x)) \quad (1)$$

$$\tilde{x} = x \odot W_{attn} \quad (2)$$

di mana $W_1 \in \mathbb{R}^{(d_{hidden} \times d_{in})}$, $W_2 \in \mathbb{R}^{(d_{in} \times d_{hidden})}$, dan σ adalah fungsi sigmoid. Perkalian element-wise ($\odot$) menghasilkan representasi fitur terbobot yang secara adaptif menekan fitur tidak relevan.

Blok Residual Dense.

Representasi terproyeksi dilewatkan melalui N blok residual berturut-turut, masing-masing terdiri atas:

$$x_{l+1} = x_l + \text{Dropout}(\text{ReLU}(\text{BN}(Wx_l + b))) \quad (3)$$

di mana BN adalah *Batch Normalization* (Huang et al., 2023), sementara Dropout berperan sebagai regularisasi (Salehin & Kang, 2023). Koneksi residual x_l memungkinkan gradien mengalir langsung ke lapisan sebelumnya, menstabilkan pelatihan dan memungkinkan jaringan yang lebih dalam.

Keluaran blok terakhir dihubungkan ke satu neuron output dengan aktivasi sigmoid menggunakan *Binary Cross-Entropy* sebagai fungsi kerugian. Seluruh komponen dapat diaktifkan atau dinonaktifkan secara independen untuk keperluan studi ablasi.

Optimasi Hyperparameter dengan Optuna

Optuna (Srinivas & Katarya, 2022) digunakan untuk mencari ruang hyperparameter berikut: dimensi *hidden* {32, 64, 96, 128}, jumlah blok residual {1–4}, tingkat *dropout* [0.1, 0.5], laju pembelajaran $[10^{-4}, 10^{-2}]$ (log-uniform), *weight decay* $[10^{-6}, 10^{-3}]$ (log-uniform), ukuran *batch* {32, 64, 128}, dan dimensi *embedding* {4, 8, 16}. Setiap *trial* Optuna melatih model hingga 50 *epoch* dengan

early stopping patience = 6, dan memaksimalkan AUC pada data validasi. Algoritma TPE digunakan dengan seed = 42 untuk reproduksibilitas.

Desain Studi Ablasi

Untuk menjawab RQ2 dan RQ3, lima varian model dilatih pada setiap dataset menggunakan hyperparameter terbaik hasil Optuna kecuali pada varian ablasi Optuna:

Tabel 3. Desain Studi Ablasi Model AGR-DNN

Kode Varian	Attention Gate	Residual Connection	SMOTE	Optuna (HP tuning)
Full Model (Proposed)	✓	✓	✓	✓
No-Attention	✗	✓	✓	✓
No-Residual	✓	✗	✓	✓
No-SMOTE	✓	✓	✗	✓
No-Optuna (default HP)	✓	✓	✓	✗

Setiap varian dievaluasi menggunakan partisi uji yang identik untuk memastikan perbandingan yang adil. Metrik evaluasi mencakup Akurasi, Presisi, Recall, F1-score, dan AUC-ROC (Chicco & Jurman, 2023). Recall mendapat perhatian khusus karena dalam konteks retensi pelanggan, biaya false negative (pelanggan churn yang tidak terdeteksi) umumnya lebih besar daripada biaya false positive.

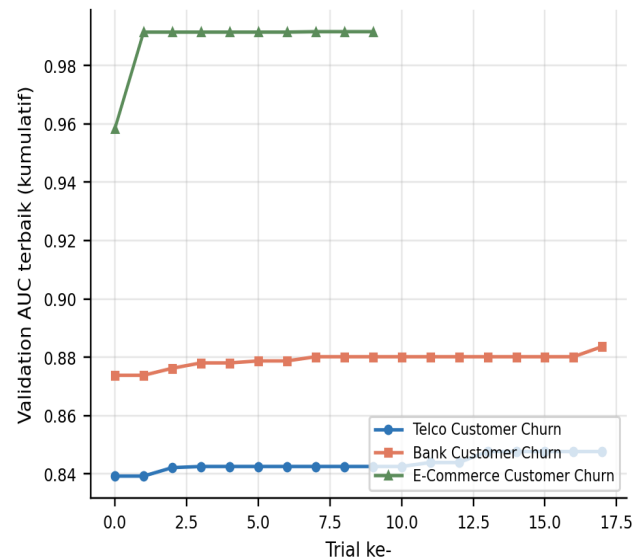
III. HASIL DAN PEMBAHASAN

Hasil Optimasi Optuna

Gambar 2 menampilkan riwayat optimasi kumulatif AUC validasi terbaik selama proses pencarian Optuna pada ketiga dataset. Dataset E-Commerce Churn menunjukkan konvergensi tercepat (trial ke-2 mencapai AUC 0.9916), sedangkan Telco Churn memerlukan lebih banyak trial hingga konvergensi (AUC akhir 0.8477). Hal ini berkorelasi dengan tingkat ketidakseimbangan kelas dataset e-commerce memiliki proporsi kelas positif terkecil (16,8%), namun fitur perilaku belanjanya sangat diskriminatif.

