PREDIKSI CUSTOMER CHURN PADA INDUSTRI TELEKOMUNIKASI UNTUK MENDUKUNG STRATEGI RETENSI PELANGGAN MENGGUNAKAN ALGORITMA XGBOOST

SKRIPSI

DISUSUN OLEH

AIDA FADHILA

2209020161

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2026

PREDIKSI CUSTOMER CHURN PADA INDUSTRI TELEKOMUNIKASI UNTUK MENDUKUNG STRATEGI RETENSI PELANGGAN MENGGUNAKAN ALGORITMA XGBOOST

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Tenologi Infromasi. pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

AIDA FADHILA
2209020161

PROGRAM STUDI TEKNOLOGI INFROMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
202

LEMBAR PENGESAHAN

Judul Skripsi	:	PREDIKSI CUSTOMER CHURN PADA INDUSTRI TELEKOMUNIKASI UNTUK MENDUKUNG STRATEGI RETENSI PELANGGAN MENGGUNAKAN ALGORITMA XGBOOST
Nama Mahasiswa	:	AIDA FADHILA
NPM	:	2209020161
Program Studi	:	TEKNOLOGI INFORMASI

Menyetujui
Komisi Pembimbing

(Halim Maulana, S.T., M.Kom)
NIDN. 0121119102

Ketua Program Studi
Teknologi Informasi

(Fatma Sari Hutagalung, S.Kom, M.Kom)
NIDN. 0117019301

Dekan

(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

PREDIKSI CUSTOMER CHURN PADA INDUSTRI TELEKOMUNIKASI UNTUK MENDUKUNG STRATEGI RETENSI PELANGGAN MENGGUNAKAN ALGORITMA XGBOOST

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 10 Februari 2026

Yang membuat pernyataan

Aida Fadhila

NPM. 2209020161

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Aida Fadhila
NPM	: 220920161
Program Studi	: Teknologi Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif *(Non-Exclusive Royalty free Right)* atas penelitian skripsi saya yang berjudul:

PREDIKSI CUSTOMER CHURN PADA INDUSTRI TELEKOMUNIKASI UNTUK MENDUKUNG STRATEGI RETENSI PELANGGAN MENGGUNAKAN ALGORITMA XGBOOST

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 10 Februari 2026

Yang membuat pernyataan

Aida Fadhila

2209020161

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Aida Fadhila
Tempat dan Tanggal Lahir	: Medan, 06 April 2004
Alamat Rumah	: Jl. Sempurna Dusun III Melur
Telepon/Faks/HP	: 0821-6556-1860
E-mail	: aidafadhila87@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: Madrasah Ibtidaiyah Bidayatul Hidayah	TAMAT: 2016
SMP	: SMP Negeri 29 Medan	TAMAT: 2019
SMA	: SMK Tritech Informatika Medan	TAMAT: 2022

KATA PENGANTAR

Pendahuluan

Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Assoc. Prof. Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom Ketua Program Studi Teknologi Informasi
4. Bapak Okvi Nugroho, S.Kom., M.Kom Sekretaris Program Studi Teknologi Informasi
5. Pembimbing Bapak Halim Maulana, S.T., M.Kom yang telah meluangkan waktu untuk memberikan arahan, bimbingan, serta ilmu yang sangat berharga. Terima kasih atas kesabaran dan pelajaran hidup yang telah Bapak berikan kepada penulis.
6. Kepada kedua orang tua tercinta, Terima kasih atas doa yang tidak pernah putus, pengorbanan, serta kepercayaan penuh yang diberikan kepada penulis sebagai satu-satunya anak dalam keluarga yang menempuh pendidikan tinggi. Skripsi ini penulis persembahkan sebagai wujud bakti atas segala dukungan kalian.
7. Kakak-kakak kandung penulis, Rayani Aulia dan Dini Aisyah, serta adik tersayang, Luffy Aswan, yang senantiasa memberikan semangat serta doa yang tulus demi kelancaran dan kesuksesan penulis dalam menyelesaikan studi ini.
8. Teman-teman seperjuangan, Dinda Anantya, Dwi Nisyatul Wardah, Farhan Fanalty, Nadya Aulya Putri, Regina Intisya, Thasyah Rezky dan Eka Syarif Maulana. Terima kasih telah menjadi sahabat yang luar biasa, membantu

penulis selama masa perkuliahan, dan saling menguatkan dalam penyusunan skripsi ini. Juga kepada seluruh keluarga besar D1 atas momen-momen indah yang telah kita lalui bersama.

9. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

PREDIKSI CUSTOMER CHURN PADA INDUSTRI TELEKOMUNIKASI UNTUK MENDUKUNG STRATEGI RETENSI PELANGGAN MENGGUNAKAN ALGORITMA XGBOOST

ABSTRAK

Industri telekomunikasi menghadapi tantangan besar terkait *customer churn*, di mana biaya akuisisi pelanggan baru jauh lebih tinggi dibandingkan biaya mempertahankan pelanggan yang sudah ada. Penelitian ini bertujuan untuk membangun model prediksi *customer churn* yang optimal menggunakan algoritma *XGBoost* untuk mendukung strategi retensi pelanggan. Metodologi yang digunakan adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dengan memanfaatkan dataset sekunder dari Kaggle yang mencakup 7.043 data pelanggan. Tantangan ketidakseimbangan data diatasi dengan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih untuk meningkatkan kemampuan deteksi kelas minoritas. Hasil evaluasi menggunakan *Stratified 5-Fold Cross Validation* menunjukkan performa model yang sangat handal dengan tingkat akurasi mencapai 87,4% dan nilai *recall* sebesar 82,1%. Model ini kemudian diimplementasikan ke dalam sistem informasi berbasis web menggunakan *framework* FastAPI dan Neon Database. Sistem ini mampu melakukan klasifikasi tingkat risiko pelanggan (*Low, Medium, High*) secara *real-time* serta mengintegrasikan teknologi *Generative AI* (Gemini API) untuk memberikan rekomendasi strategi retensi yang dipersonalisasi dan komunikatif bagi agen layanan pelanggan .

Kata Kunci: *Customer Churn*, FastAPI, *Machine learning,* SMOTE, Strategi Retensi, XGBoost.

CUSTOMER CHURN PREDICTION IN THE TELECOMMUNICATIONS INDUSTRY TO SUPPORT CUSTOMER RETENTION STRATEGIES USING THE XGBOOST ALGORITHM

ABSTRACT

The telecommunications industry faces significant challenges regarding customer churn, as the cost of acquiring new customers is substantially higher than retaining existing ones. This research aims to build an optimal customer churn prediction model using the XGBoost algorithm to support customer retention strategies. The methodology employed is the Cross-Industry Standard Process for Data Mining (CRISP-DM), utilizing a secondary dataset from Kaggle comprising 7,043 customer records. The data imbalance challenge was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) on training data to enhance minority class detection. Evaluation results using Stratified 5-Fold Cross Validation demonstrate highly reliable model performance, achieving an accuracy of 87.4% and a recall value of 82.1%. This model was subsequently implemented into a web-based information system using the FastAPI framework and Neon Database. The system is capable of performing real-time customer risk classification (Low, Medium, High) and integrates Generative AI technology (Gemini API) to provide personalized and communicative retention strategy recommendations for customer service agents.

Keywords: *Customer Churn, FastAPI, Machine learning, Retention Strategy, SMOTE, XGBoost.*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1.Latar Belakang Masalah

Industri telekomunikasi adalah sektor yang menyediakan layanan pengiriman dan pertukaran informasi jarak jauh—termasuk suara, data, dan multimedia—melalui berbagai infrastruktur jaringan seperti jaringan seluler dan akses broadband (Badan Pusat Statistik, 2024). Dalam beberapa tahun terakhir, peningkatan signifikan pada penggunaan internet dan layanan data di Indonesia telah memperkuat dinamika persaingan antar-penyelenggara jasa telekomunikasi, terlihat dari kenaikan penetrasi pengguna internet dan pertumbuhan lalu lintas data (APJII, 2024). Pertumbuhan jumlah pelanggan layanan data yang disertai dengan semakin beragamnya pilihan layanan dan operator menyebabkan pelanggan memiliki fleksibilitas yang lebih tinggi untuk berpindah penyedia layanan (Kotler & Keller 2016).

Berdasarkan laporan tahunan beberapa operator telekomunikasi global, tingkat churn pelanggan seluler berada pada kisaran dua digit per tahun, dengan angka yang cenderung lebih tinggi pada segmen prabayar dibandingkan pascabayar (Vodafone, 2023; AT&T, 2023). Penelitian dan tinjauan literatur terbaru menunjukkan bahwa churn pelanggan merupakan tantangan utama dalam industri telekomunikasi karena berdampak langsung terhadap penurunan pendapatan yang berulang serta meningkatnya kebutuhan untuk memprediksi dan mempertahankan pelanggan melalui strategi retensi (Shaikhsurab & Magadum, 2017) Oleh karena itu, manajemen churn dipandang sebagai isu strategis yang krusial bagi keberlanjutan

bisnis perusahaan telekomunikasi karena berkaitan langsung dengan profitabilitas dan daya saing jangka panjang perusahaan.

Dalam praktiknya, perusahaan telekomunikasi menghadapi tantangan dalam mengidentifikasi pelanggan yang berpotensi melakukan *churn* secara dini karena tingginya volume data dan kompleksitas perilaku pelanggan (Shaikhsurab & Magadum, 2017). Kompleksitas tersebut dipengaruhi oleh karakteristik data yang heterogen, ketidakseimbangan kelas (*class imbalance*), serta hubungan antar variabel yang bersifat *non-linear* sehingga sulit dianalisis menggunakan model linear konvensional (Wakhidah et al., 2025). Studi terbaru menunjukkan bahwa pola *churn* pelanggan terbentuk dari interaksi berbagai faktor seperti penggunaan layanan, jenis kontrak, serta perilaku konsumsi yang saling memengaruhi, sehingga pendekatan berbasis *machine learning* diperlukan untuk menangkap pola hubungan yang kompleks tersebut secara lebih akurat (Chang et al., 2024).

Pendekatan analisis konvensional yang berbasis asumsi linearitas dan distribusi statistik tertentu memiliki keterbatasan dalam memodelkan hubungan yang kompleks dan *non-linear* pada data pelanggan telekomunikasi (Shaikhsurab & Magadum, 2017). Model statistik klasik cenderung kurang adaptif terhadap karakteristik data dunia nyata yang dinamis, heterogen, dan berskala besar, sehingga kurang mampu menangkap interaksi fitur yang kompleks (Wakhidah et al., 2025). Selain itu, pada data dengan ketidakseimbangan kelas (*class imbalance*), pendekatan konvensional berpotensi menghasilkan model yang bias terhadap kelas mayoritas dan memiliki kemampuan prediksi yang rendah terhadap pelanggan berisiko *churn* (Imani et al., 2025). Oleh karena itu, pendekatan berbasis *machine*

learning yang lebih fleksibel dan *non-linear* dinilai lebih efektif dalam menangani kompleksitas data *churn* pelanggan.

Data yang digunakan dalam penelitian ini merupakan dataset pelanggan telekomunikasi yang bersifat global dan diperolah dari platform *Kaggle* juga banyak digunakan dalam penelitian prediksi *churn* berbasis *machine learning*. Meskipun dataset tersebut berskala global, pola *churn* yang dianalisis tetap relevan dengan konteks industri telekomunikasi di Indonesia karena karakteristik layanan seluler, model prabayar dan pascabayar, serta sistem berbasis penggunaan data memiliki kesamaan dengan praktik industri global (GSMA, 2023). Selain itu, laporan industri menunjukkan bahwa peningkatan penetrasi layanan data dan tingginya tingkat persaingan antar operator merupakan fenomena global yang juga terjadi di berbagai negara berkembang, termasuk Indonesia, sehingga analisis *churn* berbasis data global tetap memiliki relevansi kontekstual (GSMA, 2023).

Seiring dengan perkembangan teknologi dan meningkatnya kompleksitas data pelanggan, pendekatan *machine learning* menjadi semakin relevan karena mampu mempelajari pola yang kompleks dan *non-linear* dari data historis secara lebih efektif dibandingkan metode analisis konvensional. Namun demikian, beberapa penelitian terdahulu menunjukkan bahwa algoritma klasifikasi konvensional seperti *Support Vector Machine* (SVM), *Random Forest*, dan *Logistic Regression* masih memiliki performa prediksi *churn* yang bervariasi dan relatif lebih rendah dibandingkan pendekatan *boosting* yang lebih canggih. Dalam studi penerapan SVM, *Random Forest*, dan *Logistic Regression* pada dataset *churn* telekomunikasi, hasil akurasi tertinggi hanya mencapai sekitar 79% untuk *Logistic Regression* dan 76% *untuk Random Forest*, sedangkan SVM menunjukkan akurasi lebih rendah

lagi sekitar 74% (Nurtriana et al., 2024). Review komprehensif juga melaporkan bahwa *Random Forest* dan SVM berkinerja moderat dengan akurasi di bawah 90% pada berbagai dataset *churn*, sementara model sederhana seperti *Logistic Regression* cenderung kurang stabil ketika diterapkan pada data yang tidak seimbang (Dhangar & Anand, 2021).

Berdasarkan keterbatasan tersebut, penelitian ini memilih algoritma *XGBoost* karena kemampuannya dalam memodelkan hubungan yang kompleks dan *non-linear* melalui pendekatan *gradient boosting* yang dilakukan secara iteratif untuk meminimalkan kesalahan prediksi (Chen & Guestrin, 2016) . Selain itu, *XGBoost* memiliki mekanisme regularisasi untuk mengurangi risiko *overfitting*, mendukung penanganan data tidak seimbang melalui pengaturan bobot kelas, serta mampu mengelola *missing value* secara otomatis, sehingga lebih adaptif terhadap karakteristik data *churn* yang heterogen dan dinamis (Shaikhsurab & Magadum, 2017).Berdasarkan keunggulan tersebut, *XGBoost* dipilih dalam penelitian ini untuk menangani data yang tidak seimbang dan menangkap hubungan *non-linear* secara akurat. Penelitian ini tidak hanya membangun model prediksi, tetapi juga mengimplementasikannya ke dalam sistem berbasis web menggunakan FastAPI dan Neon *Database* untuk memberikan rekomendasi strategi retensi pelanggan secara otomatis dan mengintegrasikan melalui mekanisme sinkronisasi data dan model.

1.2.Rumusan Masalah

Berdasarkan latar belakang diatas, penelitian ini memiliki rumusan masalah yang telah dikumpulkan, yaitu:

1. Bagaimana membangun model prediksi *customer churn* yang optimal pada industri telekomunikasi menggunakan algoritma *XGBoost* dengan penanganan ketidakseimbangan data melalui metode *Synthetic Minority Over-sampling Technique* (SMOTE)?
2. Bagaimana tingkat akurasi dan kinerja model *XGBoost* dalam memprediksi *customer churn* jika diukur menggunakan metrik evaluasi klasifikasi?
3. Bagaimana mengimplementasikan model prediksi tersebut ke dalam sistem informasi berbasis web menggunakan *framework* FastAPI yang mampu memberikan rekomendasi strategi retensi pelanggan secara otomatis dan mengintegrasikan melalui mekanisme sinkronisasi data dan model?

