

MODEL PREDIKSI CHURN PELANGGAN INDIBIZ MENGGUNAKAN REGRESI LOGISTIK DAN *EXTREME GRADIENT BOOSTING* (XGBOOST)

Aulia Rahmadiva Wardana¹, Woro Isti Rahayu², Kiki Mustaqim³

^{1,2,3} Sains Data, Fakultas Logistik Teknologi dan Bisnis, Universitas Logistik dan Bisnis Internasional, Indonesia
¹auliarahmadiva4949@gmail.com, ²woroisti@ulbi.ac.id, ³kiki@ulbi.ac.id

Abstrak

Kebutuhan digitalisasi yang semakin pesat mendorong PT. Telekomunikasi Indonesia (Persero) Tbk menghadirkan IndiBiz sebagai layanan internet berbasis fiber untuk mendukung transformasi digital UMKM. Namun, meningkatnya persaingan industri telekomunikasi menimbulkan risiko churn pelanggan yang berpotensi menurunkan pendapatan dan loyalitas. Penelitian ini bertujuan membangun dan membandingkan model prediksi churn menggunakan Regresi Logistik dan Extreme Gradient Boosting (XGBoost), mengidentifikasi tingkat risiko churn pelanggan, serta mengembangkan dashboard interaktif berbasis Streamlit sebagai pendukung pengambilan keputusan. Metode yang digunakan adalah klasifikasi biner pada data pelanggan IndiBiz dengan evaluasi menggunakan akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa XGBoost memiliki performa lebih baik dengan akurasi 83%, recall 0,58, dan F1-score yang lebih tinggi dibandingkan Regresi Logistik (akurasi 81% dan recall 0,56). Namun, nilai precision yang sangat tinggi dan recall yang relatif rendah pada kedua model mengindikasikan kecenderungan model yang konservatif serta kemungkinan ketidakseimbangan kelas, sehingga masih terdapat pelanggan churn yang tidak terdeteksi. Model XGBoost mengelompokkan pelanggan ke dalam kategori risiko churn tinggi (14.255 pelanggan), sedang (3.418 pelanggan), dan rendah (889 pelanggan). Interpretasi menggunakan SHAP menunjukkan bahwa Lama_Berlangganan_Bulan, PAKET_DIGI, dan STO merupakan faktor utama yang memengaruhi churn. Implementasi model dalam dashboard berbasis Streamlit memungkinkan eksplorasi data dan prediksi secara interaktif. Penelitian ini berkontribusi dalam mendukung strategi retensi pelanggan berbasis data, meskipun masih diperlukan pengembangan lebih lanjut untuk meningkatkan kemampuan deteksi churn.

Kata kunci: churn pelanggan, regresi logistik, XGBoost, Indibiz, telekomunikasi

1. Pendahuluan

Kebutuhan digitalisasi yang semakin pesat telah mendorong perusahaan telekomunikasi untuk menyediakan layanan yang inovatif dan andal. Salah satu layanan unggulan dari PT. Telekomunikasi Indonesia (Persero) Tbk adalah IndiBiz. IndiBiz merupakan layanan konektivitas digital berbasis jaringan fiber yang dirancang untuk mendukung aktivitas usaha mikro, kecil, dan menengah (UMKM) dalam meningkatkan efisiensi operasional dan daya saing bisnis di era ekonomi digital (Putra, 2023). Layanan ini hadir sebagai respons terhadap meningkatnya kebutuhan UMKM untuk terhubung secara digital, mulai dari aktivitas pemasaran, transaksi daring, hingga kolaborasi kerja jarak jauh. Dengan memberikan konektivitas yang stabil dan layanan yang sesuai dengan kebutuhan usaha skala kecil-menengah, IndiBiz diharapkan mampu mendorong transformasi digital UMKM serta memperluas jangkauan digitalisasi ke seluruh pelosok daerah.