Hyperparameter terbaik yang ditemukan Optuna bervariasi antar dataset: Telco (hidden=32, blocks=3, dropout=0.45, lr=0.00050), Bank (hidden=96, blocks=1, dropout=0.35, lr=0.00475), E-Commerce (hidden=128, blocks=2, dropout=0.30, lr=0.00653). Keragaman konfigurasi optimal ini mengkonfirmasi

bahwa tuning hyperparameter per-dataset adalah keputusan yang tepat dan tidak dapat digantikan dengan satu konfigurasi global.



Gambar 1. Riwayat Konvergensi Optimasi Optuna pada Tiga Dataset

Kinerja Model pada Data Uji

Tabel 4, 5, dan 6 menyajikan hasil evaluasi seluruh varian model dan baseline pada partisi uji setiap dataset. Gambar 3 menyajikan perbandingan F1-score secara visual untuk memudahkan perbandingan lintas varian dan dataset.

Tabel 4. Hasil Evaluasi Model pada Dataset Telco Customer Churn

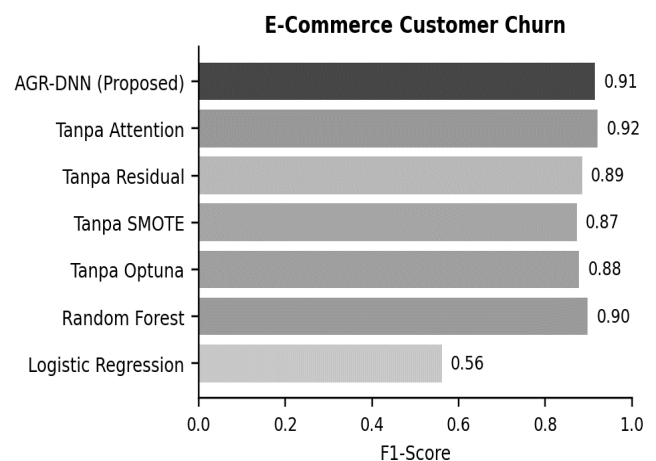
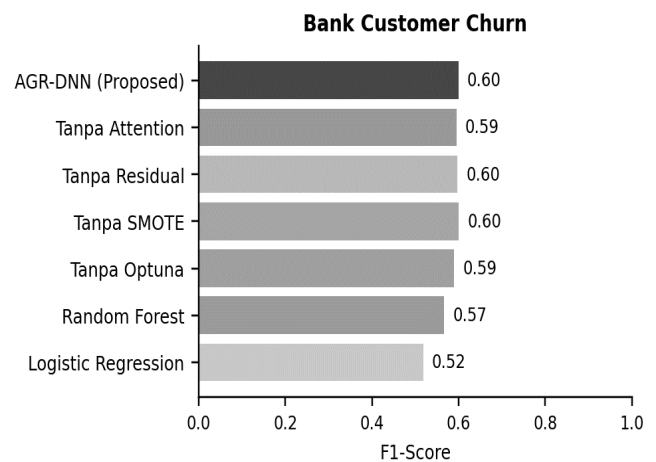
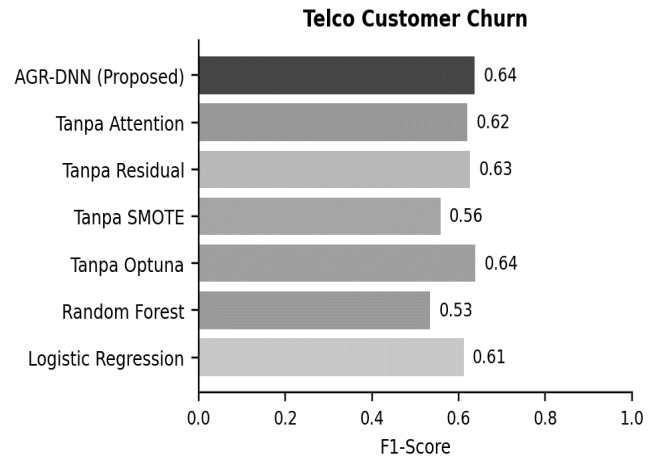
Varian / Model	Accuracy	Precision	Recall	F1-Score	AUC
AGR-DNN	0.75	0.52	0.80	0.63	0.84
Full (Proposed)	78	94	07	74	75
No-Attention	0.75	0.52	0.76	0.62	0.84
	12	18	51	05	12
No-Residual	0.74	0.50	0.81	0.62	0.84
	17	89	49	65	27
No-SMOTE	0.79	0.64	0.49	0.55	0.83
	19	06	47	82	84
No-Optuna (Default HP)	0.75	0.52	0.80	0.63	0.84
	78	93	43	84	57
Random Forest (Baseline)	0.78	0.63	0.45	0.53	0.81
	71	86	91	42	93
Logistic Regression (Baseline)	0.74	0.50	0.76	0.61	0.84
	08	82	87	19	04

