

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perusahaan telekomunikasi berkembang dengan pesat seiring kemajuan perkembangan teknologi digital dan semakin ketatnya persaingan antara penyedia layanan. Di pasar yang sangat kompetitif ini, perusahaan perlu mempertahankan pelanggan lama dan mencari pelanggan baru. Retensi pelanggan sangat penting karena sangat berkaitan dengan stabilitas pendapatan dan keberlanjutan bisnis perusahaan [1].

Masalah utama yang dihadapi perusahaan telekomunikasi yaitu churn pelanggan atau situasi ketika pelanggan berhenti menggunakan suatu layanan dan beralih ke penyedia lain. Churn pelanggan tahunan di industri ini berkisar antara 15% hingga 25% dan biaya untuk mendapatkan pelanggan baru 5 hingga 10 lebih mahal daripada mempertahankan pelanggan yang sudah ada. Hal ini menunjukkan bahwa pendapatan dan pertumbuhan suatu perusahaan dapat sangat terpengaruh oleh churn pelanggan yang tinggi [2].

Informasi penting mengenai pola perilaku konsumen dan churn pelanggan dapat ditemukan dalam data historis pelanggan [3]. Penelitian ini menggunakan algoritma Logistic Regression, Random Forest, Support Vector Machine (SVM), Gradient Boosting, dan XGBoost. Gradient Boosting dan XGBoost kemampuan yang akurat dalam klasifikasi churn pelanggan, sedangkan Random Forest dan SVM mampu menangani pola data yang kompleks, sementara Logistic Regression mudah diinterpretasikan [4].

Berdasarkan permasalahan tersebut, penelitian ini membandingkan seberapa baik berbagai algoritma machine learning dalam memprediksi churn pelanggan di perusahaan telekomunikasi, menggunakan analisis interpretabilitas model untuk mengidentifikasi penyebab utama churn pelanggan, serta mengintegrasikan model terbaik ke dalam prototipe sistem prediksi berbasis web guna mendukung strategi retensi pelanggan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Seberapa akurat berbagai algoritma machine learning dalam memprediksi churn pelanggan pada layanan telekomunikasi berdasarkan metrik evaluasi seperti Accuracy, Precision, Recall, F1-Score, dan ROC-AUC?
2. Algoritma machine learning mana yang memiliki performa terbaik dalam memprediksi churn pelanggan berdasarkan metrik evaluasi seperti Accuracy, Precision, Recall, F1-Score, dan ROC-AUC?
3. Bagaimana hasil analisis interpretabilitas model menentukan faktor utama penyebab churn pelanggan, serta bagaimana model dapat diterapkan secara optimal ke dalam prototipe sistem prediksi berbasis web untuk mendukung strategi retensi pelanggan?

1.3 Batasan Masalah

Agar penelitian ini lebih terfokus dan terarah, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Penelitian ini menggunakan dataset IBM Telco Customer Churn sebagai sumber sekunder tanpa proses pengambilan data langsung.
2. Variabel yang dianalisis dalam penelitian ini terbatas pada variabel yang tersedia dalam dataset, yaitu lama berlangganan (tenure), total biaya layanan (Total Charges), biaya bulanan (Monthly Charges), jenis kontrak pelanggan (Contract), layanan internet berbasis fiber optik (InternetService_Fiber optic), penggunaan tagihan elektronik (Paperless Billing), layanan keamanan online (OnlineSecurity), serta metode pembayaran melalui cek elektronik (PaymentMethod_Electronic check).
3. Proses pra proses data mencakup membersihkan data, menangani nilai kosong, dan mengubah data kategorikal dengan teknik encoding tanpa menambah fitur eksternal.