Namun, seiring meningkatnya kompetisi di industri telekomunikasi dan banyaknya pilihan penyedia layanan, perusahaan menghadapi tantangan

berupa churn, churn merupakan proses di mana pelanggan memutuskan jasa yang diberikan oleh suatu perusahaan dan beralih ke layanan dari pesaing (Kishendrian et al., 2023). Fenomena ini menjadi indikator penting dalam mengevaluasi kualitas layanan, efektivitas strategi pemasaran, serta kepuasan pelanggan secara umum. Churn yang tinggi tidak hanya mengurangi pendapatan, tetapi juga menunjukkan adanya ketidaksesuaian antara ekspektasi pelanggan dengan layanan yang diterima.

Dalam konteks prediksi churn pelanggan, berbagai penelitian terdahulu telah menerapkan metode klasifikasi seperti Regresi Logistik dan XGBoost. Regresi Logistik banyak digunakan karena kesederhanaannya serta kemampuannya dalam memberikan interpretasi yang jelas terhadap pengaruh variabel input terhadap probabilitas churn pelanggan (Id et al., 2024). Sementara itu, XGBoost sebagai model berbasis ensemble telah terbukti efektif dalam menangani pola data pelanggan yang kompleks dan tidak seimbang pada kasus prediksi churn di sektor telekomunikasi (Kingawa & Hailu, 2022). Meskipun demikian, penelitian yang secara khusus membandingkan kinerja Regresi Logistik dan XGBoost pada layanan internet bisnis berbasis fiber

seperti IndiBiz masih terbatas, sehingga diperlukan kajian empiris untuk menentukan metode yang paling sesuai dengan karakteristik data pelanggan IndiBiz.

Selain keterbatasan pada objek dan perbandingan metode, penelitian-penelitian sebelumnya umumnya lebih menitikberatkan pada pencapaian nilai akurasi model, tanpa disertai analisis mendalam terhadap faktor-faktor yang memengaruhi churn pelanggan. Padahal, identifikasi faktor penentu churn merupakan aspek penting dalam mendukung perumusan strategi retensi pelanggan yang efektif. Model dengan performa tinggi, seperti XGBoost, memiliki kompleksitas yang lebih besar sehingga membutuhkan pendekatan interpretasi tambahan agar hasil prediksi dapat dipahami dan dimanfaatkan secara praktis (Chen, n.d.). Dengan demikian, terdapat celah penelitian dalam mengintegrasikan analisis perbandingan performa model dengan identifikasi faktor-faktor utama penyebab churn pelanggan, khususnya pada layanan IndiBiz, sehingga hasil penelitian tidak hanya bersifat teknis tetapi juga memberikan nilai analitis dan aplikatif bagi pengambilan keputusan bisnis.

Berdasarkan research gap tersebut, penelitian ini bertujuan untuk membangun dan membandingkan model prediksi churn pelanggan IndiBiz menggunakan Regresi Logistik dan XGBoost guna mengevaluasi kinerja kedua metode dalam mendeteksi pelanggan yang berisiko churn. Selain membandingkan performa model berdasarkan metrik evaluasi yang relevan, penelitian ini juga berfokus pada identifikasi faktor-faktor utama yang memengaruhi churn pelanggan melalui pendekatan interpretasi model. Kontribusi ilmiah penelitian ini terletak pada penyediaan analisis komparatif antara metode klasik dan ensemble pada konteks layanan internet bisnis berbasis fiber, serta pada pengungkapan faktor penentu churn yang bersifat kontekstual terhadap layanan IndiBiz. Lebih lanjut, penelitian ini memberikan kontribusi praktis melalui pengelompokan pelanggan berdasarkan tingkat risiko churn dan pengembangan dashboard interaktif berbasis Streamlit sebagai alat bantu pengambilan keputusan bagi Divisi Business Service PT. Telekomunikasi Indonesia.

2. Metode

2.1 Alur Metodologi Penelitian

Gambar 1 menunjukkan alur metode penelitian yang mana diawali dengan pengumpulan data pelanggan IndiBiz yang bersumber dari data internal perusahaan, mencakup status churn serta karakteristik pelanggan seperti lama berlangganan dan jenis layanan.