Tabel 5. Hasil Evaluasi Model pada Dataset Bank Customer Churn

Varian / Model	Accuracy	Precision	Recall	F1-Score	AUC
AGR-DNN Full (Proposed)	0.8020	0.5092	0.7279	0.5992	0.8595
No-Attention	0.8020	0.5093	0.7148	0.5948	0.8597
No-Residual	0.8087	0.5221	0.6984	0.5975	0.8623
No-SMOTE	0.8687	0.7872	0.4852	0.6004	0.8677
No-Optuna (Default HP)	0.7913	0.4913	0.7377	0.5898	0.8581
Random Forest (Baseline)	0.8613	0.7771	0.4459	0.5667	0.8578
Logistic Regression (Baseline)	0.7260	0.4033	0.7246	0.5182	0.7944

Tabel 6. Hasil Evaluasi Model pada Dataset E-Commerce Customer Churn

Varian / Model	Accuracy	Precision	Recall	F1-Score	AUC
AGR-DNN Full (Proposed)	0.9704	0.8874	0.9437	0.9147	0.9923
No-Attention	0.9728	0.8993	0.9437	0.9210	0.9955
No-Residual	0.9598	0.8506	0.9225	0.8851	0.9902
No-SMOTE	0.9562	0.8523	0.8944	0.8729	0.9889
No-Optuna (Default HP)	0.9562	0.8261	0.9366	0.8779	0.9877
Random Forest (Baseline)	0.9680	0.9675	0.8380	0.8981	0.9918
Logistic Regression (Baseline)	0.7929	0.4358	0.7887	0.5614	0.8669



Gambar 2. Perbandingan F1-Score Seluruh Varian dan Baseline pada Tiga Dataset (Horizontal Bar)

Analisis Studi Ablasi

Kontribusi SMOTE.

Komponen dengan dampak paling konsisten dan substansial adalah SMOTE. Penghapusan SMOTE (varian No-SMOTE) menurunkan *recall* sebesar 30,6

poin persentase pada Telco, 24,3 poin persentase pada Bank, dan 4,9 poin persentase pada E-Commerce. Dampak lebih besar pada Telco dan Bank disebabkan oleh ketidakseimbangan kelas yang lebih tinggi pada kedua dataset tersebut (26,5% dan 20,4%). Ini menjawab RQ2: SMOTE adalah komponen paling kritis untuk optimasi *recall* dalam prediksi churn berbasis *deep learning*.

Kontribusi Attention Gate.

Penghapusan *attention gate* (varian No-Attention) menurunkan rata-rata F1-score sebesar 0,0050 poin dan rata-rata AUC sebesar 0,0010 poin di tiga dataset. Perbedaan moderat ini menunjukkan bahwa *attention gate* memberikan kontribusi nyata pada kemampuan model memprioritaskan fitur informatif, meskipun kontribusinya tidak sebesar SMOTE. Pada dataset E-Commerce yang memiliki 18 fitur (termasuk 7 kategori), *attention gate* tampak lebih diperlukan untuk menyaring noise fitur.

Kontribusi Residual Connection.

Varian No-Residual menunjukkan penurunan rata-rata F1-score sebesar 0,0141 dan AUC sebesar 0,0014. Dampak terbesar terlihat pada dataset E-Commerce (F1 turun dari 0.9147 ke 0.8851), mengindikasikan bahwa *residual connection* sangat berguna ketika model lebih dalam (3+ blok residual) dilatih pada dataset dengan struktur fitur yang kompleks.

