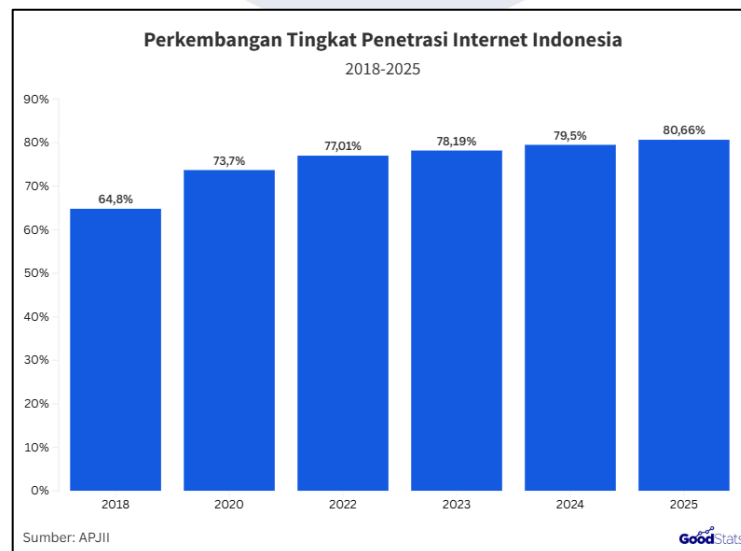


BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, industri telekomunikasi di Indonesia terus mengalami perkembangan seiring dengan meningkatnya penetrasi internet di berbagai wilayah. Berdasarkan laporan APJII, tingkat penetrasi internet nasional mencapai 80,66% pada tahun 2025 [1]. Peningkatan tersebut mencerminkan semakin tingginya kebutuhan masyarakat terhadap layanan internet dan komunikasi digital sehingga turut mendorong pertumbuhan industri telekomunikasi [2]. Seiring dengan berkembangnya industri tersebut, persaingan antarpenyedia layanan telekomunikasi juga menjadi semakin kompetitif. Kondisi ini mendorong perusahaan untuk tidak hanya meningkatkan kualitas layanan, tetapi juga mempertahankan loyalitas pelanggan agar tidak beralih ke penyedia layanan lain. Perkembangan penetrasi internet di Indonesia selama periode 2018–2025 ditunjukkan pada Gambar 1.1.



Gambar 1. 1 Persentase Pengguna Internet di Indonesia Tahun 2018–2025 [1]

Persaingan yang semakin ketat memberikan lebih banyak alternatif bagi pelanggan dalam memilih layanan telekomunikasi yang sesuai dengan kebutuhan dan preferensinya. Kondisi tersebut meningkatkan kemungkinan pelanggan berpindah ke penyedia layanan lain yang menawarkan harga maupun kualitas

layanan yang lebih sesuai. Fenomena perpindahan pelanggan tersebut dikenal sebagai *customer churn*. Pada salah satu perusahaan *fixed broadband* terbesar di Indonesia, tingkat *churn* bulanan tercatat mencapai 8,2%, yang menunjukkan bahwa risiko kehilangan pelanggan masih cukup tinggi [3]. Tingginya *customer churn* tidak hanya berdampak pada menurunnya loyalitas pelanggan, tetapi juga meningkatkan biaya yang harus dikeluarkan perusahaan untuk memperoleh pelanggan baru. Berbagai penelitian menunjukkan bahwa biaya akuisisi pelanggan baru dapat mencapai lima hingga sepuluh kali lebih tinggi dibandingkan biaya retensi pelanggan [4]. Kondisi tersebut mendorong perusahaan untuk lebih memprioritaskan strategi retensi pelanggan. Namun, strategi tersebut hanya dapat dilakukan secara efektif apabila perusahaan mampu mengidentifikasi pelanggan yang berpotensi melakukan *customer churn* sejak dini. Oleh karena itu, diperlukan suatu pendekatan yang mampu memprediksi kemungkinan pelanggan akan berhenti menggunakan layanan sehingga perusahaan dapat mengambil langkah retensi secara lebih tepat sasaran.