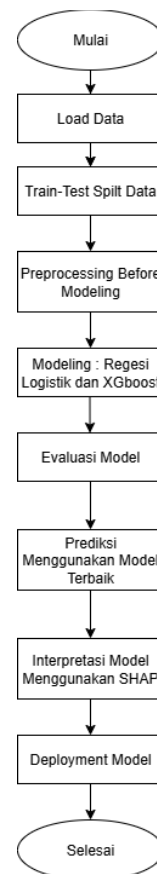


Gambar 1. Alur Metodologi Penelitian

Data yang telah dikumpulkan selanjutnya melalui tahap pra-pemrosesan untuk memastikan kualitas data sebelum digunakan dalam pemodelan (Amanda & Ridwan Yusuf, 2025). Pada tahap pemodelan, Regresi Logistik dan *Extreme Gradient Boosting* (XGBoost) diterapkan untuk membangun model prediksi churn, mengingat kedua metode telah terbukti efektif dalam penelitian (A. Hermawan et al., 2025). Kinerja kedua model kemudian dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score guna menentukan model dengan performa terbaik. Tahap akhir penelitian adalah penarikan kesimpulan dan penyusunan rekomendasi strategis berdasarkan hasil prediksi churn pelanggan IndiBiz.

2.2 Alur Metode Regresi Logistik dan XGBoost

Gambar 2 menunjukkan alur prediksi churn menggunakan 2 metode yaitu Regresi Logistik dan XGBoost pada bahasa pemrograman Python.



Gambar 2. Diagram alur metode Regresi Logistik dan XGBoost

Tiap bagian dari diagram alur tersebut dapat dijelaskan sebagai berikut:

1. Load Data

Data yang digunakan merupakan data gabungan data pelanggan keseluruhan dan data pelanggan yang churn saja, data ini juga telah melewati preprocessing. Data yang digunakan yaitu Order_Id, No_Inet, Customers_name, STO, Paket_Digi, Periode_akhir, L_Ekosistem,