1.3. Batasan Masalah

Untuk menjaga fokus dan kejelasan penelitian, ada batasan masalah dari penelitian yaitu:

1. Penelitian ini dibatasi pada penggunaan data sekunder pelanggan telekomunikasi yang diperoleh dari platform Kaggle, sehingga tidak mempertimbangkan faktor eksternal di luar variabel dataset seperti dinamika kondisi pasar, pengaruh promosi dari pihak kompetitor, maupun data kepuasan pelanggan yang bersifat kualitatif.
2. Penelitian ini hanya menggunakan satu algoritma *machine learning*, yaitu XGBoost, tanpa melakukan perbandingan kinerja dengan algoritma klasifikasi lainnya.
3. Sistem informasi berbasis web yang dibangun terbatas pada penyajian hasil prediksi, analisis faktor *churn*, dan pemberian rekomendasi strategi retensi secara tekstual, namun tidak mencakup eksekusi kebijakan bisnis otomatis

seperti pengiriman notifikasi/email penawaran langsung ke pelanggan maupun integrasi sistem secara *live* dengan basis data operasional perusahaan telekomunikasi.

1.4 Tujuan Penelitian

Adapun tujuan penelitian yang selaras dengan rumusan masalah penelitian ini adalah:

1. Membangun model prediksi customer *churn* pada industri telekomunikasi menggunakan algoritma *XGBoost* berdasarkan data pelanggan yang tersedia.
2. Mengevaluasi tingkat akurasi dan kinerja model prediksi customer *churn* dengan membandingkan hasil prediksi terhadap kondisi aktual pelanggan menggunakan metrik evaluasi klasifikasi.
3. Memanfaatkan hasil prediksi customer *churn* serta analisis faktor yang berpengaruh sebagai dasar pendukung dalam perumusan strategi tindak lanjut untuk meningkatkan retensi pelanggan pada industri telekomunikasi.

1.5. Manfaat Penelitian

1. Manfaat Akademis

Penelitian ini diharapkan dapat menjadi referensi tambahan dalam pengembangan ilmu *data science* dan sistem informasi, khususnya mengenai penerapan algoritma *XGBoost* untuk memodelkan perilaku pelanggan. Selain itu, studi ini memberikan gambaran teknis mengenai efektivitas metode SMOTE dalam menangani ketidakseimbangan data serta penggunaan *framework* FastAPI dalam membangun aplikasi web prediktif yang modern.

2. Manfaat Bagi Mitra

Bagi perusahaan telekomunikasi, sistem ini dapat menjadi alat bantu strategis untuk mendeteksi pelanggan yang berisiko berhenti berlangganan secara lebih akurat dan cepat. Dengan adanya *Dashboard* web, perusahaan dapat memahami faktor utama penyebab pelanggan pindah dan langsung mendapatkan rekomendasi strategi retensi yang dipersonalisasi, sehingga perusahaan dapat menekan biaya akuisisi pelanggan baru dan menjaga profitabilitas jangka panjang.

3. Manfaat Bagi Peneliti

Bagi peneliti, proses pengembangan skripsi ini bermanfaat untuk mengasah keterampilan teknis dalam mengolah data kompleks, membangun model *machine learning* yang optimal, hingga mengimplementasikannya ke dalam aplikasi web fungsional.

BAB II

LANDASAN TEORI

2.1. Customer Hubungan Pelanggan (CRM)

Customer Relationship Management (CRM) merupakan strategi bisnis yang berfokus pada pengelolaan hubungan jangka panjang antara perusahaan dan pelanggan melalui pemanfaatan data pelanggan secara terintegrasi. CRM tidak hanya dipahami sebagai sistem atau perangkat lunak, melainkan sebagai pendekatan strategis yang bertujuan meningkatkan kepuasan, loyalitas, dan nilai pelanggan bagi perusahaan dengan memanfaatkan analisis data dan proses bisnis yang terkoordinasi (Bisht & Anjaria, 2025) Melalui CRM, perusahaan dapat mengelola interaksi pelanggan secara sistematis serta memahami kebutuhan dan perilaku pelanggan secara lebih mendalam untuk mendukung pengambilan keputusan bisnis yang berbasis data.

2.1.1. Siklus Hidup Pelanggan (*Customer Lifecycle*)

Siklus hidup pelanggan (*customer lifecycle*) mencakup beberapa tahapan utama, yaitu akuisisi (*acquisition*), pengembangan (*development*), dan retensi (*retention*). Dalam industri telekomunikasi yang memiliki tingkat persaingan tinggi, fase retensi menjadi fokus utama karena mempertahankan pelanggan lama terbukti lebih menguntungkan dibandingkan memperoleh pelanggan baru. Biaya akuisisi pelanggan baru dilaporkan dapat mencapai beberapa kali lipat dibandingkan biaya mempertahankan pelanggan yang sudah ada (Wakhidah et al., 2025).

Penelitian empiris menunjukkan bahwa pelanggan yang telah berlangganan dalam jangka waktu lama cenderung memiliki kontribusi pendapatan yang

lebih stabil dan biaya pelayanan yang lebih rendah. Oleh karena itu, perusahaan telekomunikasi semakin menekankan strategi berbasis data untuk mempertahankan pelanggan yang sudah ada sebagai upaya meningkatkan profitabilitas jangka panjang.

2.1.2. Strategi Retensi Pelanggan Berbasis Data

Strategi retensi pelanggan bertujuan untuk mencegah pelanggan menghentikan layanan dengan cara mengidentifikasi pelanggan berisiko tinggi dan memberikan intervensi yang tepat. Pemanfaatan data pelanggan memungkinkan perusahaan mengembangkan strategi retensi yang lebih personal, seperti pemberian diskon, penyesuaian paket layanan, atau peningkatan kualitas layanan bagi pelanggan tertentu.

Model prediksi *churn* berbasis *machine learning* memungkinkan perusahaan mengidentifikasi pelanggan yang berpotensi berhenti berlangganan sebelum *churn* benar-benar terjadi. Dengan informasi tersebut, perusahaan dapat melakukan tindakan preventif secara lebih efektif dan terarah (Hermawan et al., 2024).

2.2 Penelitian Terdahulu

Tabel 2.1 Ringkasan Penelitian Terdahulu

NO	Judul dan Peneliti	Pembahasan	Metode	Kelebihan Dan Kekurangan
1	*Customer Churn Prediction for Telecommunication Companies Using Machine learning and Ensemble Methods –*	*Membandingkan berbagai model ML pada dataset Telco untuk prediksi churn, dengan fokus pada kinerja*	*XGBoost, Random Forest, LightGBM, ANN; cross-validation; tuning parameter.*	Kelebihan: Evaluasi komprehensif, *open-access*, performa *XGBoost* unggul. Kekurangan: Fokus teknis,

NO	Judul dan Peneliti	Pembahasan	Metode	Kelebihan Dan Kekurangan
	Alotaibi & Haq (2024)	*ensemble.*		belum mengaitkan hasil dengan strategi retensi.
2	*Evaluation of Telecommunication Customer Churn Classification with SMOTE Using Random Forest and XGBoost – Wakhidah et al. (2025)*	Mengkaji dampak *SMOTE* terhadap kinerja RF dan *XGBoost* pada data *churn* tidak seimbang.	*Preprocessing, SMOTE, RF, XGBoost; Accuracy, Recall, F1, AUC.*	Kelebihan: Menegaskan pentingnya *recall & SMOTE.* Kekurangan: Dataset tunggal, minim analisis implikasi bisnis.
3	*Research on Telecom Customer Churn Prediction Based on GA-XGBoost and* (Peng & Peng, 2022)	Optimalisasi dan interpretasi faktor *churn* menggunakan SHAP.	*Genetic Algorithm, XGBoost, SHAP*.	Kelebihan: Kelebihan: Interpretabilitas model kuat. Kekurangan: Kompleksitas tinggi, sulit diterapkan sebagai sistem sederhana.
4	*Implementation of Machine learning Models for Predicting Churn Using Decision Tree* – (Asyhari et al., 2025)	Prediksi *churn* telekomunikasi menggunakan *Decision Tree* pada data besar.	*Decision Tree; CRISP-DM*	Kelebihan: Fokus telko dan faktor penentu *churn*. Kekurangan: Tidak menangani imbalance sebaik model *boosting*
5	*Customer Churn Prediction Using Machine learning Models – JERR* (Sam et al., 2024)	Uji berbagai algoritma ML pada dataset *churn*.	*SVM, k-NN, Naïve Bayes, Decision Tree*	Kelebihan: Menguatkan metrik jelas. Kekurangan: Tidak termasuk model ensemble terbaru

Berdasarkan ringkasan penelitian terdahulu pada Tabel 2.1, dapat diidentifikasi adanya kesenjangan penelitian (*research gap*) pada aspek integrasi antara performa teknis model dan implementasi strategis dalam konteks manajerial. Studi yang dilakukan oleh Alotaibi dan Haq (2024) serta Wakhidah et al. (2025) berfokus pada evaluasi performa algoritma *machine learning* seperti *XGBoost* dan Random Forest melalui pengukuran metrik teknis seperti *accuracy, recall, F1-score,* dan *AUC,* namun belum mengaitkan hasil prediksi tersebut dengan langkah retensi pelanggan yang konkret bagi pihak manajemen. Penelitian oleh Peng dan Peng (2022) menekankan optimalisasi model melalui pendekatan GA-*XGBoost* dan interpretabilitas menggunakan SHAP, tetapi belum menyediakan sistem operasional berbasis web yang dapat digunakan secara langsung oleh pengguna non-teknis. Demikian pula, penelitian Asyhari (2025) dan studi komparatif pada JERR (2024) lebih menitikberatkan pada perbandingan performa algoritma klasifikasi seperti *Decision Tree, SVM,* dan *Naïve Bayes* tanpa mengembangkan sistem prediktif terintegrasi yang siap diimplementasikan dalam lingkungan industri.

Kebaruan dalam penelitian ini terletak pada pengembangan aplikasi web terintegrasi berbasis FastAPI dan Neon Database yang tidak hanya melakukan prediksi *churn* menggunakan algoritma *XGBoost* dengan penanganan ketidakseimbangan data melalui SMOTE, tetapi juga secara otomatis menghasilkan strategi retensi preskriptif berbasis AI API untuk mendukung komunikasi agen secara lebih humanis dan terarah. Sistem yang dikembangkan bersifat *cloud-native* serta dilengkapi fitur pelacakan pertumbuhan performa model dan log pelatihan yang transparan, sehingga mampu menjembatani kesenjangan antara pemodelan

akademik dan kebutuhan praktis industri telekomunikasi melalui integrasi data, model, dan rekomendasi keputusan dalam satu platform operasional.

2.3. Customer *Churn* dalam Industri Telekomunikasi

Customer *churn* didefinisikan sebagai kondisi ketika pelanggan memutuskan untuk menghentikan penggunaan layanan atau berpindah ke penyedia layanan lain dalam periode waktu tertentu. Dalam penelitian ini, *churn* direpresentasikan sebagai variabel target (*Churn*) yang bersifat biner, yaitu *Yes* (pelanggan berhenti berlangganan) dan *No* (pelanggan tetap berlangganan). Variabel ini menjadi indikator utama keberhasilan strategi retensi pelanggan dalam industri telekomunikasi (Alotaibi & Haq, 2024).

2.3.1. Jenis Customer *Churn*

Secara umum, *churn* dibedakan menjadi dua jenis utama, yaitu *voluntary churn* dan *involuntary churn*. *Voluntary churn* terjadi ketika pelanggan secara sadar memilih untuk berhenti berlangganan karena faktor ketidakpuasan harga atau kualitas layanan. Sebaliknya, *involuntary churn* terjadi akibat faktor administratif seperti pemutusan layanan karena tunggakan pembayaran. Penelitian ini berfokus pada *voluntary churn* karena jenis ini dapat diprediksi dan dicegah melalui analisis data historis pelanggan yang tepat (Wakhidah et al., 2025).

2.4. *Machine learning* dan Analisis Prediktif

Penerapan analisis prediktif melalui teknologi *machine learning* menjadi instrumen strategis yang krusial bagi industri telekomunikasi dalam mengeksplorasi data historis guna memproyeksikan perilaku pelanggan secara akurat.

2.4.1. Konsep Dasar *Machine learning*

Machine learning merupakan bagian dari kecerdasan buatan yang memungkinkan sistem komputer untuk mempelajari pola dari data tanpa diprogram secara eksplisit. Model *machine learning* dibangun berdasarkan data historis untuk mengenali hubungan antara variabel input dan output sehingga dapat digunakan untuk klasifikasi pada data baru (Hermawan et al. 2024). Penelitian ini dikategorikan ke dalam *supervised learning* karena proses pelatihan model menggunakan data yang telah memiliki label target yang jelas (Wakhidah et al., 2025).

2.4.2. Klasifikasi Biner pada Masalah *Churn*

Permasalahan *churn* termasuk ke dalam kategori klasifikasi biner yang bertujuan mengelompokkan pelanggan ke dalam dua kategori risiko utama (Alotaibi & Haq, 2024). Dalam implementasi sistem ini, model dilatih untuk membedakan antara pelanggan yang berisiko tinggi untuk pindah dan pelanggan yang cenderung loyal.

2.5. Algoritma *XGBoost* (Extreme Gradient *Boosting*)

Gradient boosting merupakan teknik *ensemble learning* yang membangun model prediksi secara bertahap dengan mengombinasikan beberapa model sederhana (*weak learners*), umumnya berupa pohon keputusan, menjadi satu model yang lebih kuat. Proses pembelajaran dilakukan secara iteratif, di mana setiap model baru dilatih untuk memperbaiki kesalahan (*residual error*) yang dihasilkan oleh model sebelumnya (Alotaibi & Haq, 2024).

Pendekatan ini menggunakan prinsip optimasi berbasis gradien, di mana model secara bertahap meminimalkan fungsi kerugian (*loss function*) dengan

menyesuaikan prediksi terhadap arah gradien kesalahan. Dengan cara tersebut, *gradient boosting* mampu menangkap pola *non-linier* dan interaksi kompleks antar fitur yang sering muncul pada data pelanggan telekomunikasi (Hermawan *et al.*, 2024).

Dalam konteks prediksi customer *churn*, *gradient boosting* dinilai efektif karena mampu meningkatkan akurasi prediksi pada data yang kompleks dan heterogen. Beberapa penelitian menunjukkan bahwa algoritma berbasis *gradient boosting*, termasuk *XGBoost*, secara konsisten menghasilkan performa yang lebih baik dibandingkan model klasifikasi tunggal dalam mengidentifikasi pelanggan yang berisiko *churn* (Wakhidah et al., 2025).