4. Model machine learning tidak membandingkan dengan metode deep learning hanya menggunakan Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting, dan XGBoost.
5. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Hanya model Gradient Boosting yang digunakan untuk penyesuaian threshold.
6. Untuk memahami model, nilai fitur pada Random Forest dan nilai permutation pada Gradient Boosting digunakan tanpa menggunakan metode interpretabilitas seperti SHAP atau LIME.
7. Aspek temporal dan data real-time perilaku pelanggan tidak dianalisis dalam penelitian ini, sehingga pola perubahan perilaku pelanggan dari waktu ke waktu tidak menjadi fokus penelitian.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengevaluasi dan membandingkan cara kerja beberapa algoritma machine learning, seperti Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting, dan XGBoost, dalam memprediksi pelanggan yang berhenti menggunakan layanan berdasarkan dataset IBM Telco Customer Churn.
2. Mengukur dan mengevaluasi kemampuan model dengan menggunakan metrik akurasi, precision, recall, dan F1-score, serta melakukan penyesuaian threshold pada model Gradient Boosting untuk memperoleh hasil prediksi yang lebih optimal.
3. Menganalisis dan mengevaluasi faktor-faktor utama yang mempengaruhi churn pelanggan berdasarkan metode Feature Importance pada Random Forest dan Permutation Importance pada Gradient Boosting agar model lebih mudah dipahami.

4. Menghasilkan model prediksi churn pelanggan yang tidak hanya memiliki performa yang baik, tetapi juga mampu memberikan interpretasi terhadap faktor-faktor dominan yang mempengaruhi keputusan pelanggan untuk berhenti berlangganan.
5. Mengintegrasikan model optimal ke dalam sistem prediksi churn pelanggan berbasis web menggunakan Streamlit untuk mendukung pengambilan keputusan terkait strategi retensi pelanggan.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Penelitian ini berpotensi untuk meningkatkan penelitian machine learning terutama prediksi kehilangan pelanggan berdasarkan data tabel.
2. Penelitian ini dapat memberikan landasan ilmiah untuk mengklasifikasikan pelanggan yang berisiko berhenti berlangganan dengan menggunakan algoritma seperti Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting, dan XGBoost.
3. Penelitian ini menjelaskan faktor-faktor yang mempengaruhi tingkat churn pelanggan dengan menggunakan teknik interpretabilitas model seperti Permutation Importance dan Feature Importance.
4. Penelitian ini dapat membantu perusahaan telekomunikasi menemukan pelanggan yang berisiko churn untuk membuat strategi retensi yang lebih baik.
5. Penelitian ini menguraikan variabel-variabel utama masa berlangganan, biaya layanan, jenis kontrak, layanan internet, dan opsi pembayaran yang mempengaruhi keputusan konsumen untuk menghentikan langganan mereka.
6. Penelitian ini dapat mendukung pengambilan keputusan berbasis data untuk menurunkan tingkat churn dan meningkatkan loyalitas pelanggan.

7. Penelitian ini dapat membantu para peneliti di masa mendatang dalam mengembangkan model prediksi churn dengan meningkatkan kinerja model dan menerapkan teknik interpretabilitas yang lebih kompleks.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran secara umum mengenai isi dari setiap bab yang dibahas dalam penelitian ini.

Adapun sistematika penulisan dalam skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah yang menjelaskan pentingnya prediksi churn pelanggan dalam industri telekomunikasi, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi secara keseluruhan.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan dasar teori dan referensi yang mendukung penelitian, termasuk konsep churn pelanggan, machine learning, jenis algoritma klasifikasi yang digunakan seperti Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting, dan XGBoost, evaluasi model klasifikasi, teknik pra proses data, serta metode interpretabilitas model seperti Feature Importance dan Permutation Importance. Selain itu, dalam bab ini juga ditampilkan beberapa penelitian sebelumnya yang berkaitan, sebagai acuan untuk dibandingkan.

BAB III METODE PENELITIAN

Bab ini menjelaskan langkah-langkah penelitian yang dilakukan, mulai dari penjelasan dataset IBM Telco Customer Churn, proses pengumpulan dan pra proses data, teknik encoding variabel kategorikal, pembagian data latih dan data uji, pembangunan model machine learning, proses tuning dan penyesuaian threshold, hingga metode evaluasi dan analisis interpretabilitas model.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil penerapan model machine learning, hasil penilaian kinerja setiap algoritma, perbandingan performa model, serta analisis faktor-faktor utama yang mempengaruhi churn pelanggan berdasarkan Feature Importance dan

Permutation Importance. Dalam bab ini juga dibahas hasil yang didapatkan serta hubungannya dengan permasalahan dan tujuan penelitian.

BAB V PENUTUP

Bab ini menjelaskan kesimpulan yang didapat dari hasil penelitian dan juga memberikan saran-saran untuk pengembangan penelitian berikutnya serta penerapan lebih lanjut dalam lingkungan industri.proses

