

Retensi Nasabah Bertarget: Pendekatan Berbasis RFM dan AI dalam Perbankan

Targeted Customer Retention: An RFM and AI-Based Approach in Banking

Ardijan Handijono¹, Zaldy Suhatman²

Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Pamulang

Jl. Raya Puspatek No. 46 Buaran, Serpong, Tangerang Selatan, Banten, Indonesia, 15417 Email :
email : dosen00853@unpam.ac.id¹, zaldy@unpam.ac.id²

Abstrak

Di tengah lanskap perbankan yang semakin kompetitif, terutama dengan munculnya bank digital, retensi nasabah menjadi prioritas kritis untuk menjamin profitabilitas jangka panjang. Penelitian ini mengintegrasikan analisis RFM (*Recency, Frequency, Monetary*) untuk segmentasi nasabah dengan model prediksi *Churn* berbasis kecerdasan buatan (AI) guna mengembangkan strategi retensi yang lebih efisien dan proaktif. Dengan menganalisis data transaksi historis dari Bank XYZ, kami mengidentifikasi nasabah bernilai tinggi yang berisiko *Churn* berdasarkan skor RFM mereka. Selanjutnya, model *machine learning* yang dikembangkan tidak hanya mampu memprediksi *Churn* secara akurat, tetapi juga secara otomatis memicu intervensi retensi yang terpersonalisasi dan terotomasi melalui sistem CRM. Melalui serangkaian eksperimen A/B testing yang ketat, studi ini menunjukkan bahwa pendekatan sinergis ini menghasilkan penurunan tingkat *Churn* yang signifikan. Tingkat keberhasilan intervensi proaktif mencapai 73,68%, angka yang jauh melampaui metode retensi konvensional. Temuan ini tidak hanya menegaskan pentingnya peran AI dan analisis RFM dalam pengambilan keputusan strategis, tetapi juga menyediakan kerangka kerja yang kuat bagi industri perbankan untuk mengoptimalkan alokasi sumber daya dan membangun strategi retensi yang berkelanjutan di era digital.

Kata Kunci: *Customer Churn, Machine Learning, Analisis RFM, Retensi Nasabah, Perbankan Digital.*

Abstract

In an increasingly competitive banking landscape, particularly with the emergence of digital banks, customer retention has become a critical priority for ensuring long-term profitability. This study integrates RFM (*Recency, Frequency, Monetary*) analysis for customer segmentation with an AI-based churn prediction model to develop a more efficient and proactive retention strategy. By analyzing historical transaction data from Bank XYZ, we identified high-value customers at risk of churn based on their RFM scores.

Subsequently, the developed machine learning model not only accurately predicts churn but also automatically triggers personalized and automated retention interventions through a CRM system. Through a series of rigorous A/B testing experiments, this study demonstrates that this synergistic approach results in a significant decrease in the churn rate. The success rate of the proactive interventions reached 73.68%, a figure that substantially exceeds conventional retention methods.

These findings not only underscore the vital role of AI and RFM analysis in strategic decision-making but also provide a robust framework for the banking industry to optimize resource allocation and build a 'sustainable' retention strategy in the digital era.

Keywords: Customer Churn, Machine Learning, RFM Analysis, Customer Retention, Digital Banking

1. Pendahuluan

Dalam industri perbankan, *customer retention* (retensi nasabah) menjadi kunci utama menjaga profitabilitas dan pertumbuhan bisnis. (Zarkesh, 2023) secara jelas menunjukkan bahwa biaya akuisisi pelanggan baru CAC (*Customer Acquisition Cost*) bisa mencapai lima kali lipat lebih mahal dibandingkan biaya mempertahankan pelanggan yang sudah ada. Angka ini semakin diperkuat oleh temuan (Tuguinay et al., 2024), di sektor perbankan digital, CAC bisa mencapai **150–300 per nasabah**, sementara biaya retensi CRC (*Customer Retention Cost*) hanya **25–50**.

Customer churn terjadi karena perusahaan tidak mampu mempertahankan pelanggan yang sudah ada. *Customer churn*, merujuk pada situasi di mana seorang pelanggan mengakhiri hubungannya dengan penyedia layanan atau produk (Văduva et al., 2024). Dalam konteks perbankan, *Churn* nasabah dapat bervariasi dari penutupan rekening tabungan, penarikan deposito, hingga pengalihan fasilitas kredit ke bank lain. Fenomena ini semakin krusial mengingat tingkat *Churn* (kehilangan nasabah) di perbankan konvensional terus meningkat, terutama karena persaingan dengan bank digital dan fintech yang menawarkan suku bunga lebih tinggi. Menurut (Mayasari & Mahadi, 2023). Data dari Otoritas Jasa Keuangan (OJK) tahun 2023 menunjukkan bahwa *Churn* rate nasabah deposito di Indonesia mencapai **19,8%**, dengan penyebab utama adalah perbedaan suku bunga yang ditawarkan kompetitor. Sebagai contoh, beberapa bank digital memberikan **bunga deposito 1-2% lebih tinggi** dibanding bank konvensional, sehingga nasabah dengan mudah berpindah penyedia jasa. Selain itu, nasabah yang sudah *Churn* memiliki kemungkinan 50% lebih kecil untuk

kembali, sehingga upaya pencegahan *Churn* jauh lebih efektif daripada mengakuisisi penggantinya. Bank yang berhasil mengurangi churn rate sebesar 5% dapat meningkatkan profitabilitas hingga **25-95%**, tergantung pada segmentasi nasabah (Otiso, 2024). Mengingat dampak destruktif *churn*, maka pendekatan reaktif (bertindak setelah nasabah pergi) menjadi tidak efisien. Di sinilah peran **prediksi churn** (*churn prediction*) menjadi krusial. menurut (Brito et al., 2024) prediksi *churn* dapat dilakukan dengan menggunakan teknik analitik khususnya AI untuk mengidentifikasi nasabah yang memiliki kemungkinan tinggi untuk churn.

Churn Labeling

Pada industri perbankan, khususnya untuk produk Tabungan, menurut (Verma, 2020) status *churn* nasabah dapat dilihat dari empat gejala sebagai berikut: a) **Rasio Debit terhadap Kredit Rata-rata Meningkat $\geq 200\%$** , Ini menunjukkan bahwa jumlah penarikan dana nasabah dalam suatu periode meningkat setidaknya dua kali lipat dibandingkan dengan setoran. b) **Jumlah Debit Rata-rata Meningkat $\geq 200\%$** , berarti jumlah total uang yang ditarik oleh nasabah dalam suatu periode meningkat setidaknya dua kali lipat. c) **Saldo Rata-rata Minimum (*Minimum Average Balance/MAB*) Menurun $\geq 70\%$** , menunjukkan bahwa saldo rata-rata minimum yang dipegang oleh nasabah dalam akunnya menurun setidaknya 70%. d) **Jumlah Kredit Menurun $\geq 70\%$** , berarti jumlah total uang yang disetor oleh nasabah dalam suatu periode menurun setidaknya 70%. Jika setidaknya tiga dari empat kondisi di atas terpenuhi, akun nasabah tersebut dapat diklasifikasikan sebagai *Churn*.