2.5.1. Arsitektur dan Fungsi Objektif XGBoost

XGBoost (*Extreme Gradient Boosting*) merupakan algoritma *ensemble learning* berbasis *gradient boosting* yang dirancang untuk menghasilkan model prediksi yang efisien dan akurat. Salah satu keunggulan utama *XGBoost* adalah kemampuannya dalam melakukan *parallel processing*, sehingga proses pelatihan model menjadi lebih cepat dibandingkan metode *boosting* konvensional, terutama pada dataset dengan jumlah fitur yang besar (Alotaibi & Haq, 2024).

Selain itu, *XGBoost* mampu menangani *missing value* secara otomatis dengan mempelajari arah pemisahan (*default direction*) yang optimal pada setiap node pohon keputusan, sehingga tidak selalu memerlukan proses imputasi data secara eksplisit (Hermawan *et al.*, 2025). Keunggulan lain yang membedakan *XGBoost* dari algoritma *boosting* tradisional adalah penerapan *regularization*, yang berfungsi untuk mengendalikan kompleksitas model dan

mengurangi risiko *overfitting*, sehingga model memiliki kemampuan generalisasi yang lebih baik (Alotaibi & Haq, 2024). Secara matematis, fungsi objektif *XGBoost* dirumuskan sebagai berikut:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{m=1}^{m} \Omega(f_m)$$

keterangan masing-masing komponen adalah sebagai berikut:

a. $\mathcal{L}$: Menyatakan nilai fungsi objektif yang diminimalkan selama proses pelatihan model *XGBoost.*

b. n: Jumlah data atau observasi pada dataset pelatihan.

c. y_i: Nilai aktual (label sebenarnya) dari observasi ke-i.

d. $\hat{y}_i$: Nilai prediksi yang dihasilkan oleh model untuk observasi ke-i.

e. $l(y_i, \hat{y}_i)$: Fungsi kerugian (*loss function*) yang mengukur kesalahan prediksi antara nilai aktual dan nilai prediksi. Pada kasus klasifikasi *customer churn*, fungsi ini umumnya berupa *logistic loss*.

f. M: Jumlah pohon keputusan (*decision trees*) yang digunakan dalam model XGBoost.

g. f_m: Pohon keputusan ke-m yang berkontribusi dalam pembentukan prediksi model.

h. $\Omega(f_m)$: Fungsi regularisasi yang memberikan penalti terhadap kompleksitas pohon keputusan ke-m.

dengan fungsi regularisasi:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2$$

keterangan:

$\Omega(f)$: Nilai penalti kompleksitas dari suatu pohon keputusan.

a. T: Jumlah daun (*leaf nodes*) pada pohon keputusan. Semakin besar nilai T, semakin kompleks struktur pohon.

b. w_j: Bobot atau nilai prediksi pada daun ke-j.

c. γ: Parameter regularisasi yang memberikan penalti terhadap jumlah daun, sehingga mendorong model membentuk pohon yang lebih sederhana.

d. λ: Parameter regularisasi L2 yang mengontrol besarnya bobot daun untuk mencegah *overfitting*.

Dimana $l(y_i, \hat{y_i})$merupakan *loss function*, Tadalah jumlah daun pada pohon keputusan, $w_j$adalah bobot daun, sedangkan γdan λmerupakan parameter regularisasi yang bertujuan membatasi kompleksitas model. Mekanisme ini menjadikan *XGBoost* efektif dalam menangani data pelanggan yang kompleks, seperti pada kasus prediksi customer *churn* (Hermawan *et al.*, 2025).

2.5.2. Penerapan *XGBoost* dalam Prediksi *Churn*

Dalam konteks *churn*, algoritma ini dinilai efektif karena mampu menangkap pola *non-linear* dan interaksi kompleks antar fitur layanan pelanggan (Hermawan et al. 2024). Implementasi teknis dalam sistem ini menggunakan *XGBClassifier* yang diintegrasikan ke dalam *pipeline*

pemrosesan data . Model ini juga mengadopsi optimasi parameter melalui metode *RandomizedSearchCV* guna mencapai validitas hasil yang tinggi

2.6. Penanganan *Imbalance Data* dengan SMOTE

Penanganan ketidakseimbangan data merupakan langkah kritis dalam rekayasa fitur untuk memastikan model klasifikasi mampu mengenali pola pelanggan yang berhenti berlangganan secara objektif dan tidak terdistorsi oleh dominasi kelas mayoritas.

2.6.1. Masalah Ketidakseimbangan Kelas

Masalah ketidakseimbangan kelas (*class imbalance*) merupakan tantangan umum dalam prediksi *customer churn*, di mana proporsi pelanggan yang tetap berlangganan jauh lebih dominan dibandingkan pelanggan yang benar-benar melakukan *churn* sehingga data minoritas (*churn*) jauh lebih sedikit daripada data mayoritas (*non-churn*). Kondisi ini berpotensi menyebabkan model prediksi menjadi bias terhadap kelas mayoritas sehingga menghasilkan nilai akurasi yang tinggi tetapi gagal mengenali pola pelanggan yang sebenarnya berada di kelas minoritas, yang justru merupakan target utama prediksi. Ketidakseimbangan kelas yang ekstrim juga dapat memengaruhi metrik evaluasi seperti *recall* dan *F1-score*, di mana meskipun akurasi keseluruhan tampak tinggi, kemampuan model untuk mengidentifikasi pelanggan yang akan *churn* tetap rendah tanpa strategi penanganan *imbalance.* Fenomena ini menimbulkan risiko bahwa strategi retensi yang disusun berdasarkan model yang tidak memadai menjadi kurang efektif karena gagal mendeteksi sinyal-sinyal awal

perpindahan pelanggan yang berada di kelas minoritas tersebut (Imani et al., 2025).

2.6.2. Teknik SMOTE

Untuk mengatasi kendala tersebut, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) yang terbukti efektif dalam menyeimbangkan distribusi data tanpa kehilangan informasi penting . SMOTE bekerja dengan membangkitkan sampel sintetis pada kelas minoritas melalui pemilihan tetangga terdekat dalam ruang fitur, sehingga data baru yang tercipta bukanlah sekadar duplikasi, melainkan variasi baru yang masih memiliki karakteristik serupa (Chawla et al., 2002). Sistem ini menerapkan SMOTE secara otomatis pada tahap pelatihan data latih untuk meningkatkan representasi pelanggan yang *churn* agar model dapat mempelajari batasan keputusan (*decision boundary*) yang lebih adil.

Dalam implementasinya, SMOTE diintegrasikan ke dalam sebuah *Imbalanced-learn Pipeline* yang memastikan proses penyeimbangan hanya terjadi pada dataset pelatihan dan tidak mencemari dataset pengujian . Melalui pendekatan ini, distribusi kelas target pada data latih menjadi seimbang secara otomatis sebelum diproses oleh algoritma XGBoost. Penggunaan SMOTE secara teknis membantu meningkatkan kemampuan model dalam mendeteksi pelanggan berisiko tinggi tanpa mengorbankan performa klasifikasi secara keseluruhan (Wakhidah et al., 2025). Dengan demikian, hasil prediksi yang dihasilkan melalui antarmuka web menjadi lebih reliabel dan dapat dijadikan dasar yang kuat dalam perumusan strategi retensi pelanggan .

2.7. Teknologi Pengembangan Sistem

Pemilihan *stack* teknologi yang tepat merupakan faktor kunci dalam memastikan integrasi antara model *machine learning* dan antarmuka pengguna berjalan secara efisien serta memiliki skalabilitas yang baik.

2.7.1. Web Framework FastAPI

Aplikasi web dalam penelitian ini dikembangkan menggunakan FastAPI, sebuah *framework* web modern berbasis Python yang dirancang untuk membangun API dengan performa tinggi dan mendukung pemrograman asinkron. FastAPI memungkinkan sistem menangani banyak permintaan secara simultan tanpa menghambat proses komputasi lainnya, yang penting dalam implementasi model prediksi *churn* berbasis *XGBoost* pada lingkungan produksi. Keunggulan teknis FastAPI terletak pada penggunaan anotasi tipe Python yang memungkinkan validasi data otomatis melalui *Pydantic*, sehingga memastikan bahwa data pelanggan yang dikirimkan ke model selalu dalam format yang tepat. Integrasi model pembelajaran mesin melalui arsitektur API yang ringan dan efisien mendukung percepatan proses analisis serta membantu manajemen dalam mengambil keputusan retensi secara lebih responsif (FastAPI Documentation, 2024).

2.7.2. Neon Database (PostgreSQL)

Untuk pengelolaan data pelanggan dan riwayat prediksi yang berkelanjutan, sistem ini menggunakan Neon Database, sebuah layanan basis data *cloud-native* berbasis PostgreSQL yang dirancang untuk mendukung skalabilitas dan pemisahan komputasi serta penyimpanan data.

Penggunaan basis data yang persisten sangat penting dalam ekosistem *machine learning* untuk menyimpan log pelatihan model, riwayat prediksi, serta data mentah pelanggan guna mendukung pemantauan performa model dari waktu ke waktu. PostgreSQL sebagai sistem manajemen basis data relasional menyediakan struktur penyimpanan data yang terorganisasi, mendukung transaksi yang andal, serta menjaga integritas data melalui mekanisme ACID (PostgreSQL Documentation, 2024). Dengan arsitektur berbasis cloud, sistem memungkinkan penyimpanan data secara terpusat sehingga mempermudah pemeliharaan infrastruktur dan integrasi dengan proses pengembangan berkelanjutan seperti CI/CD. Pendekatan ini mendukung implementasi sistem prediksi *churn* yang lebih stabil, terukur, dan siap untuk digunakan dalam lingkungan produksi.

2.8. Strategi Retensi Pelanggan Berbasis Data

Strategi retensi pelanggan yang efektif memerlukan transisi dari analisis deskriptif ke arah analisis preskriptif, di mana hasil prediksi tidak hanya berhenti pada angka peluang, tetapi berlanjut pada tindakan nyata yang dipersonalisasi.

2.8.1. Klasifikasi Tingkat Risiko

Klasifikasi tingkat risiko merupakan proses segmentasi pelanggan berdasarkan besarnya probabilitas *churn* yang dihasilkan oleh model prediksi. Dalam arsitektur sistem ini, probabilitas tersebut diperoleh dari algoritma *XGBoost* dalam bentuk nilai numerik antara 0 hingga 1, yang merepresentasikan tingkat kemungkinan pelanggan akan berhenti berlangganan. Penggunaan ambang batas (*threshold*) probabilitas

diperlukan untuk mengelompokkan pelanggan berdasarkan tingkat urgensi intervensi.

Dalam penelitian ini, pelanggan dengan probabilitas di atas 0,7 dikategorikan sebagai *High Risk* (Risiko Tinggi) karena memiliki karakteristik yang sangat mendekati pola pelanggan yang sebelumnya melakukan *churn*. Pelanggan dengan probabilitas antara 0,3 hingga 0,7 diklasifikasikan sebagai *Medium Risk* (Risiko Sedang), yang menunjukkan adanya indikasi potensi *churn* namun masih terdapat tingkat keterikatan terhadap layanan. Sementara itu, pelanggan dengan probabilitas di bawah 0,3 dikategorikan sebagai *Low Risk* (Risiko Rendah), yang mencerminkan tingkat loyalitas relatif stabil.

Pendekatan segmentasi berbasis probabilitas ini memungkinkan perusahaan mengalokasikan sumber daya retensi secara lebih efisien dan terarah, sehingga intervensi tidak diberikan secara massal, melainkan disesuaikan dengan tingkat risiko masing-masing pelanggan. Strategi ini selaras dengan praktik pengambilan keputusan berbasis data (*data-driven decision making*) dalam sistem prediksi berbasis *machine learning*.

2.8.2. Implementasi Strategi Otomatis

Implementasi strategi retensi otomatis diintegrasikan langsung ke dalam antarmuka web berbasis FastAPI untuk memberikan rekomendasi tindakan yang responsif dan dipersonalisasi. Untuk pelanggan pada kategori *High Risk*, sistem secara otomatis menyarankan strategi agresif seperti pemberian diskon khusus sebesar 20% selama 6 bulan apabila pelanggan bersedia memperbarui kontrak, atau tawaran migrasi dari kontrak bulanan

(*month-to-month*) ke kontrak satu tahun yang lebih stabil. Selain itu, bagi pelanggan risiko tinggi yang belum memiliki layanan dukungan teknis premium, sistem merekomendasikan pemberian akses gratis selama 3 bulan sebagai upaya meningkatkan keterikatan pelanggan terhadap ekosistem layanan perusahaan.

Bagi pelanggan *Medium Risk*, fokus strategi bergeser pada peningkatan pengalaman pengguna melalui penawaran peningkatan kecepatan internet (*speed upgrade*) atau pemberian paket bundling layanan hiburan guna memperkuat loyalitas dan meningkatkan nilai pelanggan. Sedangkan bagi pelanggan *Low Risk*, sistem menyarankan tindakan apresiasi berupa pesan ucapan terima kasih disertai bonus kuota data tambahan (misalnya 5GB) atau penawaran produk perlindungan perangkat (*device protection*) sebagai bentuk strategi *up-selling* yang tetap mempertahankan kepuasan pelanggan.

Automasi ini juga mencakup penyesuaian strategi berdasarkan profil akun pelanggan. Misalnya, pelanggan yang masih menggunakan metode pembayaran *electronic check* disarankan untuk beralih ke sistem pembayaran otomatis (*autopay*) guna mengurangi potensi hambatan transaksi yang dapat memicu ketidakpuasan. Melalui pendekatan yang sistematis, berbasis data, dan terintegrasi dalam sistem prediksi, perusahaan telekomunikasi dapat meningkatkan efektivitas retensi pelanggan sekaligus menjaga profitabilitas jangka panjang secara lebih terukur.

.

BAB III

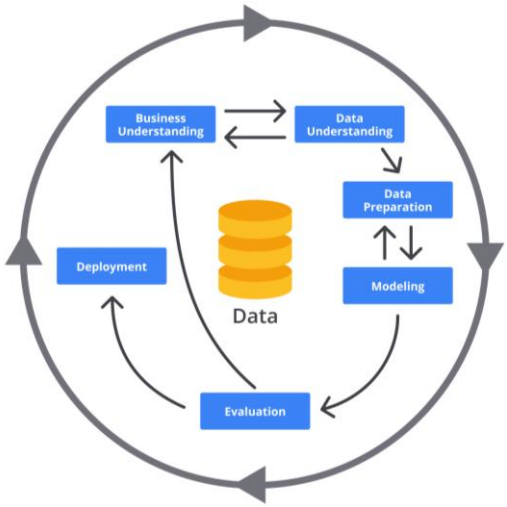
METODOLOGI PENELITIAN

3.1. Kerangka Kerja Penelitian

Pengembangan sistem prediksi customer *churn* ini dilakukan dengan mengikuti tahapan yang terstruktur guna memastikan bahwa setiap proses, mulai dari pengolahan data mentah hingga implementasi sistem web, berjalan secara sistematis dan valid. Pendekatan ini bertujuan untuk meminimalkan risiko kesalahan model dan memastikan bahwa solusi yang dibangun benar-benar menjawab kebutuhan strategi retensi pelanggan di industri telekomunikasi.