- L_Produk, Lama_Berlangganan_Bulan, Status_Churn.
2. *Train-Test Split Data*
Train-test split adalah metode membagi dataset menjadi dua bagian, yaitu data pelatihan (training) yang digunakan untuk membangun model dan data pengujian (testing) yang digunakan untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya (Umam, 2024). Pada penelitian ini data dibagi menjadi menjadi 85% data latih dan 15% data uji. Kemudian, dari data latih tersebut, 18% kembali dipisahkan sebagai data validasi. Dengan demikian, komposisi akhir data menjadi sekitar 70% untuk pelatihan, 15% untuk validasi, dan 15% untuk pengujian.
 3. *Processing Before Modeling*
Tahap ini berfokus pada pembersihan dan transformasi data agar model dapat memahami informasi yang terkandung dalam fitur. Pada tahap preprocessing ini, dilakukan transformasi terhadap fitur kategorikal dan numerikal agar data siap digunakan oleh model machine learning. Fitur kategorikal seperti *STO*, *PAKET_DIGI*, *L_PRODUK*, dan *L_EKOSISTEM* dikodekan menggunakan teknik *One-Hot Encoding* agar dapat direpresentasikan dalam format numerik biner. Sementara (Dhani et al., n.d.) itu, fitur numerikal seperti *Lama_Berlangganan_Bulan* dilakukan normalisasi dengan *StandardScaler* untuk menyamakan skala antar fitur. Semua proses ini dikemas dalam *ColumnTransformer*, yang kemudian diterapkan pada data latih, validasi, dan uji secara konsisten. Output dari proses ini berupa *DataFrame* baru yang sudah bersih, distandarkan, dan siap digunakan dalam pelatihan dan evaluasi model prediksi churn.
 4. Modeling
 - a. Regresi Logistik
Regresi logistik merupakan metode statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen dengan variabel dependen yang bersifat kategorikal, khususnya biner, melalui pendekatan probabilistik (Surabaya, 2025). Pada bagian ini, model regresi logistik dibangun menggunakan *LogisticRegression* dari pustaka *sklearn.linear_model*. Parameter *solver='liblinear'* digunakan karena cocok untuk dataset skala kecil hingga sedang dan mendukung klasifikasi biner. Model kemudian dilatih (fit) menggunakan data latih (*X_train_processed* dan *y_train*) yang telah melalui tahap preprocessing sebelumnya.
 - b. XGBoost
XGBoost atau *Extreme Gradient Boosting* XGBoost adalah model gradient boosting yang kuat dalam menyelesaikan tugas klasifikasi dan regresi karena kemampuannya menangkap hubungan kompleks antar fitur dan data yang tidak seimbang, sehingga sering kali memberikan performa yang lebih baik dibandingkan metode tradisional dalam konteks klasifikasi (Hidayat et al., 2024). Model ini melatih pohon keputusan secara berurutan pada data pelatihan. Untuk meningkatkan nilai dari fungsi tujuan, algoritma ini menambahkan pohon keputusan baru ke pohon sebelumnya pada setiap iterasi (Aydin, 2021). Model XGBoost dilatih menggunakan *XGBClassifier* dengan parameter regularisasi dan pembelajaran bertahap untuk menghindari overfitting serta meningkatkan akurasi prediksi pada data churn.
 5. Evaluasi Model
Evaluasi model bertujuan mengukur seberapa baik model dapat melakukan klasifikasi pada data yang belum pernah dilihat sebelumnya. Beberapa metrik umum yang digunakan adalah akurasi, precision, recall, dan F1-score (Dhani et al., n.d.). Evaluasi ini dilakukan baik pada data validasi maupun data uji untuk memastikan bahwa model tidak hanya baik dalam pelatihan, tetapi juga dapat digeneralisasi ke data baru (Saputra & Sugihartono, 2025). Hasil evaluasi menjadi dasar untuk memilih model terbaik.
 6. Prediksi Menggunakan Model Terbaik
Diantara model XGBoost dan Regresi Logistik dari nilai evaluasi modelnya, model terbaik diambil untuk memprediksi churn, dimana akan ditampilkan nilai probabilitas churn tiap pelanggan.
 7. Interpretasi Model Menggunakan SHAP
Selain evaluasi kuantitatif, dilakukan juga interpretasi hasil model menggunakan teknik SHAP (*SHapley Additive exPlanations*). SHAP digunakan untuk memberikan interpretasi terhadap model machine learning dengan menunjukkan kontribusi setiap fitur terhadap hasil prediksi, sehingga meningkatkan transparansi dalam pemahaman keputusan model (Iffadah & Prasetya, 2025).
 8. Deployment Model
Model terbaik yang telah teruji kemudian disimpan dan di-deploy ke dalam *dashboard* prediktif. Dashboard prediktif dikembangkan menggunakan framework Streamlit, yaitu sebuah alat berbasis Python yang memungkinkan pembuatan aplikasi web interaktif untuk menampilkan hasil prediksi dan visualisasi data model secara real-time (E. Hermawan et al., 2024). Model ini akan digunakan untuk memprediksi probabilitas churn pelanggan secara otomatis. Dalam proses

ini juga disertakan penyimpanan proses preprocessing agar prediksi pada data baru konsisten dengan data pelatihan.

3. Hasil dan Pembahasan

3.1 Perbandingan Regresi Logistik dan XGBoost

Tabel 1 menunjukkan bahwa model Regresi Logistik dan XGBoost memiliki performa yang relatif baik dalam memprediksi churn pelanggan. Model XGBoost menunjukkan performa yang lebih unggul dengan akurasi sebesar 83% dan recall sebesar 0,58, dibandingkan Regresi Logistik dengan akurasi 81% dan recall 0,56. Selain itu, nilai F1-score pada model XGBoost juga lebih tinggi dibandingkan Regresi Logistik, yang menunjukkan bahwa XGBoost memiliki keseimbangan yang lebih baik antara precision dan recall dalam mendeteksi churn pelanggan.

Tabel 1. Perbandingan Hasil Model Regresi Logistik dan XGBoost

Model	Dataset	Akurasi	Precision	Recall	F1-Score
Regresi Logistik	Validasi	80,35%	0,999	0,532	0,693
Regresi Logistik	Uji	81,35%	0,999	0,542	0,702
XGBoost	Validasi	82%	1,00	0,56	0,72
XGBoost	Uji	83%	1,00	0,58	0,74

Namun demikian, terdapat fenomena yang perlu dianalisis lebih lanjut, yaitu nilai precision yang sangat tinggi (mendekati 1,00) pada kedua model, sementara nilai recall masih relatif rendah. Kondisi ini mengindikasikan bahwa model cenderung bersifat konservatif dalam memprediksi churn, di mana model hanya mengklasifikasikan pelanggan sebagai churn ketika tingkat keyakinannya sangat tinggi.