Segmentasi berbasis RFM

Analisis RFM (*Recency, Frequency, Monetary*) adalah cara yang efektif untuk mengelompokkan pelanggan berdasarkan perilaku belanja mereka. Metode ini mengevaluasi tiga aspek kunci: *Recency* (kapan terakhir kali pelanggan membeli), *Frequency* (seberapa sering mereka membeli), dan *Monetary* (berapa banyak uang yang mereka habiskan). Dengan menggunakan ketiga metrik ini, RFM memberikan gambaran lengkap tentang loyalitas dan nilai setiap pelanggan. Ini memungkinkan perusahaan untuk mengidentifikasi pelanggan yang paling berharga dan membedakan mereka dari yang kurang bernilai. Pelanggan dengan skor RFM tinggi umumnya dianggap sebagai aset berharga yang harus dipertahankan. Analisis RFM juga sangat membantu dalam alokasi sumber daya. Perusahaan dapat memprioritaskan penanganan pelanggan, terutama yang berisiko meninggalkan layanan (*Churn*). Misalnya, perusahaan bisa memberikan insentif menarik hanya kepada pelanggan bernilai tinggi yang diprediksi akan *Churn* untuk mempertahankan mereka. Di sisi lain, perusahaan dapat memilih untuk tidak berinvestasi banyak pada pelanggan yang kurang bernilai dan membiarkan mereka pergi. Hal ini menjadikan strategi bisnis lebih efisien dan terfokus. (Handijono & Suhatman, 2024). Dengan mengetahui RFM *Score* maka dapat diidentifikasi pelanggan termasuk pada segmen yang mana dari daftar 11 segment berikut: *Champions, Loyal, Potential Loyalist, New Customers, Promising, Need Attention, About To Sleep At Risk, Cannot Lose Them, Hibernating Customers*, dan *Lost customers*. (Owoade, 2025)

Integrasi AI dan CRM

Menurut (Johrawanshi et al., 2024) *Churn prediction* berbasis AI dan integrasinya dengan CRM menjadi solusi strategis. CRM adalah strategi bisnis yang berpusat pada pelanggan, dirancang untuk mengelola dan menganalisis interaksi pelanggan dan data sepanjang siklus hidup pelanggan, dengan tujuan meningkatkan hubungan layanan pelanggan dan membantu pertumbuhan penjualan (Azeem et al., 2022). CRM bukan hanya

sekadar perangkat lunak, CRM adalah filosofi bisnis yang mengintegrasikan orang, proses, dan teknologi untuk membangun hubungan yang lebih kuat dengan pelanggan (Ijomah et al., 2024).

Menurut (Patil & Mohammad, 2023) integrasi AI dan CRM bermanfaat untuk:

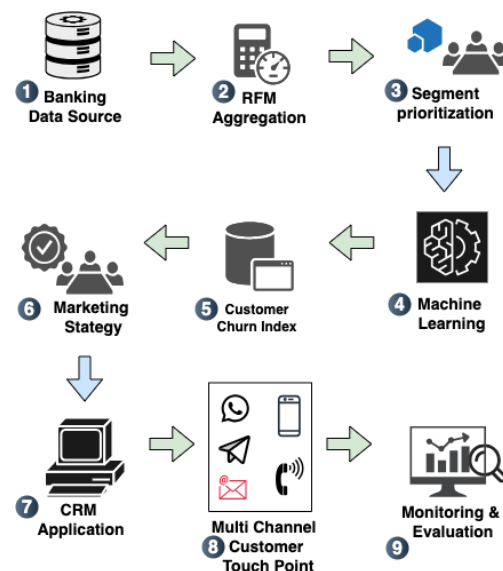
- a) Mengidentifikasi nasabah berisiko *Churn* sebelum mereka benar-benar pergi.
- b) Memberikan penawaran yang telah dipersonalisasi (seperti penyesuaian suku bunga atau layanan eksklusif) untuk meningkatkan loyalitas.
- c) Meng-optimalkan biaya retensi dengan menghindari strategi promosi yang tidak tepat sasaran.

Dengan memanfaatkan data transaksi historis dan *behavioral analytics*, bank dapat mengambil langkah proaktif guna meminimalkan *Churn* dan mempertahankan profitabilitas jangka panjang (Verma, 2020). Penelitian ini bertujuan mengukur dampak integrasi *AI Churn Prediction* dengan CRM *automation* di Bank XYZ, khususnya dalam menghadapi gempuran persaingan dari bank digital. Penelitian ini juga mengusulkan rancangan sistem terpadu untuk mengidentifikasi potensi *Churn* nasabah secara dini dan menindaklanjutinya melalui strategi pemasaran yang telah dipersonalisasi. Alur proses integrasi *AI Churn Prediction* dan CRM adalah sebagai berikut:

1) Akuisisi dan Integrasi Data - Langkah awal dari sistem ini dimulai dengan ekstraksi data nasabah dari sistem inti perbankan (*core banking system*). Data yang dikumpulkan mencakup tiga kategori utama, yaitu: a. Data demografi, seperti usia, jenis kelamin, dan domisili, b. Data akun bank, termasuk tipe rekening, saldo rata-rata, serta status aktif rekening, dan c. Riwayat transaksi, seperti frekuensi transaksi, nominal transaksi, serta jenis transaksi (debit, kredit, transfer, dsb.). Data tersebut diolah dan dibersihkan sebelum digunakan dalam proses selanjutnya. **2) RFM Aggregation** - Adalah proses agregasi RFM yang merupakan langkah fundamental untuk mengkuantifikasi nilai nasabah dan menjadi landasan strategi retensi yang

efektif. Proses ini diawali dengan menganalisis data transaksi historis untuk menghitung skor '**Recency**', '**Frequency**', dan '**Monetary**' bagi setiap nasabah. Skor yang dihasilkan kemudian dipetakan ke dalam segmen yang telah ditentukan, menghasilkan 11 segmen nasabah dengan karakteristik perilaku yang unik. Hasil segmentasi ini memungkinkan perusahaan untuk mengidentifikasi **nasabah bernilai tinggi** yang mungkin berisiko **Churn**, sehingga alokasi sumber daya dapat dioptimalkan untuk intervensi retensi yang sangat terfokus. 3) **Segment Prioritization** - Tahap ini adalah menentukan Prioritas Segmen, di mana tim pemasaran menentukan segmen mana yang akan menjadi target utama untuk upaya retensi. Tim akan mengidentifikasi **nasabah bernilai tinggi** yang diprediksi akan meninggalkan layanan. Dengan pendekatan selektif ini, sumber daya tidak dihabiskan untuk semua nasabah yang berisiko, melainkan difokuskan secara eksklusif pada segmen berharga yang paling layak dipertahankan. Hanya nasabah dalam segmen prioritas inilah yang akan menerima intervensi proaktif, seperti hadiah atau insentif bernilai tinggi, yang bertujuan mengubah mereka dari nasabah yang berisiko menjadi pelanggan setia dan loyal. 4) **Pemodelan Prediksi Churn** - Data historis yang telah disiapkan digunakan sebagai masukan (fitur) untuk *Churn Prediction* yang dikembangkan dengan pendekatan **Machine Learning**. Model ini dilatih untuk mengidentifikasi pola perilaku yang mengindikasikan potensi *Churn*. 5) **Output** dari model berupa skor **Churn Index** dengan rentang nilai 0 hingga 100, di mana semakin tinggi skor tersebut, semakin besar kemungkinan nasabah akan meninggalkan layanan bank. 6) **Strategi Retensi** - Hasil dari model prediktif kemudian dianalisis oleh tim pemasaran (*marketing*). Berdasarkan kebijakan internal atau hasil evaluasi berkala, ditetapkan ambang batas (*threshold*) skor *Churn* yang dianggap perlu ditindaklanjuti, misalnya skor ≥ 70 . Nasabah yang termasuk dalam segmen ini akan masuk ke dalam daftar prioritas penanganan *Churn*. Tim pemasaran selanjutnya menentukan strategi retensi yang relevan dan kompetitif untuk