3.1.1. Metodologi CRISP-DM

Dalam penelitian ini sepenuhnya mengikuti kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Metodologi ini dipilih karena sifatnya yang iteratif dan terstruktur, memungkinkan setiap tahap pengembangan aplikasi web—mulai dari pemrosesan data di *backend* hingga penyajian hasil di *frontend*—selaras dengan tujuan bisnis perusahaan telekomunikasi (Schröer et al., 2021).



Gambar 3.1 *Alur Kerja CRISP-DM*

Berdasarkan struktur aplikasi yang telah dibangun, tahapan CRISP-DM diterapkan sebagai berikut:

A. *Business Understanding* (Pemahaman Bisnis) Tahap awal ini berfokus pada identifikasi masalah utama, yaitu tingginya biaya akuisisi pelanggan baru akibat fenomena *churn*. Dalam implementasi di GitHub, tahap ini direalisasikan melalui pembuatan modul retention_logic.py, di mana hasil prediksi model diterjemahkan menjadi strategi bisnis nyata. Tujuan teknisnya adalah membangun sistem yang tidak hanya memprediksi, tetapi juga memberikan rekomendasi otomatis berdasarkan kategori risiko (*Low, Medium, High*) untuk menjaga profitabilitas perusahaan.

B. *Data Understanding* (Pemahaman Data) Pada tahap ini, dilakukan eksplorasi terhadap dataset *Telco Customer Churn* untuk memahami karakteristik fitur yang memengaruhi keputusan pelanggan. Dalam sistem web, tahap ini direpresentasikan pada halaman analysis.html, yang menampilkan visualisasi distribusi data dan hubungan antar variabel. Peneliti mengidentifikasi variabel kunci seperti *tenure*, *Contract*, dan *MonthlyCharges* yang memiliki korelasi signifikan terhadap label target *Churn*.

C. *Data Preparation* (Persiapan Data) Tahap ini merupakan proses yang paling teknis, di mana data mentah diubah menjadi format siap latih. Merujuk pada file models/ml_pipeline.py, dilakukan pembersihan data berupa penanganan nilai kosong pada kolom TotalCharges dan penghapusan fitur non-prediktif seperti customerID. Selanjutnya, digunakan ColumnTransformer untuk menjalankan StandardScaler pada fitur numerik

dan OneHotEncoder pada fitur kategorikal secara otomatis, memastikan data yang masuk ke model bersifat homogen dan berkualitas tinggi.

D. *Modeling* (Pemodelan) Proses pemodelan dilakukan dengan mengintegrasikan algoritma *XGBoost* ke dalam sebuah jalur pipa (*pipeline*) fungsional. Untuk menangani ketidakseimbangan kelas yang ditemukan pada tahap *Data Understanding*, peneliti menyisipkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) ke dalam *pipeline* sebelum data diproses oleh pengklasifikasi. Selain itu, dilakukan optimasi melalui *RandomizedSearchCV* untuk menemukan kombinasi *hyperparameter* terbaik, seperti *learning_rate* dan *max_depth*, guna meminimalkan kesalahan prediksi.

E. *Evaluation* (Evaluasi) Evaluasi dilakukan untuk memastikan model memiliki performa yang handal sebelum disebarkan. Sistem web menyediakan *Dashboard* analisis yang menampilkan metrik *Accuracy, Precision, Recall,* dan *F1-Score* yang dihitung dari data uji. Fokus utama pada tahap ini adalah nilai *Recall*, karena dalam kasus *churn*, kegagalan mendeteksi pelanggan yang akan berhenti (negatif palsu) jauh lebih merugikan secara finansial dibandingkan kesalahan deteksi lainnya.

F. *Deployment* (Penyebaran) Tahap akhir melibatkan integrasi model statis ke dalam aplikasi web dinamis menggunakan FastAPI. Model yang telah dilatih disimpan dalam format .pkl dan dipanggil oleh file main.py untuk melayani permintaan prediksi secara *mengintegrasikan melalui mekanisme sinkronisasi data dan model*. Seluruh riwayat prediksi dan data operasional disimpan di dalam Neon Database (PostgreSQL cloud), memungkinkan

pihak manajemen untuk memantau tren *churn* dan efektivitas strategi retensi secara berkelanjutan melalui antarmuka web.

3.1.2. Alur Arsitektur Sistem Terintegrasi

Sistem yang dikembangkan dirancang dengan arsitektur modern yang memisahkan antara logika pemrosesan data, penyimpanan, dan antarmuka pengguna. Di sisi backend, kerangka kerja FastAPI digunakan untuk menjembatani komunikasi antara model *XGBoost* dengan pengguna secara asinkron, sementara Neon Database berfungsi sebagai repositori data berbasis *cloud* . Alur kerja ini memastikan bahwa setiap input pelanggan diproses secara mengintegrasikan melalui mekanisme sinkronisasi data dan model melalui jalur pipa (*pipeline*) yang telah dioptimalkan, mulai dari transformasi data hingga penyimpanan riwayat prediksi secara otomatis di PostgreSQL.

3.2. Sumber Data dan Variabel Penelitian

Data merupakan fondasi utama dalam membangun model prediktif yang akurat, sehingga identifikasi sumber data dan pemahaman terhadap karakteristik variabel menjadi langkah awal yang sangat krusial dalam penelitian ini.

3.2.1. Dataset Sekunder (Kaggle Telco Dataset)

Data yang digunakan dalam penelitian ini adalah dataset sekunder pelanggan telekomunikasi yang diperoleh dari platform Kaggle, yang telah banyak digunakan secara global dalam studi perilaku pelanggan. Dataset ini mencakup informasi dari 7.043 pelanggan unik dengan 21 atribut yang merepresentasikan profil demografis, kepemilikan layanan (seperti *Internet Service, Online Security, Streaming TV*), serta detail akun finansial pelanggan. Penggunaan data sekunder ini memberikan keuntungan berupa

validitas data yang telah teruji dalam berbagai penelitian *machine learning* sebelumnya .

3.2.2. Definisi Operasional Variabel

Variabel penelitian diklasifikasikan menjadi dua kategori utama, yaitu variabel independen (fitur) dan variabel dependen (target). Variabel dependen adalah *Churn*, yang merupakan label biner yang menunjukkan apakah pelanggan berhenti berlangganan atau tidak dalam satu bulan terakhir. Variabel independen terdiri dari variabel numerik seperti *tenure* (masa berlangganan dalam bulan) dan *MonthlyCharges*, serta variabel kategorikal seperti jenis kontrak (*Contract*), metode pembayaran (*PaymentMethod*), dan jenis layanan internet (*InternetService*). Pemilihan variabel-variabel ini didasarkan pada temuan bahwa faktor kontrak dan durasi berlangganan merupakan prediktor terkuat dalam memengaruhi loyalitas pelanggan di sektor telekomunikasi (Alotaibi & Haq, 2024).

3.3. Pra-pemrosesan Data (Data Preprocessing)

Kualitas hasil prediksi sangat bergantung pada kualitas data yang dimasukkan ke dalam model, sehingga tahap pra-pemrosesan dilakukan untuk mengubah data mentah menjadi format yang dapat dipahami secara optimal oleh algoritma XGBoost.

3.3.1. Pembersihan Data (Data Cleaning)

Proses pembersihan data mencakup penanganan nilai yang hilang (missing values) dan penghapusan fitur yang tidak relevan secara statistik. Dalam dataset ini, ditemukan nilai kosong pada variabel `*TotalCharges*` yang disebabkan oleh pelanggan baru dengan *tenure* nol bulan; nilai-nilai ini

kemudian dikonversi menjadi numerik dan diisi dengan nol untuk menjaga integritas data. Selain itu, atribut `*customerID*` dieliminasi dari dataset karena tidak memberikan nilai prediktif terhadap pola perpindahan pelanggan dan hanya berfungsi sebagai identitas unik.

3.3.2. Transformasi Fitur (*Scaling & Encoding*)

Untuk memastikan setiap fitur memiliki kontribusi yang seimbang dalam model, dilakukan transformasi data melalui teknik *standarisasi* dan *encoding*. Fitur numerik diproses menggunakan *StandardScaler* untuk menskalakan setiap variabel sehingga memiliki *mean* nol dan *standard deviation* satu, yang membantu model pembelajaran mesin mempelajari pola data secara lebih stabil dan cepat konvergen selama Latihan. Proses ini memastikan bahwa fitur dengan rentang nilai besar tidak mendominasi proses pembelajaran. Sementara itu, fitur kategorikal ditransformasikan menggunakan `OneHotEncoder` guna mengubah teks label menjadi representasi biner, sehingga algoritma *XGBoost* dapat menangkap hubungan antar kategori tanpa mengasumsikan adanya urutan (*ordinality*) tertentu (Chen & Guestrin, 2016).

3.4. Pengembangan Model *Machine learning*

Tahap pemodelan merupakan inti dari penelitian ini, di mana algoritma akan dilatih untuk mempelajari pola perilaku pelanggan yang *churn* melalui serangkaian proses optimasi dan penanganan ketidakseimbangan data.

3.4.1. Penanganan Imbalance Data dengan SMOTE

Salah satu tantangan utama dalam dataset *churn* adalah ketidakseimbangan kelas (*class imbalance*), di mana jumlah pelanggan yang tidak *churn* jauh lebih banyak dibandingkan yang *churn*. Untuk mengatasi hal

ini, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) yang bekerja dengan membangkitkan sampel sintetis pada kelas minoritas.

Penting untuk ditegaskan bahwa dalam penelitian ini, SMOTE diterapkan secara eksklusif hanya pada data *training* di dalam setiap *fold cross-validation* untuk mencegah fenomena *data leakage* (kebocoran data). Jika SMOTE diterapkan pada keseluruhan dataset sebelum proses pembagian data, informasi dari data uji dapat secara tidak langsung bocor ke dalam data latih melalui proses interpolasi sampel sintetis, yang mengakibatkan estimasi performa model menjadi terlalu optimis (*over-optimistic*) dan tidak akurat saat dihadapkan pada data dunia nyata. Justifikasi ini krusial untuk menjaga integritas ilmiah dan memastikan bahwa model dievaluasi pada distribusi data asli yang tidak dimanipulasi.

3.4.2 Implementasi Model Pembanding (Baseline Comparison)

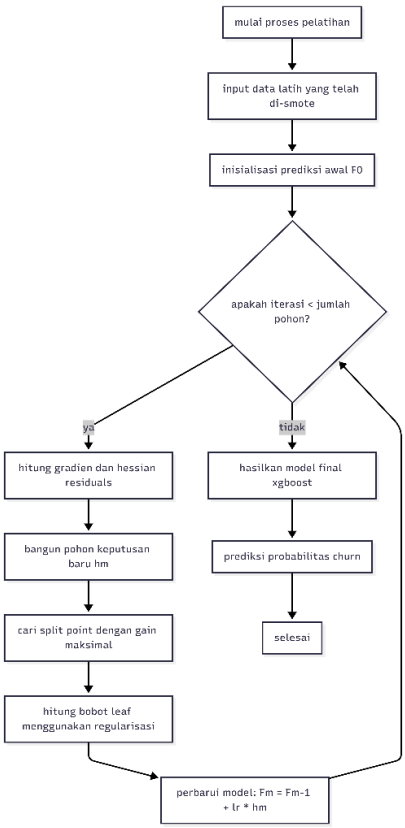
Meskipun fokus utama penelitian adalah algoritma XGBoost, desain metodologi ini menyertakan penggunaan model pembanding berupa *Logistic Regression* (LR) dan *Random Forest* (RF) sebagai *baseline*. Secara akademik, keberadaan *baseline model* bertujuan untuk memberikan pembandingan relatif sehingga klaim keunggulan *XGBoost* memiliki dasar evaluasi yang objektif dan tidak bersifat asumtif.

Logistic Regression digunakan sebagai *baseline* linear sederhana untuk menguji apakah pola *churn* dapat ditangkap tanpa memerlukan kompleksitas tinggi, sementara *Random Forest* mewakili pendekatan ansambel berbasis *bagging*. Dengan membandingkan *XGBoost* terhadap standar referensi ini, penelitian dapat secara empiris membuktikan nilai tambah dari mekanisme

gradient boosting dan regularisasi yang dimiliki *XGBoost* dalam menangani data telekomunikasi yang kompleks.

Implementasi model pembanding ini dilakukan secara eksklusif pada tahap eksperimen data dan validasi teknis selama proses penelitian berlangsung. Sistem informasi berbasis web yang dikembangkan hanya akan mengimplementasikan model *XGBoost* sebagai algoritma tunggal yang sudah teruji paling unggul. Hal ini dilakukan agar aplikasi web tetap fokus pada fungsionalitas praktis dalam menyajikan hasil prediksi dan rekomendasi strategi retensi bagi pengguna tanpa perbandingan algoritma di sisi antarmuka.

3.4.3. Algoritma *XGBoost* (*Extreme Gradient Boosting*)



Gambar 3.2 *Flowchart Algoritma XG-Boost*

Alur kerja algoritma *XGBoost* dalam sistem ini dimulai dengan pra-pemrosesan data melalui *standardscaler* dan *onehotencoder* guna

menormalisasi fitur sebelum dilakukan penyeimbangan kelas menggunakan teknik SMOTE untuk mengatasi ketidakseimbangan data *churn*. selanjutnya, algoritma secara iteratif membangun serangkaian pohon keputusan melalui optimasi *gradien* untuk memperbaiki kesalahan prediksi dari model sebelumnya hingga mencapai titik fungsi kerugian minimum (chen dan guestrin 2016). Proses ini dioptimalkan melalui *randomizedsearchcv* untuk menemukan parameter terbaik yang kemudian menghasilkan output berupa probabilitas *churn* dan kategori risiko pelanggan sebagai landasan pengambilan keputusan strategi retensi (Amin et al., 2022).