Keputusan pelanggan untuk berhenti menggunakan layanan dipengaruhi oleh berbagai faktor yang saling berkaitan, seperti karakteristik pelanggan, pola penggunaan layanan, kualitas layanan, serta kondisi persaingan pada industri telekomunikasi [5], [6]. Selain itu, faktor seperti *tenure*, kepatuhan pembayaran, serta interaksi pelanggan dengan layanan dukungan teknis juga diketahui memengaruhi kemungkinan terjadinya *customer churn* [7]. Banyaknya faktor yang saling berinteraksi menyebabkan proses identifikasi pelanggan yang berpotensi melakukan *customer churn* menjadi semakin kompleks apabila hanya mengandalkan analisis secara manual maupun pendekatan statistik konvensional [8].

Selain dipengaruhi oleh banyaknya variabel yang saling berhubungan, perusahaan telekomunikasi juga harus mengelola data pelanggan dalam jumlah besar yang mencakup informasi demografis, riwayat penggunaan layanan, hingga interaksi pelanggan terhadap layanan yang digunakan [9]. Karakteristik data tersebut menunjukkan bahwa prediksi *customer churn* merupakan permasalahan klasifikasi, yaitu menentukan apakah seorang pelanggan termasuk ke dalam

kategori *churn* atau *non-churn* berdasarkan pola historis yang dimiliki. Namun, metode analisis statistik konvensional memiliki keterbatasan dalam mengolah data dengan hubungan antarvariabel yang kompleks dan tidak linier sehingga kemampuan prediksi yang dihasilkan menjadi kurang optimal [10]. Oleh karena itu, diperlukan pendekatan yang mampu mengenali pola hubungan antarvariabel secara lebih efektif, salah satunya melalui *machine learning*. Dengan memanfaatkan *machine learning*, perusahaan dapat membangun model prediksi untuk mengidentifikasi pelanggan yang berpotensi melakukan *customer churn* sejak dini sehingga strategi retensi dapat diterapkan secara lebih tepat sasaran [11].

Prediksi *customer churn* telah banyak dikembangkan melalui penerapan berbagai algoritma *machine learning*. Salah satu studi yang menjadi acuan dalam penelitian ini adalah penelitian yang dilakukan oleh Omari et al. [12] karena membahas prediksi *customer churn* pada industri telekomunikasi menggunakan IBM Telco *Customer churn Dataset*. Hasil penelitian tersebut menunjukkan bahwa penerapan algoritma *machine learning* mampu memberikan performa yang baik dalam memprediksi *customer churn*. *Dataset* tersebut memuat data historis pelanggan telekomunikasi yang mencakup informasi demografis, layanan yang digunakan, kontrak, metode pembayaran, serta lama berlangganan (*tenure*), yaitu faktor-faktor yang berkaitan dengan kemungkinan terjadinya *customer churn*. Karakteristik tersebut menjadikan IBM Telco *Customer churn Dataset* representatif untuk memodelkan permasalahan *customer churn*. Selain itu, penggunaan *dataset* yang sama memungkinkan hasil penelitian ini dibandingkan secara lebih objektif dengan penelitian pembandingan [12]. Berdasarkan studi tersebut, penelitian ini menggunakan *Logistic Regression*, *K-Nearest Neighbors (KNN)*, *Naïve Bayes*, *Random Forest*, dan *XGBoost* sebagai algoritma yang akan dibandingkan karena mewakili pendekatan klasifikasi yang berbeda. Perbandingan tersebut dilakukan untuk memperoleh gambaran yang lebih komprehensif mengenai performa masing-masing algoritma dalam memprediksi *customer churn*.

Meskipun berbagai penelitian menunjukkan bahwa *machine learning* mampu memberikan performa yang baik dalam memprediksi *customer churn*, hasil kajian penelitian terdahulu menunjukkan bahwa pengembangan model umumnya masih