ditawarkan kepada nasabah berisiko tinggi tersebut. Strategi yang dapat diterapkan mencakup, namun tidak terbatas pada: a. Pemberian bebas biaya administrasi selama periode tertentu, b. Penghapusan biaya transaksi (misalnya transfer antarbank), c. Penawaran suku bunga kredit yang lebih rendah, atau e. Peningkatan batas transaksi dan fasilitas premium. Strategi yang dipilih dapat disesuaikan dengan segmentasi nasabah maupun histori interaksi sebelumnya. 7) **Implementasi Melalui Sistem CRM** - Setelah strategi retensi ditetapkan, proses pelaksanaannya dilakukan melalui aplikasi CRM yang terintegrasi. Sistem CRM bertindak sebagai platform pencatatan, pelacakan, dan pengukuran efektivitas intervensi yang diberikan. Setiap tindakan dan hasilnya dicatat untuk evaluasi berkala dan penyempurnaan strategi di masa mendatang. 8) **Personalisasi Saluran Komunikasi** - Sistem CRM mendukung berbagai saluran komunikasi yang disesuaikan dengan preferensi masing-masing nasabah. Saluran ini meliputi: a. Pesan instan,



Gambar 1.
Proses Bisnis Integrasi AI dan CRM

seperti WhatsApp atau Telegram, b. Email, untuk penyampaian informasi tertulis dan dokumentasi, c. Panggilan suara (telemarketing), untuk pendekatan yang lebih personal, dan d. Notifikasi dalam

aplikasi mobile banking, untuk integrasi langsung dengan ekosistem digital bank.

2. Metode Penelitian

Framework yang digunakan dalam penelitian ini menggunakan standard industri khusus untuk *Machie Learning* yakni CRISP-ML (*Cross-Industry Standard Process for Machine Learning*) (Handijono, 2022). Berikut tahapan-tahapan CRISP-ML:

1) **Business and Data Understanding** - Menetapkan tujuan bisnis, mengumpulkan data relevan, dan memahami karakteristik serta kriteria keberhasilan proyek. 2) **Data Preparation** - Membersihkan dan mentransformasi data (normalisasi, encoding, isi nilai hilang), lalu membaginya menjadi data train, data validasi, dan data uji. 3) **Model Engineering** - Memilih algoritma ML, melatih model, dan menyetel hyperparameter untuk hasil optimal. 4) **Quality Assurance** - Memvalidasi model, mengatasi bias dan *overfitting*, serta mendokumentasikan proses. 5) **Deployment** - Menerapkan model ke produksi dan mengintegrasikannya ke sistem yang ada, serta melakukan pengujian akhir. 6) **Monitoring and Maintenance** - Memantau performa model, melakukan retraining jika perlu, dan memastikan model tetap andal. Penulis membangun model *Churn Prediction* ini dirancang untuk dapat memperkirakan kemungkinan pelanggan akan berhenti berlangganan. Model ini menggunakan data historis selama enam bulan terakhir (*time window*) untuk memprediksi *Churn* dalam rentang waktu tertentu. Secara spesifik, model ini memprediksi *Churn* untuk periode satu bulan ke depan (T+1).

Aplikasi CRM

Aplikasi CRM yang digunakan adalah Zoho CRM. Zoho adalah platform manajemen hubungan pelanggan (CRM) berbasis *cloud* yang komprehensif, dirancang untuk membantu perusahaan dari berbagai skala dalam mengelola interaksi dengan pelanggan. Platform ini mengintegrasikan fitur otomatisasi penjualan untuk mengelola prospek dan *pipeline*, otomatisasi pemasaran untuk menjalankan promosi yang efektif,

Workflow Automation untuk mengotomatisasi tugas-tugas harian seperti pengiriman email tindak lanjut, penjadwalan janji temu, dan pembaruan catatan secara otomatis. dan layanan pelanggan untuk menangani tiket dukungan. Zoho CRM juga menyediakan fitur analisis dan pelaporan yang kuat, memungkinkan pengguna untuk memantau kinerja bisnis secara *real-time* dan membuat keputusan strategis berbasis data. Dengan kemampuannya yang luas, Zoho CRM berfungsi sebagai alat penting untuk meningkatkan efisiensi operasional dan memperkuat hubungan dengan pelanggan. (Shabdar, 2017)

Pertimbangan Etika dan Keamanan Data

Penelitian ini dilakukan dengan menjunjung tinggi prinsip etika dan perlindungan data pribadi nasabah. Seluruh proses pengumpulan, pengolahan, dan analisis data tunduk pada regulasi yang berlaku, termasuk **Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi** di Indonesia, serta Peraturan Otoritas Jasa Keuangan (OJK) dan Bank Indonesia (BI) terkait kerahasiaan data nasabah (Setiawan & Najicha, 2022). Sebagai acuan standar global, penelitian ini juga mengadopsi prinsip-prinsip yang selaras dengan **General Data Protection Regulation (GDPR)**. Untuk memastikan privasi nasabah terjaga sepenuhnya, semua data identitas dianonimkan secara ketat dan hanya data transaksi yang relevan yang diekstrak. Dengan demikian, kami menjamin bahwa penelitian ini telah dilakukan secara etis dan bertanggung jawab, memastikan data nasabah aman dan terlindungi dari segala bentuk penyalahgunaan.

3. Hasil dan Pembahasan

Data mentah diekstraksi dari sistem *Core Banking* Bank XYZ yaitu salah satu bank swasta. Semua informasi yang bersifat rahasia telah disamarkan. Data mentah diekstraksi dari sistem *Core Banking* dengan *ETL tool*. Data transaksi Tabungan selama 6 bulan sebesar **10.474.817** raws dan Data nasabah yang ditarik **675.494** raws, namun yang masih aktif hanya **85.925** nasabah, karena Bank tidak pernah menghapus data nasabah

Dorman yang sudah bertahun-tahun tidak aktif dan Bank hanya mengubah statusnya saja. Program ETL juga telah melakukan perhitungan RFM untuk mendapatkan Segment nasabah. Distribusi segment nasabah ditunjukkan pada Tabel-1.