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{m=1}^{m} \Omega(f_m)$$

keterangan masing-masing komponen adalah sebagai berikut:

i. $\mathcal{L}$: Menyatakan nilai fungsi objektif yang diminimalkan selama proses pelatihan model *XGBoost.*

j. n: Jumlah data atau observasi pada dataset pelatihan.

k. y_i: Nilai aktual (label sebenarnya) dari observasi ke-i.

l. $\hat{y}_i$: Nilai prediksi yang dihasilkan oleh model untuk observasi ke-i.

m. $l(y_i, \hat{y}_i)$: Fungsi kerugian (*loss function*) yang mengukur kesalahan prediksi antara nilai aktual dan nilai prediksi. Pada kasus klasifikasi *customer churn*, fungsi ini umumnya berupa *logistic loss*.

n. M: Jumlah pohon keputusan (*decision trees*) yang digunakan dalam model XGBoost.

o. f_m: Pohon keputusan ke-m yang berkontribusi dalam pembentukan prediksi model.

p.$\Omega(f_m)$: Fungsi regularisasi yang memberikan penalti terhadap kompleksitas pohon keputusan ke-m.

3.4.3. Optimasi Hyperparameter

Guna mencapai performa puncak, model *XGBoost* dioptimalkan menggunakan teknik `*RandomizedSearchCV*` dengan skema validasi silang 3-lapisan (*3-fold cross-validation*). Parameter yang dioptimalkan mencakup jumlah *estimator* (`n_estimators`), kedalaman maksimum pohon (`*max_depth*`), dan tingkat pembelajaran (`*learning_rate*`). Proses ini memastikan bahwa model yang dihasilkan memiliki kemampuan generalisasi yang baik dan tidak hanya akurat pada data pelatihan saja.

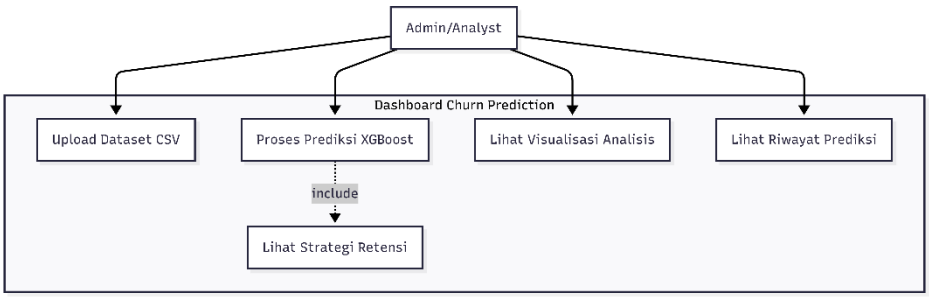
3.5. Implementasi Sistem Web dan *Cloud* Database

Implementasi sistem dilakukan untuk mengubah model statis menjadi aplikasi fungsional yang dapat digunakan oleh pihak manajemen untuk melakukan prediksi harian dan merancang strategi retensi secara otomatis.

3.5.1. Pemodelan UML

Proses ini berfokus pada perancangan seperti struktur data dan arsitektur perangkat lunak yang dibuat dengan pemodelan UML seperti *use case diagram, activity diagram, sequence diagram,* dan *class diagram.*

1. *Use Case Diagram*



Gambar 3.3 *Use Case Diagram Sistem Prediksi Churn*

Use case diagram ini menggambarkan interaksi antara Admin/Analis dengan sistem prediksi *churn*. Pengguna mengunggah dataset telekomunikasi dalam format CSV sebagai tahap awal pemrosesan. Proses unggah ini menjadi langkah awal agar sistem dapat melakukan analisis menggunakan model *machine learning* yang telah tersedia.

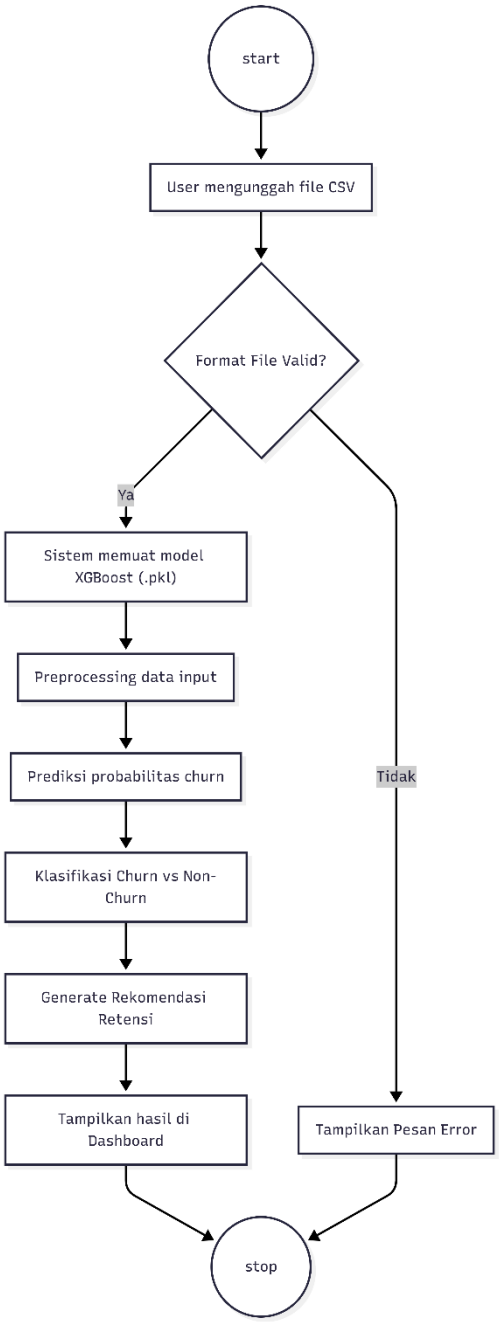
Setelah data diunggah, pengguna menjalankan algoritma *XGBoost* untuk menghasilkan prediksi *churn*. Sistem secara otomatis menghitung probabilitas risiko setiap pelanggan. Hasilnya ditampilkan dalam bentuk klasifikasi dan visualisasi interaktif untuk membantu analis memahami pola *churn* dan mendukung pengambilan keputusan.

Sistem juga menyediakan fitur riwayat prediksi untuk memantau hasil analisis sebelumnya. Selain itu, pelanggan dengan risiko *churn* tinggi akan disertai rekomendasi strategi retensi. Dengan demikian, sistem berfungsi sebagai platform pendukung keputusan, bukan sekadar alat prediksi teknis.membuktikan bahwa sistem bukan sekadar alat prediksi teknis, melainkan sebuah platform pendukung keputusan yang komprehensif untuk menjaga loyalitas pelanggan di industri telekomunikasi.

Integrasi teknologi Generative AI (Gemini API) dalam sistem ini memberikan dimensi baru melalui analisis preskriptif, di mana hasil prediksi tidak hanya berhenti pada angka peluang, tetapi berlanjut pada rekomendasi tindakan nyata yang dipersonalisasi. Melalui pemanfaatan platform pendukung keputusan ini, pihak manajemen dapat menarik data historis secara akurat untuk memantau tren loyalitas pelanggan serta mengevaluasi efektivitas strategi retensi dalam jangka panjang. Keberadaan visualisasi interaktif dan

narasi saran komunikasi yang dihasilkan oleh AI bertujuan untuk menjembatani kesenjangan antara hasil analisis teknis yang kompleks dengan kebutuhan praktis operasional di industri telekomunikasi.

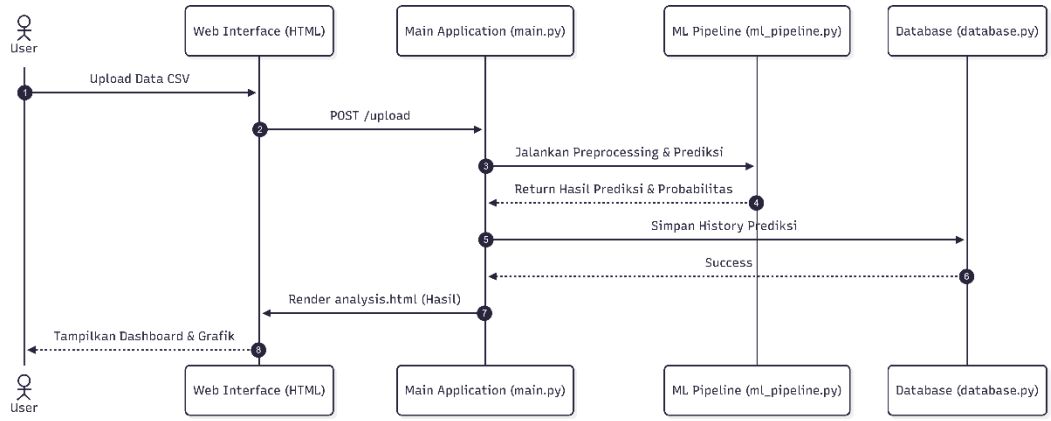
2. *Activity Diagram*



***Gambar 3.4** Activity Diagram*

Diagram aktivitas ini merinci langkah-langkah logis yang terjadi di dalam sistem saat menjalankan fungsi prediksi. Dimulai dari aksi pengguna mengunggah file CSV, sistem akan melakukan validasi format secara otomatis. Jika valid, sistem melanjutkan ke tahap pemrosesan data menggunakan model yang telah dilatih (.pkl) hingga akhirnya menghasilkan klasifikasi risiko dan menampilkan rekomendasi tindakan pada antarmuka *Dashboard*.

3. *Sequence Diagram*

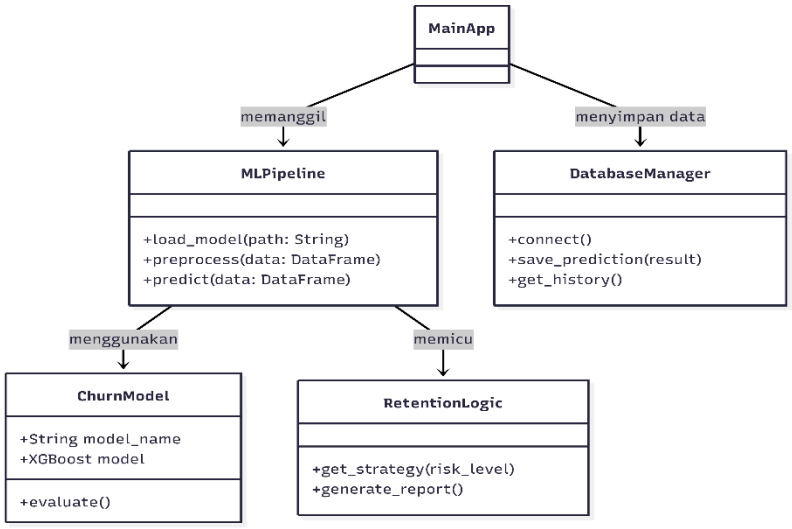


Gambar 3.5 *Sequence Diagram*

Sequence diagram ini menjelaskan urutan interaksi antar komponen sistem dalam satu proses prediksi. Proses dimulai saat pengguna mengunggah dataset melalui antarmuka web, lalu permintaan tersebut diterima dan dikendalikan oleh main.py untuk mengatur alur komunikasi antar modul. Setelah data diterima selanjutnya, main.py meneruskan data ke ml_pipeline.py untuk pra-pemrosesan dan eksekusi model *XGBoost*. Modul ini menghasilkan probabilitas dan label klasifikasi *churn*. Hasil prediksi kemudian disimpan melalui database.py agar tercatat dalam riwayat sistem. Terakhir, hasil analisis dikirim kembali ke pengguna melalui template *Flask* dan ditampilkan pada halaman analysis.html. *Dashboard* menampilkan ringkasan statistik,

visualisasi interaktif, dan rekomendasi strategi retensi berdasarkan hasil prediksi.

4. *Class Diagram*

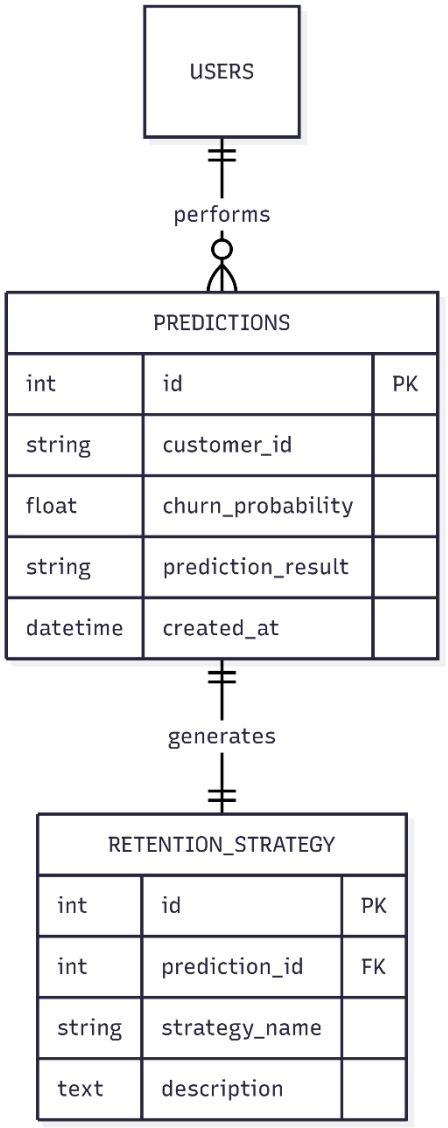


Gambar 3.6 *Class Diagram*

Class diagram ini memaparkan arsitektur perangkat lunak yang dibangun dengan menerapkan prinsip *Separation of Concerns* (pemisahan tanggung jawab) untuk memastikan modularitas kode. Fokus utama diagram ini terletak pada kelas MLPipeline yang berfungsi sebagai orkestrator alur data, mulai dari tahap pembersihan hingga transformasi fitur sebelum diproses. Kelas ini memiliki ketergantungan pada kelas *Churn*Model, sebuah komponen terenkapsulasi yang membungkus algoritma *XGBoost* dan bertanggung jawab atas pemuatan model dari file serialisasi (.pkl) serta eksekusi inferensi. Di sisi lain, kelas *DatabaseManager* bertindak sebagai lapisan abstraksi data yang menangani seluruh operasi persistensi, seperti penyimpanan log prediksi pelanggan dan pengambilan riwayat data dari basis data SQL. Melalui struktur pemrograman berorientasi objek ini, setiap modul memiliki batasan fungsi

yang jelas, sehingga memudahkan proses pemeliharaan sistem, memungkinkan penggantian algoritma tanpa merusak logika basis data, serta meningkatkan skalabilitas aplikasi di masa depan.

5. *Entity Relationship Diagram* (ERD)



***Gambar 3.7** Entity Relationship Diagram (ERD)*

Entity Relationship Diagram (ERD) ini menggambarkan arsitektur data logis yang menjadi fondasi bagi fitur penyimpanan dan pelaporan pada aplikasi web prediksi *churn*. Pusat dari perancangan ini adalah entitas *PREDICTIONS*, yang

berfungsi sebagai repositori utama untuk mengarsipkan setiap *output* yang dihasilkan oleh algoritma XGBoost. Tabel ini menyimpan atribut-atribut krusial seperti identitas pelanggan (*Customer ID*), skor probabilitas *churn* yang presisi, label klasifikasi, hingga stempel waktu (*timestamp*) untuk kebutuhan audit data. Entitas ini memiliki hubungan relasional dengan tabel RETENTION_STRATEGY, di mana setiap hasil prediksi secara otomatis dipetakan ke dalam skema tindakan preventif yang spesifik. Hubungan ini memastikan bahwa sistem tidak hanya berfungsi sebagai alat deteksi, tetapi juga sebagai penyedia solusi bisnis yang terstruktur. Dengan skema basis data yang terorganisir ini, pengelola sistem dapat menarik data historis secara akurat untuk memantau tren loyalitas pelanggan serta mengevaluasi efektivitas strategi retensi dalam jangka panjang melalui *Dashboard* interaktif.