Selain itu, perbedaan yang cukup signifikan antara precision dan recall juga mengindikasikan kemungkinan adanya ketidakseimbangan kelas (*class imbalance*), di mana jumlah pelanggan tidak churn lebih dominan dibandingkan pelanggan churn. Hal ini menyebabkan model lebih fokus pada pengurangan kesalahan prediksi positif (*false positive*), namun kurang optimal dalam menangkap seluruh pelanggan yang benar-benar churn.

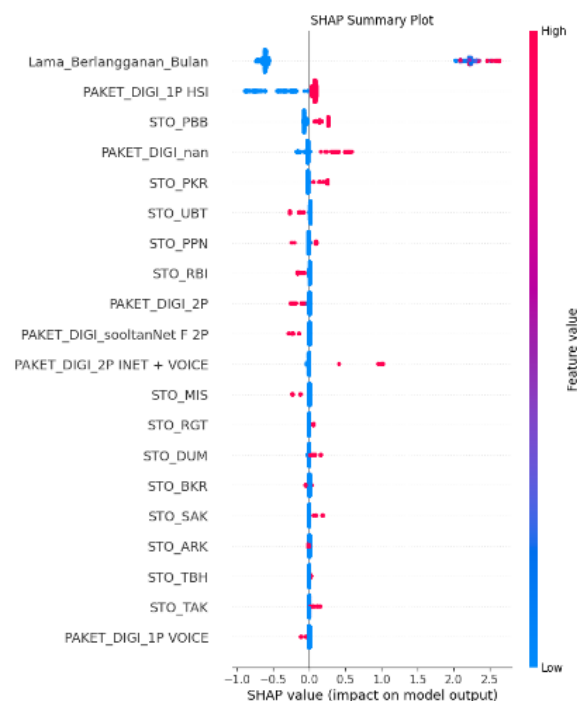
Dalam konteks bisnis, nilai recall yang relatif rendah memiliki implikasi yang cukup penting. Model yang memiliki recall rendah berpotensi melewatkan sejumlah pelanggan yang sebenarnya berisiko churn (*false negative*), sehingga perusahaan kehilangan peluang untuk melakukan intervensi atau strategi retensi secara dini. Dengan kata lain, meskipun prediksi churn yang dihasilkan sangat akurat (precision tinggi), model belum sepenuhnya efektif dalam mendukung strategi pencegahan churn secara menyeluruh.

Secara keseluruhan, meskipun XGBoost menunjukkan performa yang lebih baik dibandingkan Regresi Logistik, hasil evaluasi ini menunjukkan bahwa peningkatan recall serta keseimbangan antara precision dan recall yang tercermin pada nilai F1-

score masih menjadi aspek penting yang perlu diperhatikan agar model lebih optimal dalam mendukung pengambilan keputusan bisnis yang bersifat proaktif.

3.2 Interpretasi SHAP

Selanjutnya, melihat SHAP (*SHapley Additive exPlanations*) Summary Plot untuk menunjukkan kontribusi masing-masing fitur terhadap prediksi churn pada seluruh pelanggan.



Gambar 3. Boxplot hasil model XGBoost setelah tuning

Hasil interpretasi menggunakan metode SHAP menunjukkan bahwa fitur Lama_Berlangganan_Bulan merupakan faktor yang paling dominan dalam memengaruhi prediksi churn pelanggan. Pelanggan dengan masa berlangganan yang lebih singkat cenderung memiliki probabilitas churn yang lebih tinggi dibandingkan pelanggan dengan masa berlangganan yang lebih lama.

Temuan ini mengindikasikan bahwa fase awal siklus pelanggan merupakan periode yang krusial dalam menentukan loyalitas pelanggan. Oleh karena itu, perusahaan perlu memfokuskan strategi retensi pada pelanggan baru, misalnya melalui program onboarding yang lebih optimal, peningkatan kualitas layanan di awal penggunaan, serta pemberian insentif atau promosi khusus untuk meningkatkan engagement pelanggan.