Tabel 1. Hasil Segmentasi RFM

No	Segment	R	F	M	Customer	Percent
1	Champions	4 - 5	4 - 5	4 - 5	-	0,00
2	Loyal Customers	2 - 4	3 - 4	4 - 5	-	0,00
3	Potential Loyalist	3 - 5	1 - 3	1 - 3	5	0,01
4	New Customers	4 - 5	< 2	< 2	23	0,03
5	Promising	3 - 4	< 2	< 2	500	0,58
6	Needing Attention	3 - 4	3 - 4	3 - 4	17	0,02
7	About To Sleep	2 - 3	< 3	< 3	282	0,33
8	At Risk	< 3	2 - 5	2 - 5	26.947	31,36
9	Can't Lose Them	< 2	4 - 5	4 - 5	24.835	28,90
10	Hibernating	2 - 3	2 - 3	2 - 3	2.252	2,62
11	Lost	< 2	< 2	< 2	31.064	36,15
	Total				85.925	100,00

Dalam rangka optimalisasi biaya, Bank menerapkan strategi dengan menawarkan insentif khusus kepada nasabah yang teridentifikasi memiliki **nilai tinggi** dan diprediksi berpotensi untuk *Churn*. Pemberian insentif ini bertujuan untuk meningkatkan retensi nasabah. Setelah analisis segmentasi RFM menggunakan strategi seperti ditunjukkan pada Tabel-2, Bank mengidentifikasi empat segmen prioritas untuk upaya retensi, yaitu segmen *Can't Lose Them*, *Promising*, *Needing Attention*, dan *Potential Loyalist*. Secara total, terdapat **25.357** nasabah yang termasuk dalam segmen-segmen tersebut dan menjadi target utama program retensi ini. Setelah menentukan segmen yang akan diselamatkan dari *Churn*, selanjutnya akan diproses dengan *Machine Learning* mengikuti kerangka kerja **CRISP-ML** :

1) Business and Data Understanding – Tujuan dari *Machine Learning* yang dibangun adalah untuk memprediksi apakah seorang nasabah bank ada indikasi akan *Churn* atau tidak. Algoritma yang akan digunakan adalah **Supervised Learning** sehingga diperlukan label untuk *Train* data. Pada umumnya jumlah nasabah yang berstatus *Churn* akan sangat sedikit sekali.

2) Data Preparation - Persiapan data adalah tahap krusial dalam proyek

Machine Learning (ML) Proses ini harus dilakukan secara sistematis untuk memastikan model yang dikembangkan valid dan dapat digeneralisasi dengan baik.

Tabel 2.

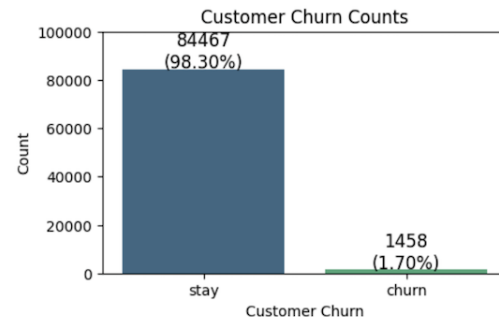
RFM dan Strategi Bisnis (Robo et al., 2024)

No	Segment	Characteristic	Suggested Next Steps
1	Champions	Bought recently, buy often and spend the most!	Reward them. Can be early adopters for new products. Will promote your brand.
2	Loyal Customers	Spend good money with us often. Responsive to	Upsell higher value products. Ask for reviews.
3	Potential Loyalist	Recent customers, but spent agood amount and bought more than once.	Offer membership /loyalty program, recommend other products.
4	Recent Customers	Bought most recently, but not often.	Provide onboarding support, give them early success, start building relationship.
5	Promising Customers	Recent shoppers, but haven't spent much.	Create brand awareness, offer free trials
6	Needing Attention	Above average recency, frequency and. monetary values. May not have bought	Make limited time offers, Recommend based on past purchases. Reactivate
7	About To Sleep	Below average recency, frequency and. monetary values. Will lose them if not reactivated.	Share valuable resources, recommend popular products /renewals at discount, reconnect with
8	At Risk	Spent big money and purchased often. But long time ago. Need to bring	Send personalized emails to reconnect, offer renewals, provide helpful resources.
9	Can't Lose Them	Made biggest purchases, and often. But haven't returned for along time.	in them back via renewals or newer products, don't lose them to competition,
10	Hibernating	Last purchase was long back, low spenders and bought seldomly.	Offer other relevant products and special discounts.Recreate brand
11	Lost	Lowest recency, frequency and monetaryscores (RFM	Revive interest with reach out campaign, ignore

Kehati-hatian sangat penting untuk mencegah *Data Leakage*, di mana informasi dari Test Dataset secara tidak sengaja "bocor" ke *Training Dataset*. *Data Leakage* sering kali menjadi penyebab utama *overfitting*, sebuah kondisi di mana model bekerja sangat baik pada Training Dataset yang sudah dikenalnya tetapi gagal berkinerja baik pada data baru atau *Test Dataset*. (Domnik & Holland, 2022). Berikut adalah kerangka kerja dalam Data Preparation: **a) Data Acquisition and Integration** – dalam hal ini adalah **25.357** raws data nasabah yang sudah diseleksi untuk segmen tertentu. **b) Exploratory Data Analysis (EDA)** – untuk memahami data akan dilakukan Identifikasi Pola dan Distribusi: Menganalisis sebaran variabel-variabel kunci, seperti usia pelanggan, jumlah transaksi, dan nilai rata-rata transaksi, Deteksi Anomali: Mengidentifikasi *outlier* atau nilai-nilai yang tidak biasa, yang mungkin mengindikasikan kesalahan entri data

atau fenomena yang signifikan dan Penilaian Kualitas Data: Mengevaluasi keberadaan nilai yang hilang (*null values*), inkonsistensi, atau kesalahan format. c) **Pre-Splitting Preprocessing** – Menyiapkan seluruh Dataset meliputi penanganan Nilai Hilang: dengan teknik Imputasi antara lain, dengan menggunakan nilai rata-rata, median, atau modus dari fitur yang relevan pada seluruh dataset, pengodean fitur kategorikal: mengubah variabel kategorikal menjadi format numerik menggunakan metode seperti *one-hot encoding* atau *label encoding*. d) **Data Splitting (Pembagian Dataset)** - Setelah data nasabah siap, Dataset dibagi menjadi dua bagian dengan rasio, *Training Dataset* (80%) dan *Test Dataset* (20%). Splitting dilakukan dengan menggunakan **stratified sampling** untuk memastikan bahwa distribusi kelas dalam setiap subset sama dengan distribusi dalam dataset asli. Hasil dari proses *Splitting* ini untuk *Traning Dataset* sebanyak 20.287 nasabah dan *Test Dataset* sebanyak 5.070 nasabah. e) **Post-Splitting Preprocessing** - Setelah dataset dibagi dua, akan dilakukan langkah-langkah *Preprocessing* berikut : **Scaling** (Penskalaan) dan Normalisasi: proses ini dilakukan secara terpisah untuk setiap subset, **Scaling** yang dilakukan secara terpisah ini mencegah kebocoran data karena informasi statistik dari *Test Dataset* tidak bocor ke *Training Dataset*. Skema fitur yang digunakan, beserta deskripsi dan tipe datanya, disajikan pada Tabel-1. e) **Feature Engineering** adalah proses kreatif untuk membuat fitur baru dari fitur yang sudah ada, tujuannya adalah untuk memberikan informasi yang lebih kaya kepada model. (Katya, 2023). Pada sistem ini yang menggunakan metode **Supervised, Learning** maka fitur utama yang harus ditentukan adalah *Label*, ini adalah fitur penting dalam proses untuk menentukan label seorang nasabah dalam status *Churn* atau *Not Churn*, berdasarkan empat kondisi (Verma, 2020). Data transaksi selama enam bulan dipilah menjadi dua kelompok transaksi yaitu transaksi dalam tiga bulan pertama dan transaksi dalam tiga bulan terakhir, kemudian di populasi : Total kredit tiga bulan pertama, Total kredit tiga bulan terakhir, Total debit tiga bulan

pertama tiga Total debit tiga bulan terakhir untuk menghitung rasio, tingkat kenaikan atau penurunannya sehingga dapat ditentukan status *Churn* nya. Gambar-2 menunjukkan bahwa nasabah dengan label *Churn* hanya 1.70%.