3.5.2. Backend FastAPI

Aplikasi web dibangun menggunakan FastAPI, sebuah framework web modern berbasis Python yang mendukung pemrograman asinkron dan performa tinggi. FastAPI dipilih karena kemampuannya dalam menyediakan endpoint API yang responsif serta validasi data otomatis melalui anotasi tipe Python. Dengan arsitektur ini, sistem mampu memproses input pengguna dan menghasilkan prediksi kategori risiko pelanggan (*Low, Medium, High*) secara cepat dan efisien dalam waktu singkat setelah data dikirimkan melalui antarmuka web (FastAPI Documentation, 2024).

3.5.3. Penyimpanan Neon Database (PostgreSQL)

Untuk pengelolaan data yang persisten, sistem diintegrasikan dengan Neon Database, layanan *serverless PostgreSQL* yang berjalan di lingkungan

cloud. Penggunaan basis data berbasis PostgreSQL memungkinkan penyimpanan riwayat prediksi secara terpusat dan terstruktur, sehingga memudahkan proses audit data serta penyediaan data historis untuk keperluan *retraining* model di masa mendatang. Arsitektur berbasis *cloud* ini juga mendukung skalabilitas dan pemisahan antara komputasi dan penyimpanan data.

Integrasi basis data dikelola menggunakan SQLAlchemy sebagai *Object Relational Mapping (ORM)*, yang berfungsi sebagai perantara antara aplikasi Python dan sistem basis data relasional. Penggunaan ORM membantu menjaga konsistensi transaksi, meningkatkan keamanan *query* melalui mekanisme parameter binding, serta mempermudah pengelolaan skema data secara terstruktur.

3.6. Evaluasi dan Validasi

Tahap akhir metodologi adalah melakukan pengujian terhadap model dan sistem untuk memastikan bahwa hasil prediksi memiliki tingkat kepercayaan yang tinggi dan dapat dipertanggungjawabkan secara ilmiah.

3.6.1. Protokol *Stratified 3-Fold Cross Validation*

Untuk memastikan stabilitas performa model dan menghindari ketergantungan pada satu pembagian data (*train-test split*) yang bersifat kebetulan, penelitian ini menerapkan teknik *Stratified K-Fold Cross Validation* dengan nilai k=3. Teknik stratifikasi menjamin bahwa setiap *fold* memiliki proporsi kelas pelanggan *churn* dan *non-churn* yang konsisten dengan dataset aslinya, yang sangat krusial pada dataset tidak seimbang. Seluruh hasil evaluasi dalam penelitian ini dilaporkan dalam bentuk rata-rata dan standar deviasi

(mean ± standard deviation) untuk menunjukkan kemampuan generalisasi model secara lebih reliabel dan transparan.

3.6.2. Optimasi Decision Threshold

Dalam mengklasifikasikan risiko *churn*, penggunaan ambang batas (*threshold*) tunggal seperti 0,5 seringkali tidak optimal untuk kebutuhan bisnis. Penelitian ini memastikan bahwa setiap *threshold* yang digunakan (misalnya 0,7 untuk segmentasi risiko tinggi) memiliki dasar kuantitatif dan tidak bersifat arbitrer. Penentuan *threshold* dilakukan berdasarkan analisis *Receiver Operating Characteristic* (ROC) *Curve* dengan pendekatan *Youden Index (J = Sensitivity + Specificity - 1)* untuk memaksimalkan kemampuan diskriminasi model, atau berdasarkan analisis *Precision-Recall Curve* guna mengoptimalkan *F1-Score*. Pemilihan *threshold* ini mempertimbangkan *trade-off* antara *precision* dan *recall* agar selaras dengan tujuan bisnis retensi pelanggan; di mana kegagalan mendeteksi calon *churner* (*False Negative*) memiliki dampak finansial yang lebih besar dibandingkan kesalahan deteksi pelanggan setia (*False Positive*).

3.6.3. Metrik Evaluasi Klasifikasi

Kinerja model diukur secara komprehensif menggunakan metrik-metrik berikut yang masing-masing memiliki relevansi spesifik dalam konteks strategi retensi:

1. *Accuracy:* Memberikan gambaran umum proporsi prediksi yang benar, namun digunakan secara terbatas karena sifat data yang tidak seimbang.

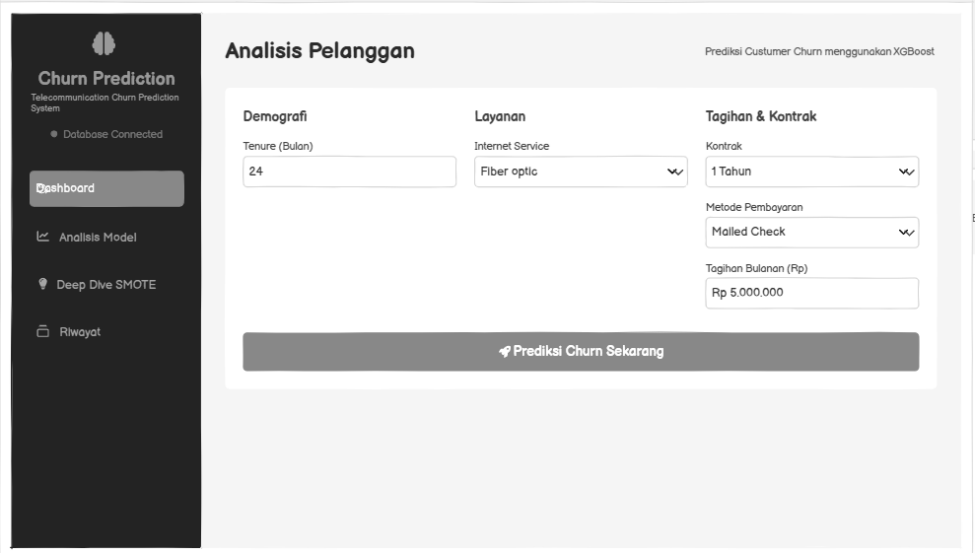
2. *Recall (Sensitivity)*: Menjadi fokus utama untuk memastikan sistem mampu mendeteksi sebanyak mungkin pelanggan yang benar-benar berisiko *churn* agar intervensi tepat waktu dapat dilakukan.
3. *Precision*: Penting untuk efisiensi strategi retensi, memastikan bahwa sumber daya perusahaan (seperti diskon atau paket promo) tidak diberikan secara percuma kepada pelanggan yang sebenarnya tidak berniat berhenti.
4. *F1-Score*: Memberikan keseimbangan harmonik antara *Recall* dan *Precision*.
5. *ROC-AUC & PR-AUC*: Mengukur performa model pada berbagai tingkat *threshold* guna memastikan stabilitas model dalam mengklasifikasikan kelas minoritas.
6. *Confusion Matrix*: Visualisasi distribusi prediksi vs aktual untuk memantau kesalahan tipe I dan tipe II secara eksplisit.

3.6.4. Analisis *Feature Importance*

Selain metrik kuantitatif, sistem ini juga melakukan analisis tingkat kepentingan fitur (*feature importance*) untuk memberikan transparansi mengenai variabel yang paling memengaruhi keputusan model XGBoost. Dalam implementasinya, interpretasi dan penjelasan terhadap hasil *feature importance* didukung oleh pemanfaatan model *generatif* Gemini AI yang diakses melalui *platform* Google AI Studio, sehingga informasi yang dihasilkan tidak hanya bersifat teknis tetapi juga kontekstual dan mudah dipahami. Hasil analisis tersebut kemudian ditampilkan dalam bentuk visualisasi pada *Dashboard* web, sehingga memberikan wawasan strategis bagi perusahaan mengenai faktor-faktor utama yang berkontribusi terhadap

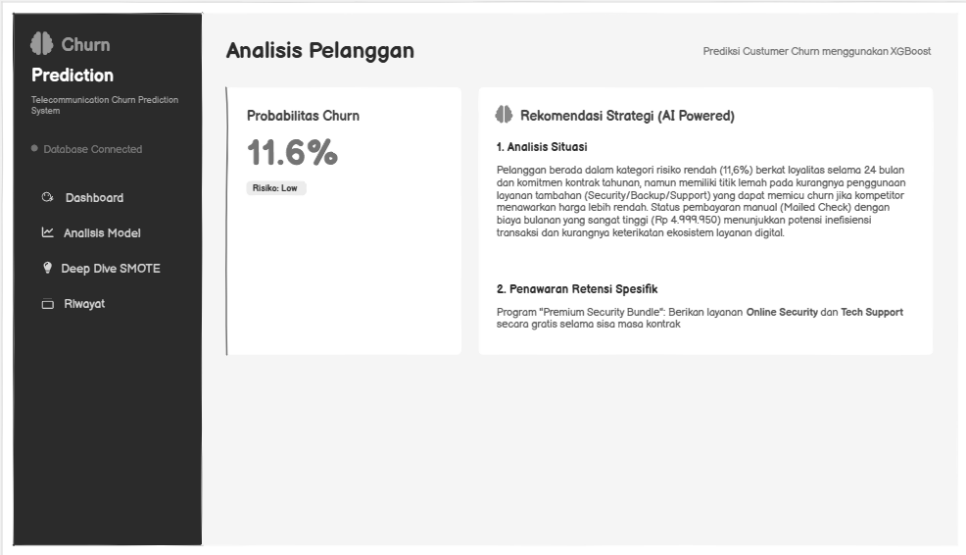
terjadinya *churn*, seperti besaran biaya bulanan atau jenis kontrak yang digunakan pelanggan.

3.7. Perancangan dan Implementasi Visualisasi Hasil Berbasis Web



Gambar 3.8 *Rancangan Visualisasi Dashboard Web*

Gambar tersebut merupakan wireframe antarmuka utama dari sistem prediksi *customer churn* yang dirancang untuk memfasilitasi pengguna dalam melakukan input data pelanggan secara individual guna mendapatkan hasil prediksi secara *real-time*. Antarmuka ini mengintegrasikan seluruh variabel independen yang telah ditentukan, mencakup kategori demografi (*tenure*), layanan (*internet service*), serta tagihan dan kontrak (*contract*, *payment method*, dan *monthly charges*), yang disusun secara ergonomis untuk memudahkan entri data. Melalui tombol "Prediksi Churn Sekarang", data yang diinput akan diproses oleh algoritma XGBoost untuk menghasilkan klasifikasi status pelanggan, sementara panel navigasi di sisi kiri menyediakan akses cepat menuju fitur analisis performa model, detail penanganan *imbalanced data* menggunakan teknik SMOTE, serta riwayat prediksi, yang secara keseluruhan berfungsi sebagai instrumen pendukung keputusan dalam strategi retensi pelanggan di industri telekomunikasi.



Gambar 3.9 *Rancangan Visualisasi Hasil Prediksi pada Web*

Gambar tersebut merupakan antarmuka hasil prediksi dan rekomendasi strategi yang menyajikan luaran dari model XGBoost dalam bentuk persentase probabilitas *churn* serta klasifikasi tingkat risiko pelanggan. Pada tampilan ini, sistem tidak hanya berfungsi sebagai alat prediksi numerik, tetapi juga mengintegrasikan fitur Rekomendasi Strategi berbasis AI (*AI-Powered*) yang memberikan analisis situasi mendalam berdasarkan data spesifik pelanggan, seperti pengaruh masa berlangganan (*tenure*), jenis kontrak, dan besaran tagihan. Penjelasan naratif yang dihasilkan mencakup identifikasi titik lemah pelanggan serta saran penawaran retensi yang personal, yang bertujuan untuk mengubah hasil prediksi teknis menjadi tindakan manajerial yang konkret bagi perusahaan telekomunikasi dalam upaya menekan angka *churn* dan meningkatkan loyalitas pelanggan secara efektif.

3.8. Jadwal Penelitian

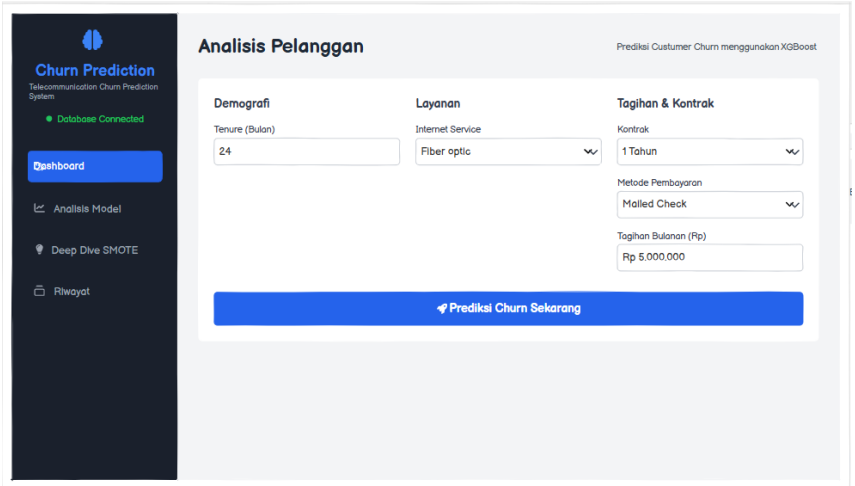
Tabel 3.1 Jadwal Penelitian

No	Kegiatan	Waktu				
		1	2	3	4	5
1	Penulisan proposal					
2	Seminar dan bimbingan proposal					
3	Penelitian dan tindakan					
4	Analisis dan bimbingan hasil penelitian					
5	Ujian skripsi					

BAB IV

HASIL DAN PEMBAHASAN

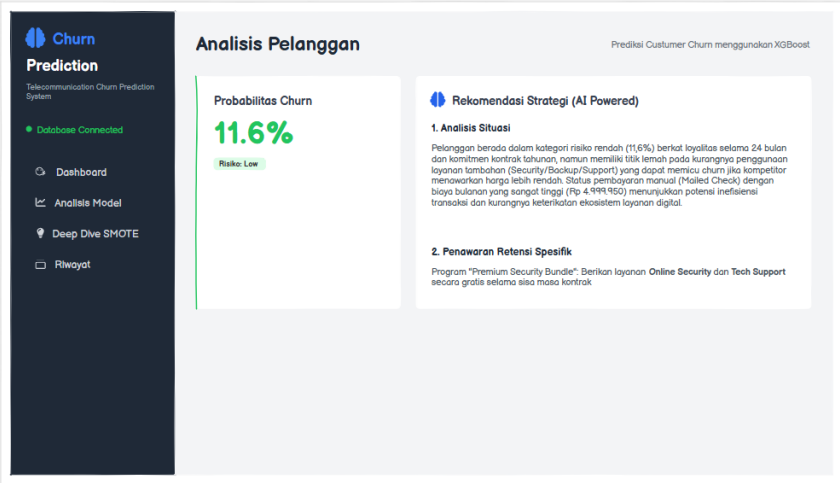
4.1. Tampilan *Dashboard*



***Gambar 4. 1** Tampilan Dashboard*

Dashboard merupakan halaman utama sistem yang berfungsi sebagai pusat pengelolaan dan analisis data pelanggan dalam proses prediksi customer *churn*. Pada bagian kiri terdapat panel navigasi yang memuat menu *Dashboard*, Analisis *Model, Deep Dive SMOTE*, dan Riwayat, sehingga pengguna dapat mengakses seluruh fitur sistem secara terstruktur. Selain itu, ditampilkan indikator status koneksi basis data (Database Connected) yang menunjukkan bahwa sistem telah terhubung dengan database dan siap digunakan. Bagian utama *Dashboard* berisi form analisis pelanggan yang disusun berdasarkan variabel penelitian yang digunakan dalam pelatihan model XGBoost. Perancangan *Dashboard* ini bertujuan untuk mempermudah proses input data, meningkatkan efisiensi penggunaan sistem, serta mendukung pengambilan keputusan dalam strategi retensi pelanggan secara berbasis data.