Kontribusi Optuna.

Optimasi hyperparameter memberikan peningkatan F1-score paling signifikan pada E-Commerce (+3,7 poin persentase vs. default hyperparameter), sementara perbedaannya relatif kecil pada Telco (-0,1 poin) dan Bank (+0,9 poin). Ini mengindikasikan bahwa manfaat tuning Optuna lebih menonjol pada dataset dengan karakteristik yang lebih kompleks.

Perbandingan dengan Baseline

Model AGR-DNN penuh secara konsisten mengungguli Logistic Regression pada semua metrik dan semua dataset. Dibandingkan Random Forest, AGR-DNN unggul terutama pada *recall*: +34,2 poin persentase pada Telco, +28,2 poin persentase pada Bank, dan +10,6 poin persentase pada E-Commerce. Keunggulan *recall* ini sangat penting dalam konteks CRM karena pelanggan yang benar-benar akan churn

(kelas positif) lebih kritis untuk diidentifikasi daripada pelanggan yang tidak churn. Random Forest cenderung bias terhadap kelas mayoritas dan menghasilkan *recall* rendah meskipun akurasi keseluruhannya tinggi.

Rata-rata AUC AGR-DNN (0,900) melampaui Random Forest (0,890) dan Logistic Regression (0,834), menjawab RQ1 secara afirmatif: AGR-DNN mampu mengungguli baseline konvensional pada tiga domain dataset yang berbeda.

Implikasi bagi Sistem Informasi dan Transformasi Digital

Dari perspektif Sistem Informasi, temuan ini memiliki beberapa implikasi praktis. Pertama, arsitektur AGR-DNN kompatibel dengan data tabular yang tersedia pada sistem CRM, ERP, dan *transaction processing* berbasis basis data relasional tidak memerlukan data citra atau teks. Kedua, proses Optuna dapat diintegrasikan sebagai komponen *automated machine learning* (AutoML) dalam *pipeline* SI prediktif yang berjalan secara terjadwal untuk memperbarui model seiring perubahan perilaku pelanggan. Ketiga, *recall* yang tinggi menghasilkan program retensi yang lebih efisien: sumber daya intervensi terkonsentrasi pada pelanggan yang benar-benar berisiko churn daripada tersebar ke seluruh basis pelanggan.

Konsistensi kinerja AGR-DNN lintas tiga domain (telekomunikasi, perbankan, e-commerce) menjawab RQ3 secara positif: komponen SMOTE dan *residual connection* memberikan kontribusi yang konsisten di semua domain, sehingga arsitektur ini dapat diadaptasi pada berbagai jenis organisasi jasa yang menghadapi tantangan transformasi digital.

IV. KESIMPULAN

Penelitian ini mengusulkan arsitektur AGR-DNN (*Attention-Gated Residual Deep Neural Network*) untuk prediksi *customer churn* dalam konteks transformasi digital organisasi jasa, dilengkapi optimasi hyperparameter berbasis Optuna dan teknik penyeimbangan kelas SMOTE. Model diuji melalui studi ablasi terstruktur pada tiga dataset sekunder publik dari domain telekomunikasi, perbankan, dan e-commerce.

Tiga temuan utama penelitian ini menjawab pertanyaan penelitian yang diajukan: (1) AGR-DNN

secara konsisten mengungguli baseline konvensional (Random Forest dan Logistic Regression) terutama pada *recall*, dengan rata-rata AUC 0,900 versus 0,890 dan 0,834 masing-masing menjawab RQ1; (2) SMOTE adalah komponen dengan kontribusi terbesar terhadap *recall* (penurunan 4,9-30,6 poin persentase ketika dihapus), diikuti *residual connection* (penurunan rata-rata F1 0,0141) dan *attention gate* (penurunan rata-rata F1 0,0050) menjawab RQ2; (3) Kontribusi SMOTE dan *residual connection* bersifat konsisten di tiga domain, sementara keuntungan Optuna paling menonjol pada dataset yang lebih kompleks menjawab RQ3.

Keterbatasan penelitian ini meliputi: eksperimen hanya mencakup tiga dataset dengan rentang ukuran 5.630–10.000 sampel, sehingga generalisasi ke dataset sangat besar belum divalidasi. Selain itu, interpretabilitas prediksi AGR-DNN bagi pemangku kepentingan bisnis masih memerlukan integrasi teknik *explainability* seperti SHAP [35]. Penelitian lanjutan disarankan menggabungkan AGR-DNN dengan teknik *explainable AI* untuk mendukung adopsi model pada SI produksi, serta menguji arsitektur pada dataset *streaming* untuk mendekati skenario transformasi digital secara *real-time*.