Gambar 2. Proporsi Customer Churn

Penanganan *Imbalanced Dataset*

Training Dataset memiliki proporsi kelas yang tidak seimbang (*imbalanced dataset*), di mana jumlah pelanggan yang *churn* hanya kecil sekali dari total data. Kondisi ini dapat menyebabkan bias pada model *Machine Learning*, di mana algoritma cenderung mengabaikan kelas minoritas dan menghasilkan prediksi yang kurang akurat untuk kasus tersebut, untuk mengatasi ketidakseimbangan ini, kami menerapkan teknik **SMOTE** (*Synthetic Minority Over-sampling Technique*). Metode SMOTE hanya dilakukan pada *Training Dataset* saja. Metode ini bekerja dengan menciptakan sampel sintetis dari kelas minoritas, sehingga distribusi kelas menjadi lebih seimbang. Dengan demikian, model dapat belajar dari kedua kelas secara efektif, meningkatkan kemampuannya dalam mengidentifikasi pelanggan yang berpotensi *churn*. (Mqadi et al., 2021).

3 Model Engineering - Algoritma yang dipilih dari keluarga *ensemble learning* berbasis **Decision Tree**, yang sangat efektif untuk tugas prediksi seperti *Churn*.

Algoritma tersebut antara lain: a) **RandomForest** - algoritma ini membangun banyak **Decision Tree** secara independen dan menggabungkan hasilnya. Ini membuatnya sangat tangguh terhadap **overfitting** dan mampu mengidentifikasi fitur-fitur penting yang memengaruhi *churn* tanpa memerlukan penyesuaian yang rumit. b). **XGBoost** - Performa dan

akurasi tinggi. Algoritma ini membangun Tree secara berurutan, di mana setiap Tree baru memperbaiki kesalahan dari Tree sebelumnya. Hal ini menjadikannya sangat kuat dalam menangani data yang tidak seimbang, yang umum terjadi dalam kasus *churn*. c) **LightGBM** - ideal untuk *Dataset* yang besar dan butuh efisiensi komputasi. Algoritma ini menggunakan teknik unik untuk membangun *Tree* lebih cepat daripada algoritma *boosting* lainnya, sehingga menghemat waktu pelatihan tanpa mengorbankan banyak akurasi. d) **ExtraTrees** - varian dari **RandomForest** yang dirancang untuk kecepatan. Dengan memilih pemisahan node secara acak, algoritma ini mempercepat proses *Training* secara signifikan, sekaligus efektif dalam mengurangi *overfitting*.

Hyperparameter Tuning

Untuk mencapai kinerja model yang optimal, diperlukan *hyperparameter tuning*. Proses ini bertujuan untuk menemukan kombinasi nilai *hyperparameter* terbaik guna menyeimbangkan bias dan varians, serta mencegah *overfitting* atau *underfitting* (a Ilemobayo et al., 2024). Pada penelitian ini, kami menggunakan dua metode *tuning* yang berbeda untuk setiap algoritma:

RandomizedSearchCV - Metode ini digunakan untuk algoritma **RandomForest**, **XGBoost**, dan **LightGBM**. Kami memilih metode ini karena ketiga algoritma tersebut memiliki ruang pencarian *hyperparameter* yang sangat luas. **RandomizedSearchCV** secara efisien menguji kombinasi nilai secara acak, yang lebih praktis daripada menguji semua kemungkinan kombinasi. Ruang pencarian untuk setiap algoritma adalah sebagai berikut:

RandomForest:

```
random_param = {"max_depth": [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12],
                 'n_estimators': [50, 100, 200],
                 'min_samples_leaf': [1, 2, 4],
                 'min_samples_split': [2, 5, 10],
                 "criterion": ["gini", "entropy"]
}
```

XGBoost:

```
param_grid = {
    'n_estimators': [100, 200],
    'max_depth': [3, 5, 7],
```

```
'learning_rate': [0.1, 0.5, 1.0],
'subsample': [0.7, 1.0],
'gamma': [0, 0.1, 0.2],
'colsample_bytree': [0.8, 1.0],
'reg_alpha': [0, 0.1, 0.5],
'reg_lambda': [1, 1.5, 2]
}
```

LightGBM:

```
param_dist = {
    'num_leaves': stats.randint(20, 41),
    'max_depth': stats.randint(5, 11),
    'learning_rate': stats.uniform(0.05, 0.15),
    'min_child_samples': stats.randint(20, 51),
    'subsample': stats.uniform(0.7, 0.2),
    'colsample_bytree': stats.uniform(0.7, 0.2),
    'reg_lambda': stats.uniform(0.1, 0.9)
}
```

GridSearchCV - Metode ini diterapkan pada algoritma **ExtraTrees**. Kami memilih **GridSearchCV** karena algoritma **ExtraTrees** memiliki jumlah *hyperparameter* penting yang lebih sedikit dibandingkan algoritma lain. Metode ini akan menguji setiap kombinasi yang mungkin dari ruang pencarian yang ditentukan, sehingga menjamin hasil optimal secara matematis. (Handijono & Suhatman, 2024).

Ruang pencarian untuk algoritma **ExtraTrees** adalah sebagai berikut:

ExtraTrees:

```
param_grid = {
    'n_estimators': [50, 100, 200],
    'max_depth': [10, 20, 30, None],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}
```

4. Model Validation and Evaluation

Metrik-metrik berikut digunakan untuk mengukur seberapa baik model dapat memprediksi nasabah yang akan *churn* atau tidak.

- a) **Confusion Matrix** - adalah tabel yang menunjukkan performa prediksi model menjadi empat kategori seperti dijelaskan pada Tabel- 4:

Tabel 3. Skema Fitur

No.	Feature	Type	Description
1	customer id	Numeric	ID unik untuk setiap pelanggan bank.
2	gender	Categorical	Jenis kelamin pelanggan (M = Male, F = Female).
3	marital_status	Categorical	Status perkawinan pelanggan.
4	place_of_birth	Categorical	Tempat kelahiran pelanggan.
5	post_code	Numeric	Kode pos tempat tinggal pelanggan.
6	date_of_birth	Date	Tanggal lahir pelanggan.
7	town_country	Categorical	daerah (kota atau kabupaten) tempat tinggal pelanggan.
8	kyc_incom_rng	Numeric	Rentang pendapatan pelanggan.
9	open_date	Date	Tanggal pelanggan terdaftar.
10	branch	Numeric	ID cabang bank tempat rekening dibuka.
11	job	Categorical	Tipe Pekerjaan pelanggan.
12	education	Categorical	Tingkat pendidikan terakhir pelanggan.
13	homeownership	Categorical	Status kepemilikan rumah pelanggan.
14	generation	Categorical	Generasi pelanggan berdasarkan usia (Gen X, Gen Z, Millennial, dst.).
15	inactive_marker	Categorical	Indikator apakah akun pelanggan tidak aktif.
16	segment	Categorical	Segment pelanggan yang dihitung dari RFM Score
17	total_bal	Numeric	Total saldo keseluruhan pelanggan.
18	total_sav	Numeric	jumlah rekening yang dimiliki pelanggan.
19	mean_in_taken	Numeric	Rata-rata pinjaman yang diambil oleh pelanggan.
20	ln_count	Numeric	berapa kali pelanggan meminjam.
21	avg_collect	Numeric	Rata rata definisi kolektabilitas pelanggan dalam melakukan pembayaran hutang. (Lancar, WriteOffs)
22	wo_total	Numeric	jumlah gagal bayar yang dilakukan oleh pelanggan.
23	wo_mark	Categorical	Pernah mengalami writeoff (gagal bayar) atau tidak (No dan Yes)
24	current_in_count	Numeric	Jumlah pinjaman yang sedang berjalan.
25	current_in_amo nt	Numeric	Jumlah total pinjaman yang sedang berjalan.
26	total_depo	Numeric	Total deposito yang dimiliki oleh pelanggan (jumlah uang).
27	avg_depo_taken	Numeric	Rata-rata deposito yang diambil oleh pelanggan (jumlah uang).
28	avg_rollover_ter m	Numeric	Rata-rata masa jangka waktu deposito oleh pelanggan.
29	avg_bal	Numeric	Rata-rata saldo harian pelanggan.
30	trans_count	Numeric	Jumlah transaksi yang dilakukan oleh pelanggan (frequency).
31	credit_trans_amo unt	Numeric	Jumlah total transaksi kredit pelanggan (jumlah uang monetary).
32	recency_score	Numeric	Skor pelanggan terakhir kali melakukan transaksi
33	collect_score	Numeric	Skor pelanggan dalam melakukan pembayaran hutang dalam 3 bulan terakhir dengan range 0-36.
34	credit_first_3_mo nth	Numeric	jumlah credit pelanggan pada 3 bulan pertama
35	credit_last_3_mo nth	Numeric	jumlah credit pelanggan pada 3 bulan terakhir
36	debit_first_3_mo nth	Numeric	jumlah debit pelanggan pada 3 bulan pertama
37	debit_last_3_mon th	Numeric	jumlah debit pelanggan pada 3 bulan terakhir
38	bal_first_3_mont h	Numeric	jumlah balance atau saldo rata rata diakhir bulan pada 3 bulan pertama
39	bal_last_3_month	Numeric	jumlah balance atau saldo rata rata diakhir bulan pada 3 bulan terakhir.

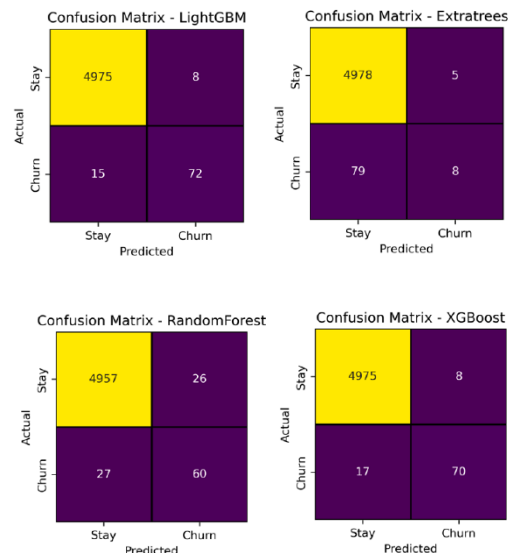
Tabel 4. Confusion Matrix Indicator

Indicator	Meaning
<i>True-Positive (TP)</i>	Model dengan benar memprediksi nasabah akan churn.
<i>True-Negative (TN)</i>	Model dengan benar memprediksi nasabah tidak akan churn.
<i>False-Positive (FP)</i>	Model salah memprediksi nasabah akan churn (padahal tidak).
<i>False-Negative (FN)</i>	Model salah memprediksi nasabah tidak akan churn (padahal seharusnya churn).

b) *Precision* dan *Recall*. Keduanya dapat dihitung dari nilai-nilai dalam *Confusion Matrix*: ***Precision*** - Mengukur seberapa banyak dari semua prediksi positif yang benar-benar positif. ***Recall*** (Ingatan): Mengukur seberapa banyak dari semua kasus positif yang berhasil ditangkap oleh model.

$$Precision = \frac{TP}{(TP+FP)}$$

$$Recall = \frac{TP}{(TP+FN)}$$



Gambar 3. Confusion Matrix

1. Precision (kelas churn = positif)

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{60}{60 + 26} = \frac{60}{86} \approx 0.6976744186$$

2. Recall

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{60}{60 + 27} = \frac{60}{87} \approx 0.6896551724$$

3. F1-score

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

Substitusi angka:

$$F1 = 2 \times \frac{0.6976744186 \times 0.6896551724}{0.6976744186 + 0.6896551724}$$

$$F1 = 2 \times \frac{0.4809}{1.3873} = \frac{120}{173} \approx 0.6936416185$$

Gambar 4. Perhitungan RandomForest

1. Precision

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{70}{70 + 8} = \frac{70}{78} \approx 0.8974358974358975$$

2. Recall

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{70}{70 + 17} = \frac{70}{87} \approx 0.8045977011494253$$

3. F1-score

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

Masukkan nilai:

$$F1 = 2 \times \frac{0.8974358974358975 \times 0.8045977011494253}{0.8974358974358975 + 0.8045977011494253} \approx 0.84848484848485$$

Gambar 5. Perhitungan XGBoost

2) Hitung metrik (kelas churn = positif)

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{72}{72 + 8} = \frac{72}{80} = 0.9$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{72}{72 + 15} = \frac{72}{87} = \frac{24}{29} \approx 0.8275862069$$

F1-score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.9 \times \frac{24}{29}}{0.9 + \frac{24}{29}}$$

Sederhanakan → hasilnya

$$F1 = \frac{144}{167} \approx 0.8622754491017964$$

Gambar 6. Perhitungan LightGBM

3) Hitung manual F1

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{8}{8 + 5} = \frac{8}{13} \approx 0.615384615$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{8}{8 + 79} = \frac{8}{87} \approx 0.091954023$$

F1:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.615384615 \times 0.091954023}{0.615384615 + 0.091954023}$$

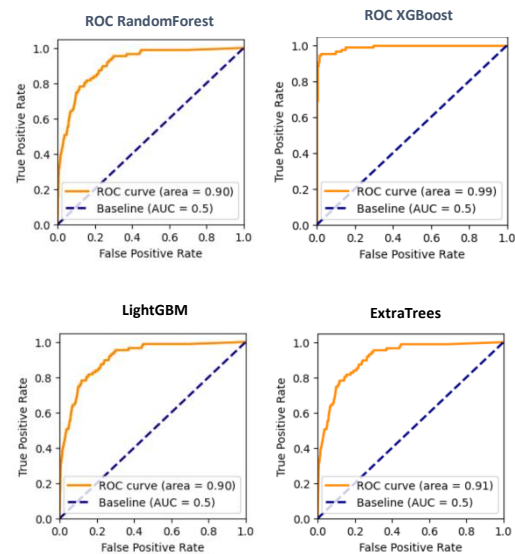
Hitung:

- Numerator = 0.0566
- Denominator = 0.7073
- Fraction = 0.0800

$$F1 \approx 0.16$$

Gambar 7. Perhitungan ExtraTrees

b) ROC (Receiver Operating Characteristic) & AUC (Area Under Curve) - Kurva ROC adalah grafik yang memvisualisasikan performa model pada semua ambang batas klasifikasi. Sesuai Gambar-8 AUC adalah area di bawah kurva ROC. Nilai AUC juga berkisar antara 0 hingga 1. Semakin tinggi nilai AUC (mendekati 1), semakin baik kemampuan model dalam membedakan antara nasabah yang akan churn dan yang tidak.