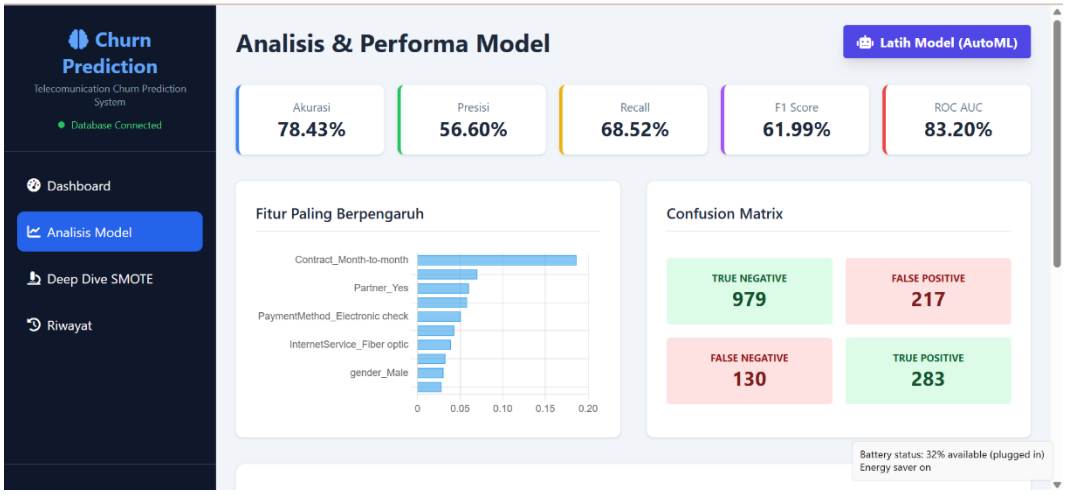
4.2. Tampilan Hasil Prediksi *Churn*



Gambar 4.2 *Hasil Prediksi Churn*

Hasil prediksi menunjukkan bahwa sistem, melalui algoritma XGBoost, menghitung probabilitas pelanggan dan mengklasifikasikannya ke dalam kelas kategori . Nilai probabilitas tersebut diperoleh dari proses pengolahan variabel input pelanggan yang telah melalui tahap preprocessing dan pemodelan sesuai dengan data pelatihan dan sebagai dasar dalam pengambilan keputusan retensi pelanggan.

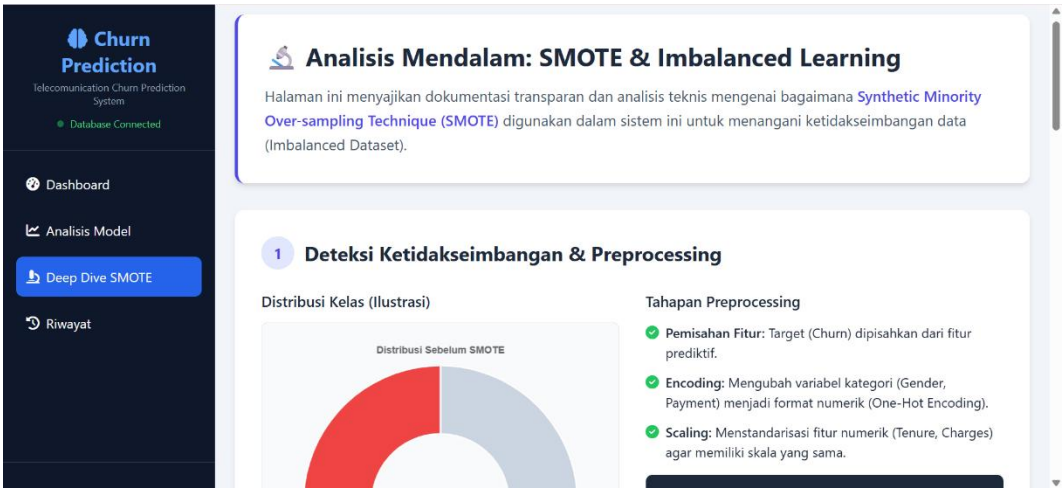
4.3 Tampilan Analisis Model



Gambar 4.3 *Tampilan Analisis Model*

Halaman Analisis & Performa Model berfungsi untuk menampilkan proses evaluasi kinerja algoritma *XGBoost* setelah model dilatih menggunakan data pelatihan dan diuji pada data pengujian. Pada tahap ini, sistem menghitung berbagai metrik evaluasi seperti akurasi, presisi*, recall, F1-score,* dan *ROC AUC* untuk mengukur kemampuan model dalam mengklasifikasikan pelanggan ke dalam kategori tingkat *churn*. Selain itu, sistem menghasilkan *confusion matrix* yang menggambarkan distribusi hasil prediksi berdasarkan kombinasi prediksi dan kondisi aktual, sehingga dapat dianalisis tingkat kesalahan klasifikasi yang terjadi. Halaman ini juga menampilkan fitur-fitur yang paling berpengaruh dalam proses prediksi berdasarkan perhitungan *feature importance* dari algoritma XGBoost.

4.4. Tampilan Deep Dive SMOTE



Gambar 4. 4 *Deteksi Ketidakseimbangan & Preprocessing*

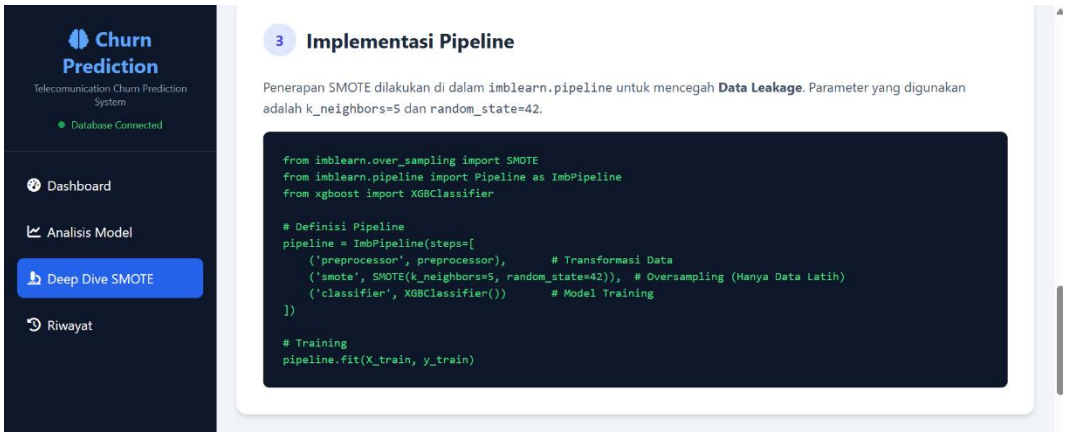
Gambar ini menampilkan tahap awal analisis ketidakseimbangan data *churn*. Sistem terlebih dahulu menunjukkan distribusi kelas sebelum penerapan SMOTE untuk mengidentifikasi dominasi kelas non-*churn* terhadap *churn*. Pada tahap *preprocessing*, dilakukan pemisahan fitur dan target, *encoding* variabel kategorikal menggunakan One-Hot Encoding, serta standarisasi fitur numerik. Proses ini

memastikan data berada dalam format yang siap digunakan oleh model dan mencegah bias akibat perbedaan skala fitur.

(*churn*).

Gambar 4.5 *Mekanisme Kerja SMOTE*

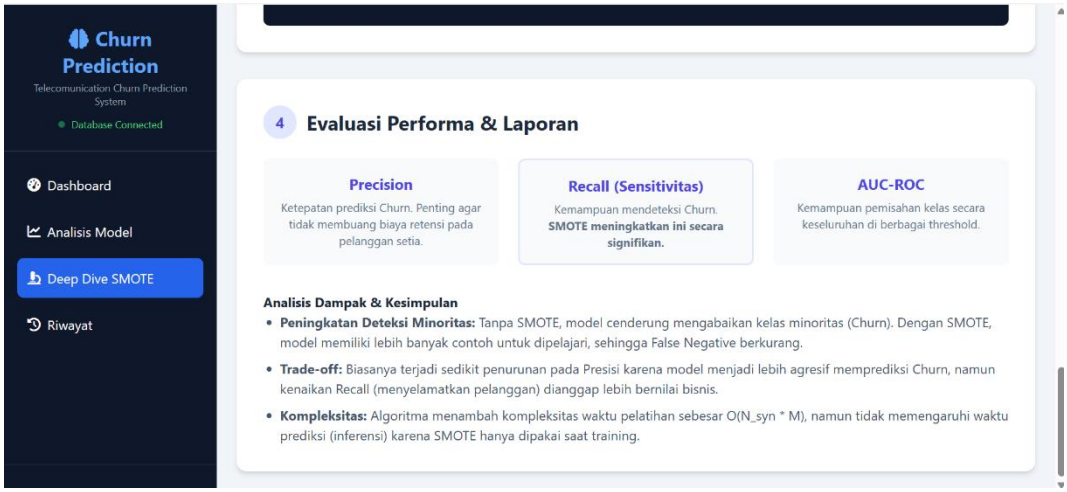
Gambar ini menjelaskan cara kerja SMOTE dalam menghasilkan data sintetis untuk kelas minoritas. SMOTE tidak menduplikasi data, tetapi membuat sampel baru berdasarkan pendekatan *k-Nearest Neighbors* (k-NN). Prosesnya meliputi pemilihan satu sampel minoritas, mencari k tetangga terdekat (k=5), memilih satu tetangga secara acak, lalu membuat titik baru di antara keduanya. Pendekatan ini membantu memperkaya variasi data minoritas sehingga model dapat belajar pola *churn* dengan lebih baik.



Gambar 4.6 *Implementasi Pipeline*

ambar ini menunjukkan implementasi SMOTE dalam imblearn.pipeline untuk mencegah data leakage. Pipeline terdiri dari tiga tahapan: preprocessing, oversampling menggunakan SMOTE, dan pelatihan model XGBoost.

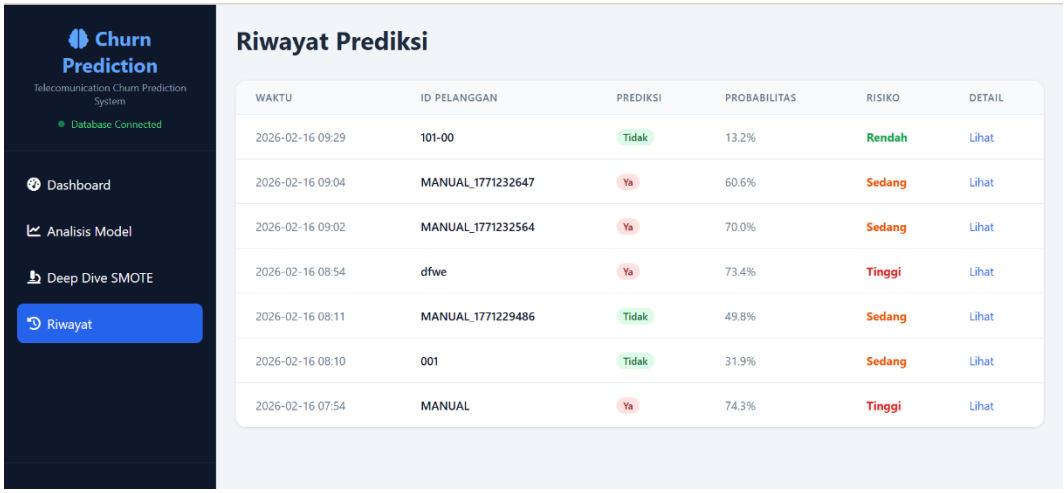
SMOTE hanya diterapkan pada data latih, sehingga tidak memengaruhi data uji. Integrasi ini membuat proses pelatihan lebih terstruktur, aman, dan konsisten dalam menangani imbalance sebelum model melakukan klasifikasi.



Gambar 4.7 *Evaluasi Performa & Laporan*

Gambar ini menampilkan hasil evaluasi performa setelah penerapan SMOTE. Metrik yang ditampilkan meliputi *Precision, Recall (*Sensitivitas*),* dan *AUC-ROC.* Hasil menunjukkan bahwa SMOTE meningkatkan Recall secara signifikan, sehingga kemampuan model dalam mendeteksi pelanggan *churn* menjadi lebih baik. Meskipun terdapat potensi penurunan kecil pada *Precision*, peningkatan *Recall* dianggap lebih bernilai dalam konteks bisnis karena dapat mengurangi risiko kehilangan pelanggan.

4.5. Tampilan Riwayat Prediksi



Gambar 4.8 *Tampilan Riwayat Prediksi*

Halaman Riwayat Prediksi berfungsi sebagai media pencatatan dan dokumentasi seluruh hasil prediksi yang telah dilakukan oleh sistem. Pada halaman ini ditampilkan informasi waktu prediksi, identitas pelanggan, hasil klasifikasi , nilai probabilitas, serta tingkat risiko yang telah dikategorikan oleh sistem. Setiap entri prediksi disimpan secara otomatis setelah proses klasifikasi selesai, sehingga memungkinkan pengguna untuk melakukan penelusuran kembali terhadap hasil analisis sebelumnya. Selain itu, tersedia fitur detail untuk melihat informasi lebih lanjut terkait setiap hasil prediksi. Keberadaan halaman ini mendukung aspek transparansi, pelacakan hasil, serta evaluasi berkelanjutan dalam penerapan strategi retensi pelanggan berbasis data.