Selain itu, fitur PAKET_DIGI juga menunjukkan pengaruh yang signifikan terhadap churn. Hal ini menunjukkan bahwa kesesuaian antara paket layanan yang dipilih dengan kebutuhan pelanggan menjadi faktor penting dalam mempertahankan pelanggan. Oleh karena itu, perusahaan perlu melakukan evaluasi terhadap variasi paket layanan yang ditawarkan serta

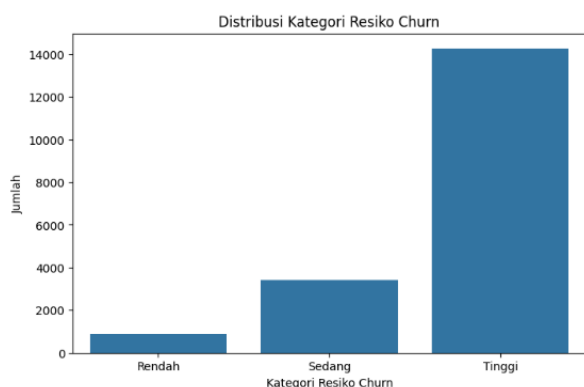
mempertimbangkan personalisasi layanan berdasarkan karakteristik pelanggan.

Fitur STO juga berkontribusi terhadap prediksi churn, yang mengindikasikan adanya perbedaan risiko churn berdasarkan wilayah atau lokasi layanan. Hal ini dapat berkaitan dengan kualitas infrastruktur jaringan, stabilitas koneksi, maupun kualitas layanan di masing-masing wilayah. Oleh karena itu, perusahaan perlu melakukan analisis lebih lanjut secara geografis untuk mengidentifikasi wilayah dengan tingkat churn tinggi dan melakukan peningkatan kualitas layanan secara terarah.

Dengan demikian, hasil interpretasi SHAP tidak hanya memberikan pemahaman terhadap faktor-faktor yang memengaruhi churn, tetapi juga dapat digunakan sebagai dasar dalam merumuskan strategi retensi pelanggan yang lebih efektif dan berbasis data.

3.3 Persebaran Pelanggan Churn

Selanjutnya, model terbaik digunakan untuk memprediksi churn pelanggan, di mana setiap pelanggan memperoleh skor probabilitas churn pada rentang 0,0 hingga 1,0. Nilai probabilitas tersebut kemudian dikelompokkan ke dalam kategori risiko berdasarkan pendekatan kuartil, sehingga diperoleh hasil pengelompokan risiko churn pelanggan sebagaimana disajikan pada Gambar 4.



Gambar 4. Grafik Persebaran Risiko Churn

Pada Gambar 4 terlihat hasil prediksi model XGBoost, setiap pelanggan memperoleh nilai probabilitas churn yang kemudian dikelompokkan ke dalam tiga kategori risiko, yaitu tinggi, sedang, dan rendah. Pengelompokan ini dilakukan menggunakan pendekatan kuartil untuk membagi distribusi probabilitas churn secara proporsional.

Namun demikian, pendekatan berbasis kuartil ini bersifat statistik dan belum sepenuhnya mempertimbangkan konteks bisnis, seperti kapasitas tim dalam melakukan intervensi atau nilai ekonomi dari masing-masing pelanggan. Oleh karena itu, batasan antar kategori risiko yang dihasilkan masih berpotensi bersifat arbitrer.

Meskipun demikian, hasil pengelompokan menunjukkan bahwa sebagian besar pelanggan berada pada kategori risiko churn tinggi, yaitu sebanyak 14.255 pelanggan, diikuti oleh 3.418

pelanggan pada kategori sedang, dan 889 pelanggan pada kategori rendah. Distribusi ini menunjukkan bahwa perusahaan perlu memprioritaskan strategi retensi pada kelompok pelanggan berisiko tinggi dengan pendekatan yang lebih proaktif.