berfokus pada aspek tertentu. Sebagian penelitian lebih menitikberatkan pada perbandingan algoritma, sedangkan penelitian lainnya berfokus pada penanganan *class imbalance*, optimasi hyperparameter, peningkatan interpretabilitas model menggunakan *Explainable Artificial Intelligence (XAI)*, maupun implementasi model. Kondisi tersebut menunjukkan bahwa masih terdapat peluang untuk mengintegrasikan berbagai tahapan pengembangan model ke dalam satu alur penelitian yang sistematis sehingga proses pembangunan model dapat dilakukan secara lebih komprehensif. Berdasarkan kondisi tersebut, penelitian ini mengembangkan model prediksi *customer churn* menggunakan *framework* CRISP-DM dengan mengintegrasikan proses perbandingan performa beberapa algoritma *machine learning*, penanganan ketidakseimbangan kelas menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)*, optimasi hyperparameter, evaluasi kemampuan konsistensi model, interpretasi hasil prediksi menggunakan SHAP (Shapley Additive Explanations), serta implementasi model ke dalam aplikasi berbasis *Streamlit*. Dengan pendekatan tersebut, penelitian ini tidak hanya berfokus pada peningkatan performa model, tetapi juga mengembangkan proses pembangunan model yang lebih komprehensif dibandingkan penelitian terdahulu.

Selain meningkatkan performa model, penelitian ini juga memberikan perhatian terhadap kemampuan model dalam mengidentifikasi pelanggan yang benar-benar berpotensi melakukan *customer churn*. Oleh karena itu, evaluasi model tidak hanya mempertimbangkan nilai *accuracy*, tetapi juga memberikan perhatian terhadap nilai *recall* sebagai salah satu indikator penting dalam menilai kemampuan model mendeteksi pelanggan yang berpotensi melakukan *customer churn*, tanpa mengabaikan metrik evaluasi lainnya. Melalui pendekatan tersebut, penelitian ini diharapkan menghasilkan model prediksi *customer churn* yang memiliki performa klasifikasi yang baik, kemampuan konsistensi yang optimal, mudah diinterpretasikan, serta dapat mendukung pengambilan keputusan pada industri telekomunikasi.

1.2 Rumusan Masalah

Penelitian ini dirancang untuk menjawab beberapa permasalahan berikut:

1. Bagaimana hasil analisis dan optimasi algoritma *Logistic Regression*, *K-Nearest Neighbors (KNN)*, Naive Bayes, *Random Forest*, dan *XGBoost* dalam memprediksi *customer churn* pada industri telekomunikasi?
2. Bagaimana pengaruh optimasi *hyperparameter* menggunakan metode *Grid search* terhadap peningkatan performa *model* prediksi *churn*?
3. Algoritma manakah yang menghasilkan performa terbaik setelah dilakukan optimasi *hyperparameter* berdasarkan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC, serta bagaimana hasil interpretasi model menggunakan pendekatan *Explainable Artificial Intelligence (XAI)* dalam mengidentifikasi faktor-faktor yang memengaruhi *customer churn*?

1.3 Batasan Masalah

Batasan masalah adalah sebagai berikut:

1. Penelitian ini menggunakan data sekunder berupa IBM Telco *Customer churn Dataset* yang diperoleh dari platform Kaggle, sehingga analisis hanya terbatas pada variabel dan karakteristik yang tersedia dalam *dataset* tersebut.
2. *Dataset* yang digunakan merupakan *dataset* telekomunikasi bersifat historis dan bukan data *real-time*, sehingga hasil prediksi hanya merepresentasikan pola *churn* berdasarkan data masa lalu.
3. Pemilihan algoritma tersebut bertujuan untuk membandingkan performa beberapa pendekatan klasifikasi pada permasalahan prediksi *customer churn*. Oleh karena itu, penelitian ini tidak mencakup pendekatan *deep learning* seperti CNN maupun RNN.
4. Optimasi model dalam penelitian ini dibatasi pada optimasi *hyperparameter* menggunakan metode *Grid search* untuk meningkatkan performa model. Optimasi yang dilakukan hanya mencakup penyesuaian *hyperparameter* pada algoritma *Logistic Regression*, *K-Nearest Neighbors (KNN)*, Naive Bayes, *Random*

Forest, dan *XGBoost*, serta tidak mencakup pengembangan arsitektur model maupun metode optimasi lanjutan lainnya.