Gambar 8. ROC/AUC Chart

Tabel 5. Final Score Comparison

Algorithm	Precision	Recall	F1-Score	ROC Area
RandomForest	0,6977	0,6897	0,6936	0,9000
XGBoost	0,8974	0,8046	0,8485	0,9900
LightGBM	0,9000	0,8276	0,8623	0,9000
ExtraTrees	0,6154	0,0920	0,1600	0,9100

Berdasarkan evaluasi performa model pada Tabel-4, algoritma *XGBoost* menunjukkan kinerja terbaik dalam memprediksi customer churn dengan nilai ROC Area yang superior sebesar 0.9900. Meskipun *LightGBM* juga memiliki presisi yang sangat tinggi, keunggulan *XGBoost* pada metrik ROC Area mengindikasikan kemampuannya yang luar biasa dalam membedakan antara nasabah yang

Setelah mendapatkan nilai **Precision** dan **Recall**, maka **F1 Score** dapat dihitung dengan rumus berikut:

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

berpotensi churn dan tidak. Sebaliknya, model *ExtraTrees* memiliki performa terlemah, ditandai dengan F1-Score yang rendah dan Recall yang sangat minim, menunjukkan bahwa model ini tidak efektif dalam mengidentifikasi nasabah yang akan churn. Hasil ini mengukuhkan *XGBoost* sebagai pilihan model yang paling optimal untuk implementasi dalam sistem prediksi churn di sektor perbankan.

5. Deployment - Pada tahap deployment, model prediksi *Churn* yang telah lolos validasi diimplementasikan ke dalam lingkungan produksi menggunakan arsitektur layanan berbasis API, sehingga dapat diintegrasikan dengan sistem CRM atau data warehouse perusahaan. Proses deployment mencakup pembuatan *pipeline* otomatis yang mengekstrak data nasabah terbaru, melakukan preprocessing sesuai skema yang digunakan saat pelatihan, kemudian mengirimkan data tersebut ke model untuk memperoleh skor probabilitas *Churn*. Hasil prediksi disimpan ke basis data operasional dan diakses oleh tim retensi melalui dashboard analitik.

6. Monitoring and Maintenance - Tahap ini bertujuan memastikan model prediksi *Churn* tetap akurat dan relevan seiring berjalannya waktu. Pemantauan dilakukan secara berkala terhadap metrik kinerja utama seperti *Area Under Curve* (AUC), *F1-score*, dan akurasi prediksi pada data terbaru. Selain itu, dilakukan deteksi *data drift* dan *concept drift* untuk mengidentifikasi perubahan distribusi data atau perilaku pelanggan yang dapat memengaruhi performa model. Proses monitoring dapat diotomatisasi menggunakan *model performance dashboard* yang terintegrasi dengan pipeline data produksi. Jika terdeteksi penurunan kinerja di bawah ambang batas yang telah ditentukan, prosedur *model retraining* diaktifkan. *Retraining* dilakukan dengan memanfaatkan kombinasi data historis dan data terbaru, diikuti evaluasi ulang untuk memastikan model yang diperbarui memberikan peningkatan kinerja sebelum diganti di lingkungan produksi.

Integrasi AI dan Aplikasi CRM

1) Persiapan Arsitektur Data di Zoho CRM

- Tahap ini berfokus pada penyesuaian struktur data Zoho CRM. memanfaatkan *Custom field*, menjadi field *Indeks Churn*, dibuat di modul *Contacts* dengan tipe data persentase. Selanjutnya, **API Zoho CRM** diaktifkan dan dikonfigurasi untuk memungkinkan akses dan transfer data dari aplikasi eksternal. **2) Integrasi Model AI dan Transfer Data** - Model AI akan memproses data historis nasabah dan menghasilkan **nilai prediksi Churn**. Nilai persentase ini kemudian dikirim secara terprogram ke Zoho CRM melalui **API**, memperbarui bidang *Indeks Churn* pada profil nasabah yang sesuai. Proses ini memastikan data prediksi yang dihasilkan oleh model AI terintegrasi secara *real-time* dengan data pelanggan di CRM. **3) Otomatisasi Alur Kerja Mitigasi Churn** - Zoho CRM dimanfaatkan untuk mengotomatisasi respons berbasis data prediksi. *workflow rules* dibuat dengan *trigger* berupa pembaruan pada field *Indeks Churn*. *Rules* ini juga akan mengelompokkan nasabah menjadi tiga segmen risiko, seperti **Hi-Risk (> 70%)**, **Mid-Risk (40%-70%)** dan **Lo-Risk (< 40%)**. Berdasarkan segmen tersebut, sistem akan memicu **tindakan otomatis** seperti: a) **Pengiriman email promosi** yang dipersonalisasi. b) **Penugasan tugas** spesifik kepada *relationship manager* untuk intervensi langsung. c) **Notifikasi internal** ke tim terkait.

Melalui metodologi ini, prediksi *Churn* diubah menjadi **tindakan proaktif** yang terotomatisasi, memungkinkan bank menargetkan nasabah berisiko tinggi dengan penawaran khusus secara efisien.

A/B-Test untuk Validasi Efektivitas Intervensi

Untuk mengukur efektivitas intervensi promosi dalam mencegah *Churn*, penelitian ini menggunakan metode A/B-Test yang dirancang secara sistematis. Berdasarkan model prediktif, sebanyak 46 nasabah dengan indeks risiko churn tinggi (>70%) dipilih sebagai subjek eksperimen. Populasi ini kemudian dibagi secara acak menjadi dua kelompok yang setara: **Churn Test Group** (CTG) yang menerima intervensi promosi dan **Churn**

Control Group (CCG) yang tidak menerima intervensi. Setelah periode 30 hari, tingkat *Churn* aktual pada kedua kelompok dicatat, seperti yang ditunjukkan pada tabel berikut.

Tabel 6. Churn Prediction vs Actual

Group	Churn Prediction	Actual Churn	Percent (%)
CTG	23	5	21.74%
CCG	23	19	82.60%

Perhitungan Metrik & Uji Statistik

Untuk mengukur dampak praktis dari intervensi, kami menghitung tingkat keberhasilan (sering disebut sebagai reduksi relatif risiko) dengan membandingkan tingkat churn antara kedua kelompok. Tingkat Keberhasilan (*Success Rate*) adalah:

$$\begin{aligned}
 SR &= \frac{(CCG - CTG)}{CCG} \times 100 \\
 &= \frac{(82.60 - 21.74)}{82.60} \times 100 \\
 &= 73.68 \%
 \end{aligned}$$

Berdasarkan data yang ada, perhitungan menunjukkan tingkat keberhasilan **73,68%**, sebuah angka yang mencerminkan penurunan dari **82,6%** menjadi **21,7%**. Mengingat ukuran sampel yang kecil dan data kategoris, **Fisher's Exact Test** dipilih sebagai metode uji utama (Sprenst, 2025). Tes ini menghitung probabilitas eksak dari hasil yang diamati jika tidak ada efek dari intervensi (hipotesis nol). Hasil pengujian menunjukkan nilai-p (*p-value*) yang sangat kecil, yaitu **< 0.0001**. Nilai ini menunjukkan bahwa probabilitas untuk mengamati perbedaan sebesar itu secara kebetulan sangatlah kecil, memungkinkan kami untuk menolak hipotesis nol. Lebih lanjut, analisis statistik juga mengungkapkan **Odds Ratio** (OR) sebesar 17,1, dengan interval kepercayaan 95% antara [3.95, 73.95]. Angka ini mengindikasikan bahwa kemungkinan nasabah *Churn* di kelompok control adalah 17,1 kali lebih tinggi dibanding di kelompok *treatment*. Karena interval

kepercayaan tidak mencakup angka 1, hal ini semakin memperkuat kesimpulan bahwa efek intervensi adalah signifikan secara statistik dan substansial.