Ke depan, pendekatan pengelompokan risiko churn dapat dikembangkan dengan mempertimbangkan aspek *business threshold* atau *cost-sensitive analysis*, sehingga kategori risiko yang dihasilkan lebih relevan dengan kebutuhan strategis perusahaan.

3.4 Implikasi Strategis

Berdasarkan hasil analisis model dan interpretasi fitur, strategi retensi pelanggan dapat difokuskan pada beberapa aspek utama. Pertama, pelanggan dengan masa berlangganan yang rendah perlu mendapatkan perhatian khusus melalui pendekatan onboarding dan peningkatan pengalaman awal penggunaan layanan. Kedua, evaluasi terhadap paket layanan perlu dilakukan untuk memastikan kesesuaian dengan kebutuhan pelanggan. Ketiga, perusahaan perlu melakukan pemantauan dan peningkatan kualitas layanan pada wilayah tertentu yang memiliki tingkat churn tinggi.

Dengan memanfaatkan hasil prediksi churn secara optimal, perusahaan dapat mengimplementasikan strategi retensi yang lebih terarah, efisien, dan berbasis data, sehingga mampu mengurangi tingkat churn dan meningkatkan loyalitas pelanggan.

3.5 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, model yang digunakan belum secara khusus menangani ketidakseimbangan kelas (*class imbalance*), yang dapat memengaruhi nilai recall dalam mendeteksi pelanggan churn. Kedua, penentuan *threshold* klasifikasi masih menggunakan nilai default, sehingga belum dioptimalkan untuk meningkatkan keseimbangan antara *precision* dan *recall*.

Selain itu, fitur yang digunakan dalam penelitian ini masih terbatas pada data statis pelanggan, sehingga belum sepenuhnya merepresentasikan perilaku dinamis pelanggan, seperti frekuensi penggunaan layanan atau interaksi pelanggan dengan perusahaan. Keterbatasan lainnya adalah pendekatan pengelompokan risiko churn yang masih berbasis kuartil dan belum mempertimbangkan aspek bisnis secara mendalam.

Penelitian selanjutnya disarankan untuk mengatasi keterbatasan tersebut dengan menerapkan teknik penyeimbangan data, optimalisasi *threshold*, serta penggunaan fitur yang lebih kompleks dan dinamis untuk meningkatkan performa model.

3.6 Deploy Model Kedalam Dashboard

Untuk mendukung kebutuhan analisis dan pengambilan keputusan tim *Business Service*,

dibangun sebuah dashboard berbasis *Streamlit* yang bersifat interaktif dan mudah digunakan. Dashboard ini bertujuan untuk memvisualisasikan hasil prediksi churn, mengeksplorasi karakteristik pelanggan yang berisiko churn, serta menyediakan fitur prediksi manual terhadap pelanggan baru. Dengan pendekatan ini, tim non-teknis dapat lebih intuitif memahami pola perilaku pelanggan dan mengambil tindakan berbasis data secara cepat dan tepat.

Dashboard ini dapat diakses secara daring melalui tautan <https://churnindibiz-app.streamlit.app/>, sehingga memudahkan tim *Business Service* IndiBiz dalam melakukan pemantauan, analisis, dan pengambilan keputusan berbasis data secara real-time dan fleksibel di mana pun dan kapan pun diperlukan.

4. Kesimpulan

Hasil penelitian menunjukkan bahwa model XGBoost memiliki performa lebih baik dibandingkan Regresi Logistik dengan akurasi 83% dan recall 0,58, serta nilai F1-score yang lebih tinggi yang menunjukkan keseimbangan yang lebih baik antara precision dan recall. Meskipun demikian, nilai precision yang sangat tinggi dan recall yang relatif rendah pada kedua model mengindikasikan bahwa model cenderung konservatif serta berpotensi dipengaruhi oleh ketidakseimbangan kelas, sehingga masih terdapat pelanggan churn yang tidak terdeteksi.