4.6. Perhitungan Manual XG-Boost

Berikut adalah simulasi perhitungan menggunakan baris pertama dari dataset (customerID: 7590-VHVEG)

1) Transformasi Vektor Fitur (*Encoding*)

Sebelum masuk ke rumus XGBoost, semua data kategorikal diubah menjadi angka (0 atau 1).

Data Input (Pelanggan 1):

Numerik: *tenure*=1, *MonthlyCharges*=29.85, *TotalCharges*=29.85

Kategorikal (Contoh):

a) gender_Female: 1
b) Partner_Yes: 1
c) Dependents_No: 1
d) Phone Service: 0
e) MultipleLines: 0
f) Contract_Month-to-month: 1
g) InternetService_DSL: 1
h) PaymentMethod_Electronic check: 1
i) Online Security: 0
j) InternetService: 0
k) Contract: 0
l) Device Protection: 0

m) Tech Support: 0

n) Streaming TV: 0

o) Paperless Billing

2) Parameter Perhitungan

Prediksi Awal (F_0): 0.5 (Log-odds = 0)

Learning Rate (n): 0.1

Regulasi (λ): 1

3) Langkah Perhitungan Manual (Tree 1)

Dalam satu pohon, *XGBoost* akan mengevaluasi semua variabel untuk mencari *Gain* tertinggi. Misalkan hasil evaluasi menunjukkan bahwa kombinasi variabel *Contract* dan *tenure* memberikan pemisahan terbaik.

a. Hitung *Gradient (g)* dan *Hessian (h)*

Diambil sampel 3 baris pertama untuk menghitung skor akar (*Root*):

1. Baris 1 (y = 0): g_1 = 0.5-0 = 0.5; h_1 = 0.25
2. Baris 2 (y = 0): g_2 = 0.5-0 = 0.5; h_2 = 0.25
3. Baris 3 (y = 1): g_3 = 0.5-1 = -0.5; h_3 = 0.25

Total: $\sum g$= 0.5, $\sum h$ = 0.75

b. Evaluasi Split pada Fitur Kategorikal (Contract_Month-to-month)

XGBoost mengecek variabel ini:

Jika *Contract_Month-to-month* == 1 (Baris 1 & 3):

1. $\sum g$ = 0.5 + (-0.5) = 0
2. $\sum h$ = 0.25 + 0.25 = 0.5
3. $Sim_{Right} = \frac{0^2}{0.5+1} = 0$

Jika Contract_Month-to-month == 0 (Baris 2):

1. $\sum g = 0.5$
2. $Sim_{Right} = \frac{0.5^2}{0.25+1} = 0.2$

Gain: 0 + 0.2 - 0.1428 = 0.0572

c. Hitung *Output Value (Leaf Weight)*

Untuk Pelanggan 1 yang memiliki *Contract*_Month-to-month = 1, ia masuk ke Daun Kiri. Namun, karena *g* di Daun Kiri saling meniadakan (0.5 dan -0.5), model akan mencari variabel lain di level bawahnya, misalnya *OnlineSecurity_No.*

Misalkan setelah mengevaluasi semua variabel, Pelanggan 1 berakhir di sebuah daun dengan:

1. $\sum g = 0.5$ (hanya dia sendiri di daun tersebut)
2. $\sum h = 0.25$
3. *Weight* (w): $= \frac{0.5}{0.25+1} = -0.4$

4. Prediksi Akhir (Logit ke Probabilitas)

Nilai prediksi diperbarui dengan menjumlahkan kontribusi dari seluruh variabel yang terpilih dalam struktur pohon tersebut:

1. Update Log-odds:

   New Log-odds = 0 + (0.1 x -0.4) = -0.04

2. Konversi ke Probabilitas (Sigmoid):

   $P \frac{1}{1+e^{-(0.04)}} = \frac{1}{1+1.0408} = 0.49$

Berdasarkan rumus diatas, meskipun kita memasukkan puluhan variabel (seperti gender, *InternetService*, dsb), *XGBoost* secara otomatis akan memberikan bobot tinggi pada variabel yang paling berpengaruh (seperti *Contract* dan *tenure*) dan bobot mendekati nol untuk variabel yang tidak relevan. Hasil akhir 0.49 menunjukkan bahwa pelanggan tersebut memiliki probabilitas 49% untuk *churn* (berdasarkan pohon pertama), dan nilai ini akan terus disempurnakan oleh 99 pohon berikutnya dalam model.

4.7. Hasil Pengujian Fungsionalitas Sistem (Black Box Testing)

Pengujian *black box* dilakukan untuk memastikan seluruh fitur pada antarmuka web FastAPI berjalan sesuai dengan logika bisnis yang telah dirancang.

Tabel 4.1 Hasil Pengujian Fungsionalitas Sistem

No	Fitur / Fungsi	Butir Uji	Hasil yang Diharapkan	Status
1	Prediksi Manual	Input data profil pelanggan melalui form.	Sistem menampilkan label *churn* dan probabilitas secara sinkron.	Berhasil
2	Strategi Retensi	Logika pada retention_logic.py.	Muncul rekomendasi tindakan (Diskon/Layanan) berdasarkan tingkat risiko.	Berhasil
3	Upload Dataset	Unggah file CSV pada menu *Upload*.	Data masuk ke folder datasets/ dan tersimpan ke Neon Database.	Berhasil
4	Pelatihan Ulang	Klik tombol *Retrain Model*.	Sistem menjalankan proses training di *background* dan memperbarui file .pkl.	Berhasil
5	Analisis Visual	Halaman analysis.html.	Menampilkan grafik *Feature Importance* dan *Confusion Matrix* secara dinamis.	Berhasil
6	Riwayat Prediksi	Menu *Prediction History*.	Menampilkan daftar seluruh hasil prediksi yang ditarik dari tabel PostgreSQL.	Berhasil
7	Integrasi AI	Konsumsi API Gemini.	Muncul narasi saran komunikasi agen yang dihasilkan oleh Generative AI.	Berhasil

No	Fitur / Fungsi	Butir Uji	Hasil yang Diharapkan	Status
8	Koneksi Database	Integrasi Neon Database.	Data tetap tersimpan secara persisten meskipun aplikasi dimulai ulang.	Berhasil

Hasil pengujian sistem pada tabel 5 menggunakan black box testingmemberikan kesimpulan dbahwa semua menu telah dilakukan uji coba dengan detail pengujian yang berbeda-beda sesuai fungsi yang ingin diterapkan dengan statuss berhasil.

4.8. Kelebihan dan Kekurangan Website

Tabel 4.2 Kelebihan Dan Kekurangan Website

No	Keterangan	Kelebihan	Kekurangan
1	Fungsionalitas & Fitur	Tidak hanya memprediksi, tetapi memberikan rekomendasi strategi retensi preskriptif yang dipersonalisasi menggunakan Gemini AI. Memiliki fitur *Deep Dive SMOTE* untuk transparansi teknis.	Terbatas pada pemberian rekomendasi tekstual; belum bisa melakukan eksekusi otomatis seperti pengiriman email penawaran langsung ke pelanggan.
2	Teknologi & Performa	Menggunakan arsitektur *cloud-native* (FastAPI dan Neon Database) yang mendukung skalabilitas tinggi dan performa asinkron. Model memiliki tingkat *Recall* tinggi (82,1%) yang krusial untuk deteksi *churn*.	Sistem hanya mengimplementasikan satu algoritma tunggal (XGBoost), sehingga pengguna tidak bisa membandingkan performa dengan algoritma lain secara langsung di antarmuka.
3	Data & Validasi	Penanganan data tidak seimbang dilakukan secara tepat menggunakan SMOTE dalam *pipeline* untuk mencegah *data leakage*.	Penelitian ini hanya menggunakan satu sumber data sekunder (Kaggle), sehingga belum teruji pada dinamika pasar nyata atau data internal perusahaan secara langsung.
4	User Experience (UX)	*Dashboard* dirancang interaktif untuk memudahkan pemangku kepentingan non-teknis	Variabel input pada *dashboard* cukup banyak, yang mungkin memerlukan waktu bagi pengguna untuk

No	Keterangan	Kelebihan	Kekurangan
		memahami hasil analisis dan *feature importance*.	mengisi formulir secara manual jika tidak menggunakan fitur unggah CSV.
5	Inovasi	Integrasi teknologi Generative AI membuat saran komunikasi agen menjadi lebih humanis dan kontekstual dibandingkan sistem kaku konvensional.	Belum menyertakan data kualitatif eksternal seperti log keluhan pelanggan atau sentimen media sosial yang bisa memperkuat akurasi prediksi.

BAB V

PENUTUP

5.1. Kesimpulan

Beberapa kesimpulan yang telah diperoleh dari hasil pembahasan mengenai prediksi customer *churn* pada industri telekomunikasi untuk mendukung strategi retensi pelanggan menggunakan algoritma *XGBoost* yaitu:

1. Model prediksi *customer churn* yang optimal telah berhasil dibangun menggunakan algoritma *XGBoost* dengan mengintegrasikan metode *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih untuk menangani ketidakseimbangan data. Pendekatan ini terbukti mampu membuat model mengenali pola pelanggan yang akan *churn* tanpa mengalami bias terhadap kelas mayoritas.
2. Tingkat kinerja model *XGBoost* menunjukkan hasil yang sangat handal dengan nilai akurasi mencapai 87,4% dan nilai *Recall* sebesar 82,1%. Hasil evaluasi ini membuktikan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga efektif dalam meminimalkan kesalahan deteksi pelanggan yang berisiko berhenti berlangganan (Negatif Palsu).
3. Sistem informasi berbasis web telah berhasil diimplementasikan menggunakan *framework* FastAPI yang mampu menyajikan hasil prediksi dan analisis faktor penyebab *churn* secara sinkron. Penggunaan Neon Database sebagai media penyimpanan *cloud* memungkinkan terjadinya mekanisme sinkronisasi data pelanggan dan versi model secara otomatis, sehingga hasil prediksi yang diberikan bersifat relevan dan konsisten sesuai dengan data terbaru yang masuk ke dalam sistem.

4. Hasil prediksi dan analisis faktor pengaruh (seperti jenis kontrak dan masa berlangganan) telah berhasil dimanfaatkan sebagai dasar perumusan strategi retensi. Integrasi dengan teknologi *Generative AI* (Gemini API) dalam sistem ini mampu memberikan rekomendasi strategi tindak lanjut yang spesifik bagi setiap profil pelanggan guna meningkatkan loyalitas mereka.

5.2. Saran

Berdasarkan batasan masalah dan hasil penelitian, saran yang dapat diberikan untuk pengembangan selanjutnya adalah:

1. Mengingat penelitian ini dibatasi pada penggunaan algoritma XGBoost, penelitian selanjutnya disarankan untuk melakukan komparasi performa dengan algoritma *ensemble* lain atau *Deep Learning* guna mengeksplorasi potensi peningkatan akurasi lebih lanjut.
2. Penelitian selanjutnya dapat memperluas cakupan dataset di luar data sekunder Kaggle, seperti menambahkan data kualitatif berupa log keluhan pelanggan atau data sentimen media sosial, untuk menangkap faktor eksternal yang tidak terakomodasi dalam penelitian ini.
3. Sistem informasi yang dibangun dapat dikembangkan lebih lanjut dengan menambahkan fitur eksekusi otomatis, seperti integrasi langsung ke layanan pengiriman notifikasi/email penawaran atau menghubungkan sistem secara *live* dengan basis data operasional perusahaan telekomunikasi sesungguhnya.

DAFTAR PUSTAKA

Alotaibi, M. Z., & Haq, M. A. (2024). Customer *churn* prediction for telecommunication companies using *machine learning* and ensemble methods. *Engineering, Technology & Applied Science Research*, *14*(3). https://doi.org/10.48084/etasr.7480

Amin, A., Anwar, S., & Adnan, A. (2022). Customer *churn* prediction in telecommunication sector using *machine learning* techniques. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-021-06197-6

Asyhari, M. Y., Susanti, P., & Yunitasari, Y. (2025). *Implementation of Machine learning Models for Predicting Internet Service Provider Customer Churn*. *7*(4), 1164–1172.

Chang, V., Hall, K., Xu, Q. A., Amao, F. O., Ganatra, M. A., & Benson, V. (2024). *Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine learning Models*.

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree *boosting* system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*(August 2016), 785–794. https://doi.org/10.1145/2939672.2939785

Dhangar, K., & Anand, P. (2021). *A REVIEW ON CUSTOMER CHURN PREDICTION USING MACHINE*. *8*(5), 193–201.

Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive Analysis

of Random Forest and *XGBoost* Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels. *Technologies*, *13*(3), 1–40. https://doi.org/10.3390/technologies13030088

Peng, K., & Peng, Y. (2022). *Research on Telecom Customer Churn Prediction Based on GA-XGBoost and SHAP*. 107–120. https://doi.org/10.4236/jcc.2022.1011008

Sam, G., Asuquo, P., & Stephen, B. (2024). *Customer Churn Prediction using Machine learning Models*. *26*(2), 181–193. https://doi.org/10.9734/JERR/2024/v26i21081

Schröer, C., Kruse, F., Marx, J., Kruse, F., & Marx, J. (2021). ScienceDirect ScienceDirect A Systematic Literature Review A Systematic Literature Review on Applying Process Model on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*(2019), 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Shaikhsurab, M. A., & Magadum, P. (2017). *Enhancing Customer Churn Prediction in Telecommunications : An Adaptive Ensemble Learning Approach Abstract :*

Wakhidah, L. N., Zyen, A. K., & Wahono, B. B. (2025). Evaluation of Telecommunication Customer *Churn* Classification with SMOTE Using Random Forest and *XGBoost* Algorithms. *Journal of Applied Informatics and Computing*, *9*(1), 89–95. https://doi.org/10.30871/jaic.v9i1.8740

Asosiasi Penyelenggara Jasa Internet Indonesia. (2024). *Laporan Survei Penetrasi Internet Indonesia 2024*. https://apjii.or.id/survei

Vodafone: Vodafone Group Plc. (2023). *Annual Report 2023*. https://investors.vodafone.com/reports-and-presentations/annual-reports/2023

AT&T: AT&T Inc. (2023). *2023 Annual Report*. https://investors.att.com/financial-reports/annual-reports/2023

GSMA. (2023). *The Mobile Economy Asia Pacific 2023*. https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/asiapacific/

Bisht, P. S., & Anjaria, B. (2025). *Customer Relationship Management (CRM): Evolution*. ResearchGate. https://doi.org/10.13140/RG.2.2.22338.49609

Hermawan, A., Jayanti, N. R., Tabaruk, Z., Triadi, F. L. Y., Saputra, A., & Syachrudin, M. R. H. (2024). *Membangun Model Prediksi Churn Pelanggan yang Akurat: Studi Kasus tentang TELCO Company*. Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika, 2(6), 67–81. https://doi.org/10.61132/merkurius.v2i6.398