4. Kesimpulan

Penelitian ini berhasil menunjukkan bahwa integrasi antara sistem *Churn Prediction* berbasis Machine Learning dengan otomasi CRM (*Customer Relationship Management*) secara signifikan dapat meningkatkan efektivitas retensi nasabah dalam industri perbankan. Tujuan penelitian untuk mengukur dampak implementasi sistem terpadu ini dapat dijawab dengan hasil A/B Test yang menunjukkan penurunan tingkat *Churn* yang substansial pada kelompok nasabah yang diberikan intervensi proaktif. Temuan ini menyajikan kontribusi baru dalam ranah ilmu pengetahuan, khususnya dalam bidang sains dan teknologi, dengan mengkonfirmasi bahwa penerapan kecerdasan buatan tidak hanya sebatas analisis prediktif, tetapi juga mampu menggerakkan tindakan bisnis yang terotomatisasi dan personal, menghasilkan dampak kuantitatif yang jelas. Hasil penelitian menunjukkan bahwa tingkat keberhasilan intervensi proaktif mencapai 73,68%, sebuah peningkatan signifikan dibandingkan tidak adanya intervensi. Angka ini secara kuat memperkuat pentingnya pendekatan proaktif yang didorong oleh data, alih-alih pendekatan reaktif, sebagai strategi fundamental untuk mengatasi masalah churn yang terus meningkat di tengah persaingan ketat dengan bank digital.

Daftar Pustaka

- A.lemobayo, J., Durodola, O., Alade, O., Awotunde, O. J., Olanrewaju, A. T., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., & Ifeanyi, A. (2024). Hyperparameter tuning in machine learning: a comprehensive review. *Journal of Engineering Research and Reports*, 26(6), 388–395.
- Azeem, M., Aziz, S., Shahid, J., Hayat, A., Ahmed, M., & Khan, M. I. (2022). The impact of customer-centric business strategies on customer relationship management: pharmaceutical field-

- force perspective. *International Journal of Pharmaceutical and Healthcare Marketing*, 16(1), 138–154.
- Brito, J. B. G., Bucco, G. B., Heldt, R., Becker, J. L., Silveira, C. S., Luce, F. B., & Anzanello, M. J. (2024).** A framework to improve churn prediction performance in retail banking. *Financial Innovation*, 10(1), 17.
- Domnik, J., & Holland, A. (2022).** On data leakage prevention and machine learning. 35th Bled EConference Digital Restructuring and Human (Re) Action, 695.
- Handijono, A. (2022).** Memanfaatkan Data Mining dan Knowledge Management System Untuk Mengoptimalkan Strategi Pemasaran pada Bank. *Jurnal Ilmiah Akuntansi Universitas Pamulang*, 5(1), 268517.
- Handijono, A., & Suhatman, Z. (2024).** Optimisasi Hybrid Recommendation System dengan Clustering menggunakan Machine Learning. *Sainstech: Jurnal Penelitian Dan Pengkajian Sains Dan Teknologi*, 34(4), 1–11.
- Ijomah, T. I., Nwabekee, U. S., Agu, E. E., & Abdul-Azeez, O. Y. (2024).** The impact of customer relationship management (CRM) tools on sales growth and customer loyalty in emerging markets. *International Journal of Management & Entrepreneurship Research*, 6(9), 2964–2988.
- Johrawanshi, A., Verma, P., Vyas, H., Gupta, M., Kumar, S., & Pandey, A. (2024).** Optimizing Customer Relationship Management (CRM) Based on Machine Learning-Based Churn Prediction Models. 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS), 1–6.
- Katya, E. (2023).** Exploring feature engineering strategies for improving predictive models in data science. *Research Journal of Computer Systems and Engineering*, 4(2), 201–215.
- Mayasari, S., & Mahadi, T. (2023, April 5).** Bank digital tawarkan bunga tinggi hingga 9%, OJK angkat bicara. <https://Keuangan.Kontan.Co.Id/News/Bank-Digital-Tawarkan-Bunga-Tinggi-Hingga-9-Ojk-Angkat-Bicara>.
- Mqadi, N., Naicker, N., & Adeliyi, T. (2021).** A SMOTe based oversampling data-point approach to solving the credit card data imbalance problem in financial fraud detection. *International Journal of Computing and Digital Systems*, 10(1), 277–286.
- Otiso, K. N. (2024).** Impact Of Network Quality On Customer Retention; Case Study Of Selected Universities In Kenya, A Customer Relationship Management Approach. *African Journal of Emerging Issues*, 6(20), 41–50.
- Owoade, A. A. (2025).** Development Of Customer Segmentation System Using Supervised And Unsupervised Machine Learning Algorithms. *Science World Journal*, 20(2), 554–565.
- Patil, S., & Mohammad, A. S. (2023).** Proactive CRM: Predicting Customer Behavior And Churn Using Machine Learning Models.
- Robo, S., Melani, P. I., Fernetnyanan, P., Widianoro, M. R., & Bariyyah, S. K. (2024).** The Application of Customers Segmentation Using RFM Analysis Method and K-Means Clustering to Improve Marketing Strategy. *IJISTECH (International Journal of Information System and Technology)*, 8(3), 200–211.
- Setiawan, H. B., & Najicha, F. U. (2022).** Perlindungan data pribadi warga negara Indonesia terkait dengan kebocoran data. *Jurnal Kewarganegaraan*, 6(1), 976–982.
- Shabdar, A. (2017).** Mastering Zoho CRM: Manage Your Team, Pipeline, and Clients Effectively. Apress.
- Sprent, P. (2025).** Fisher Exact Test, The. In *International encyclopedia of statistical science* (pp. 961–962). Springer.
- Tuguinay, J., Prentice, C., Moyle, B., Vada, S., & Weaven, S. (2024).** A journey from customer acquisition to retention: An integrative model for guiding future gaming marketing research. *Cornell Hospitality Quarterly*, 65(3), 335–353.

- Văduva, A.-G., Oprea, S.-V., Niculae, A.-M., Bâra, A., & Andreescu, A.-I. (2024).** Improving churn detection in the banking sector: a machine learning approach with probability calibration techniques. *Electronics*, 13(22), 4527.
- Verma, P. (2020).** Churn prediction for savings bank customers: A machine learning approach. *Journal of Statistics Applications & Probability*, 9(3), 535–547.
- Zarkesh, B. (2023).** Exploring the Impact of AI-Driven Pricing on Customer Loyalty and Churn Rates in the Banking Industry