V. REFERENSI

- AL-Najjar, D., Al-Rousan, N., & AL-Najjar, H. (2022). Machine learning to develop credit card customer churn prediction. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(4), 1529-1542.
- Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, 13(2), e1484.
- Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2024). Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6), 7499-7519.
- Brauwiers, G., & Frasnica, F. (2023). A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(4), 3279-3298.
- Chicco, D., & Jurman, G. (2023). The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Mining*, 16, 4.
- de Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to customer churn in a financial institution: A machine learning approach. *Neural Computing and Applications*, 34(14), 11751-11768.
- Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, 113(7), 4903-4923.
- Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, 11(1), 185-198.
- Geiler, L., Affeldt, S., & Nadif, M. (2022). A survey on machine learning methods for churn prediction. *International Journal of Data Science and Analytics*, 14(3), 217-242.
- Gorishniy, Y., Rubachev, I., & Babenko, A. (2022). On embeddings for numerical features in tabular deep learning. *Advances in Neural Information Processing Systems*, 35, 24991-25004.
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems*, 35, 507-520.
- Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., Schirrmeyer, R. T., & Hutter, F. (2025). Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045), 319-326.
- Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., & Shao, L. (2023). Normalization techniques in training DNNs: Methodology, analysis and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8), 10173-10196.
- IBM. (2023). *Telco Customer Churn Dataset* [Data set]. Kaggle. <https://www.kaggle.com/datasets/blastchar/telco-customer-churn> (diakses Januari 2025).
- Jajam, N., Challa, N. P., Prasanna, K. S. L., & Deepthi, C. H. V. S. (2023). Arithmetic optimization with ensemble deep learning SBLSTM-RNN-IGSA model for customer churn prediction. *IEEE Access*, 11, 93111-93128.

- Kaggle. (2023). *Bank Customer Churn Modelling Dataset* [Data set]. <https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling> (diakses Januari 2025).
- Khattak, A., Mehak, Z., Ahmad, H., Asghar, M. U., Asghar, M. Z., & Khan, A. (2023). Customer churn prediction using composite deep learning technique. *Scientific Reports*, 13, 17294.
- Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management*, 63, 102466.
- Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: A machine learning approach. *Computing*, 104(2), 271-294.
- Matuszelański, K., & Kopczewska, K. (2022). Customer churn in retail e-commerce business: Spatial and machine learning approach. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(1), 165-198.
- McElfresh, D., Khandagale, S., Valverde, J., Prasad C., V., Ramakrishnan, G., Goldblum, M., & White, C. (2023). When do neural nets outperform boosted trees on tabular data? *Advances in Neural Information Processing Systems*, 36, 76336-76369.
- Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: A comprehensive review. *Artificial Intelligence Review*, 55(5), 3503-3568.
- Nematzadeh, S., Kiani, F., Torkamanian-Afshar, M., & Aydin, N. (2022). Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases. *Computational Biology and Chemistry*, 97, 107619.
- Plekhanov, D., Franke, H., & Netland, T. H. (2023). Digital transformation: A review and research agenda. *European Management Journal*, 41(6), 821-844.
- Prasetyo, E., Widyantoro, D. H., & Sarno, R. (2023). Transformasi digital dan sistem informasi prediktif pada industri jasa di Indonesia. *Jurnal Sistem Informasi*, 19(1), 45-58.
- Rezvani, S., & Wang, X. (2023). A broad review on class imbalance learning techniques. *Applied Soft Computing*, 143, 110415.
- Salehin, I., & Kang, D.-K. (2023). A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics*, 12(14), 3106.
- Setiani, N., & Mustafa, A. (2024). Prediksi churn nasabah perbankan menggunakan CNN-BiLSTM dengan SMOTE. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(2), 201-212.
- Shafiq, M., & Gu, Z. (2022). Deep residual learning for image recognition: A survey. *Applied Sciences*, 12(18), 8972.
- Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84-90.
- Srinivas, P., & Katarya, R. (2022). hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomedical Signal Processing and Control*, 73, 103456.
- Verma, A. (2023). *E-Commerce Customer Churn Analysis and Prediction* [Data set]. Kaggle. <https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction> (diakses Januari 2025).
- Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, 14, 100342.
- Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information*, 14(1), 54.
- Xiahou, X., & Harada, Y. (2022). B2C e-commerce customer churn prediction based on K-means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), 458-475.