Model XGBoost mengelompokkan 18.562 pelanggan ke dalam tiga kategori risiko churn, meskipun pendekatan berbasis kuartil yang digunakan masih bersifat statistik. Interpretasi SHAP menunjukkan bahwa Lama_Berlangganan_Bulan, PAKET_DIGI, dan STO merupakan faktor utama yang memengaruhi churn. Secara praktis, hasil ini dapat dimanfaatkan untuk mendukung strategi retensi pelanggan, khususnya pada pelanggan baru dan wilayah tertentu. Namun, penelitian ini masih memiliki keterbatasan, sehingga diperlukan pengembangan lebih lanjut untuk meningkatkan kemampuan deteksi churn secara lebih optimal.

Daftar Pustaka:

Amanda, A., & Ridwan Yusuf. (2025). ANALISIS SENTIMEN MASYARAKAT INDONESIA PADA PLATFROM X TERHADAP ISU FUFUFABA MENGGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMER. *Jurnal Informatika Teknologi Dan Sains (JINTEKS)*, 7, 72–80.

Aydin, Z. E. (2021). *Performance Analysis of XGBoost Classifier with Missing Data*. March.

Chen, T. (n.d.). *XGBoost: A Scalable Tree Boosting System*.

Dhani, C. R., Trisnawarman, D., Budiyantra, A., Studi, P., Informasi, S., Teknologi, F., Universitas, I., Barat, J., Teknologi, F., Universitas, I., & Barat, J. (n.d.).

MENINGKATKAN PENGELOLAAN INSIDEN LAYANAN TI. i, 1–5.

- Hermawan, A., Saputra, A., Rafi, M. D., & Basmallah, S. (2025). *Implementing XGBoost Model for Predicting Customer Churn in E-Commerce Platforms*. 3(2).
- Hermawan, E., Panjaitan, S. D., & Ripanti, E. F. (2024). *Sistem Prediksi Banjir Rob Kota Pontianak Berbasis Machine Learning Menggunakan Framework Streamlit*. 10(3), 351–361.
- Hidayat, R., Mahdiana, D., & Fergina, A. (2024). *Comparative Analysis of Logistic Regression, SVM, Xgboost, and Random Forest Algorithms for Diabetes Classification*. 7(1), 281–291. <https://doi.org/10.32493/jtsi.v7i1.38258>
- Id, M. I., Id, R. I., Kawakami, E., Michihata, N., Goto, T., Shinozaki, T., & Sun, Y. (2024). *PLOS DIGITAL HEALTH Comparison of machine-learning and logistic regression models for prediction of 30-day unplanned readmission in electronic health records: A development and validation study*. 1–16. <https://doi.org/10.1371/journal.pdig.0000578>
- Iffadah, A. S., & Prasetya, D. A. (2025). *Shapley Additive Explanations Interpretation of the XGBoost Model in Predicting Air Quality in Jakarta*. 7(3).
- Kingawa, E. D., & Hailu, T. T. (2022). *Customer Churn Prediction Using Machine Learning Techniques: the case of Lion Insurance*. 4(4), 60–73.
- Kishendrian, H., Hanum, N., Prianto, C., & Rahayu, W. I. (2023). Penerapan Metode Clustering Dalam Segmentasi Pelanggan Perusahaan Logistik. *SINTECH (Science and Information Technology) Journal*, 6(3), 137–146. <https://doi.org/10.31598/sintechjournal.v6i3.1432>
- Putra, I. S. (2023). *Inovasi dan Strategi Pemasaran Produk Indibiz di PT. Telkom Indonesia Prosiding Seminar Nasional Teknologi Komputer dan Sains*. 1(1), 299–306.
- Saputra, M. A., & Sugihartono, T. (2025). *Evaluasi Kinerja Model LSTM Untuk Prediksi Risiko Penyakit Jantung Menggunakan Dataset Program Studi Teknik Informasi, Fakultas Teknologi Informasi, ISB Atma Luhur, Indonesia Performance Evaluation of the LSTM Model for Heart Disease Risk Prediction Using a Dataset*. 5(7), 1823–1833.
- Surabaya, U. N. (2025). *Journal of Emerging Information Systems and Business Intelligence ISSN: 2774-3993*. 6(2), 230–240.
- Umam, K. (2024). *MENGANALISIS RESPONS NETIZEN TWITTER TERHADAP PROGRAM MAKAN SIANG GRATIS MENERAPKAN NLP METODE NAÏVE BAYES*. 14(3), 201–208.