5. Evaluasi kinerja *model* terbatas pada metrik klasifikasi, yaitu *Accuracy*, *Precision*, *Recall*, *F1-score*, dan *ROC-AUC*.
6. Penelitian ini tidak membahas aspek ekonomi secara rinci, seperti perhitungan kerugian finansial akibat *churn*, dan hanya berfokus pada aspek prediksi berbasis data.
7. Penelitian ini menggunakan teknik penanganan data imbalance berupa *Synthetic Minority Over-sampling Technique (SMOTE)* untuk meningkatkan keseimbangan distribusi kelas pada data pelatihan.
8. Penelitian ini hanya berfokus pada penggunaan metode SHAP dalam pendekatan *Explainable Artificial Intelligence (XAI)* untuk membantu memahami hasil prediksi *model*, sehingga metode interpretabilitas lainnya tidak dibahas.
9. Hasil penelitian ini hanya berlaku pada IBM Telco *Customer churn Dataset* serta skenario pengujian yang digunakan dalam penelitian, termasuk pembagian data, proses preprocessing, penanganan data imbalance menggunakan SMOTE, dan konfigurasi evaluasi model. Oleh karena itu, performa model yang diperoleh tidak dapat digeneralisasikan secara langsung pada *dataset*, kondisi operasional, maupun lingkungan telekomunikasi yang berbeda tanpa dilakukan proses pelatihan dan evaluasi ulang.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan, yaitu :

1. Menganalisis dan mengoptimalkan performa algoritma *Logistic Regression*, *K-Nearest Neighbors (KNN)*, Naive Bayes, *Random Forest*, dan *XGBoost* dalam memprediksi *customer churn* pada industri telekomunikasi dengan menitikberatkan pada peningkatan kemampuan model dalam mengidentifikasi pelanggan yang berpotensi *churn*.

2. Menganalisis pengaruh optimasi hyperparameter menggunakan metode *Grid search* terhadap peningkatan performa model prediksi *customer churn*.
3. Mengevaluasi dan membandingkan performa *model* sebelum dan sesudah optimasi berdasarkan metrik *Accuracy*, *Precision*, *Recall*, *F1-score*, dan *ROC-AUC* untuk menentukan *model* terbaik, serta menganalisis dan menginterpretasikan hasil prediksi *model* menggunakan pendekatan *Explainable Artificial Intelligence (XAI)* guna mengidentifikasi faktor-faktor yang mempengaruhi *customer churn*.

1.4.2 Manfaat Penelitian

Manfaat Penelitian ini sebagai berikut:

1. Memberikan hasil analisis dan optimasi performa model *machine learning* dalam memprediksi *customer churn* pada industri telekomunikasi melalui penerapan optimasi hyperparameter.
2. Memberikan analisis perbandingan performa model sebelum dan sesudah optimasi berdasarkan metrik evaluasi yang relevan sehingga dapat diketahui model yang paling efektif dalam mendeteksi pelanggan *churn*.
3. Menjadi bahan pertimbangan bagi perusahaan telekomunikasi dalam menyusun strategi retensi pelanggan berbasis data melalui pemanfaatan hasil prediksi *customer churn* serta interpretasi faktor-faktor yang memengaruhi *customer churn*.

1.5 Sistematika Penulisan

Penyusunan laporan skripsi ini terdiri atas lima bab yang disusun secara berurutan untuk menjelaskan alur penelitian secara sistematis.

1. BAB I PENDAHULUAN

Bagian ini memuat latar belakang, rumusan masalah, tujuan penelitian, batasan penelitian, serta gambaran umum sistematika penulisan skripsi.

2. BAB II TINJAUAN PUSTAKA

Pada bab ini dijelaskan teori-teori pendukung yang berkaitan dengan *customer churn*, *machine learning*, dan penelitian terdahulu yang relevan dengan topik penelitian.

3. BAB III METODOLOGI PENELITIAN

Bab ini menguraikan metode penelitian yang digunakan, meliputi proses pengumpulan data, *preprocessing*, pembangunan *model*, hingga tahap evaluasi dan optimasi *model machine learning*.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen, evaluasi performa model sebelum dan sesudah optimasi, interpretasi hasil menggunakan *Explainable Artificial Intelligence (XAI)* berbasis SHAP, serta pembahasan terhadap hasil penelitian.

5. BAB V KESIMPULAN DAN SARAN

Bab terakhir berisi kesimpulan penelitian dan rekomendasi yang dapat dijadikan acuan untuk pengembangan penelitian berikutnya.

