

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Pada perusahaan penyedia layanan mana pun, seorang pelanggan menjadi "*churner*" ketika mereka menghentikan langganannya mereka atau berpindah ke penyedia layanan pesaing. Dengan kata lain, *churning* adalah proses perputaran pelanggan. Hal ini adalah masalah besar bagi perusahaan dimana sebagian besar pelanggan dapat dengan mudah berpindah ke pesaing lain di pasar. Contohnya termasuk perbankan, perusahaan asuransi, perusahaan game, dan perusahaan telekomunikasi (Ullah dkk., 2019).

*Churn* ini merupakan masalah penting yang sering dikaitkan dengan siklus hidup industri. Ketika industri berada dalam fase pertumbuhan dalam siklus hidupnya, penjualan meningkat secara eksponensial dan jumlah pelanggan baru jauh lebih banyak daripada jumlah pelanggan yang berpindah. Di sisi lain, perusahaan yang berada dalam fase matang dalam siklus hidupnya, menetapkan fokus mereka untuk mengidentifikasi tingkat *churn* pelanggan salah satunya memprediksi *churn* pelanggan tersebut (Hammoudeh dkk., 2019).

Alasan utama yang menyebabkan *churn* dibagi menjadi dua kelompok: tidak disengaja dan disengaja. *Churn* yang tidak disengaja terjadi ketika keadaan berubah sehingga mencegah pelanggan untuk menggunakan layanan di masa depan. Contoh yang tidak disengaja adalah keadaan ekonomi yang membuat layanan menjadi terlalu mahal bagi pelanggan. *Churn* yang disengaja terjadi ketika pelanggan memilih untuk beralih ke perusahaan lain yang menyediakan layanan

serupa. Contoh yang disengaja adalah penawaran yang lebih baik dari pesaing, layanan yang lebih canggih, dan harga yang lebih baik untuk layanan yang sama. Jenis *churn* ini adalah salah satu yang berusaha dicegah oleh sebagian besar perusahaan(Labhsetwar, 2020).

Membangun sistem untuk mengelola *churn* sangatlah penting. Ada dua pendekatan dasar untuk mengelola perputaran pelanggan: terarah dan tidak terarah. Dalam pendekatan tidak terarah, perusahaan mengandalkan produk yang unggul dan iklan massal untuk meningkatkan loyalitas terhadap merek dan mempertahankan pelanggan. Dalam pendekatan terarah, perusahaan mengandalkan identifikasi pelanggan yang cenderung berpindah, dan kemudian menyesuaikan kebutuhan mereka untuk mencegah terjadinya *churn* pelanggan(Lalwani dkk., 2022).

Dalam beberapa tahun terakhir, prediksi *churn* menjadi isu yang sangat penting dalam industri telekomunikasi(Zhu, 2023). Untuk mengatasi masalah ini, operator telekomunikasi harus mengenali pelanggan sebelum mereka melakukan *churn*. Oleh karena itu, mengembangkan pengklasifikasi unik yang dapat memprediksi *churn* di masa depan sangatlah penting(Jain dkk., 2021). Pengklasifikasi ini harus dapat mengenali pengguna yang memiliki kecenderungan untuk melakukan churn dalam waktu dekat, sehingga operator dapat bereaksi dengan cepat dengan memberikan diskon dan promosi yang sesuai(Adhikary & Gupta, 2021). Teknik yang paling sering digunakan untuk tujuan ini adalah algoritma pembelajaran untuk prediksi, seperti *Decision Tree*, *Logistics Regression*, *K-Nearest Neighbors*, *Naive Bayes*, *Neural Networks*, dll.(Han dkk., 2022).

Penelitian harus fokus pada identifikasi fitur baru yang paling efektif dalam memprediksi *churn* pelanggan(Singh dkk., 2021). Selain itu, algoritma atau metode yang memiliki performa yang baik dengan tingkat akurasi yang tinggi sangat diperlukan oleh industri telekomunikasi dalam prediksi *churn*(Zhu, 2023).

Pada penelitian-penelitian sebelumnya telah dilakukan penelitian terkait prediksi *churn* pelanggan telekomunikasi dengan algoritma atau metode yang beragam, algoritma atau metode yang sudah digunakan pada penelitian serupa yakni algoritma *Naive Bayes* dan *K-Nearest Neighbor* pada penelitian (Kaharudin dkk., 2019), algoritma *Adaptive Boost*, *Extreme Gradient Boost* dan *Random Forest* pada penelitian(Latief dkk., 2021), algoritma *Adaptive Boost*, *Extreme Gradient Boost* dan *Categorical Boost* pada penelitian (Lalwani dkk., 2022) dan algoritma *Classification And Regression Tree(CART)* dan *Naive Bayes* pada penelitian (Adhitya dkk., 2023).

Data yang tidak seimbang atau lebih sering disebut *Imbalanced Data*, adalah suatu kondisi dimana data memiliki rasio yang tidak seimbang antara satu kelas dengan kelas yang lain, sehingga terdapat kelas mayoritas dan kelas minoritas(Cosmas Haryawan & Yosef Muria Kusuma Ardhana, 2023). Sulit untuk melakukan prediksi pada dataset yang tidak seimbang karena pengklasifikasi akan cenderung mendeteksi kelas mayoritas daripada kelas minoritas. Oleh karena itu, keluaran dari klasifikasi akan menjadi bias(Y. R. Chen dkk., 2021). Beberapa metode dapat digunakan untuk mengatasi hal tersebut. Metode *resampling* menjadi salah satu yang paling efektif dalam menyelesaikan permasalahan *imbalanced data* ini(Ghorbani & Ghousi, 2020). Dalam metode *resampling* terdapat teknik

*oversampling* dan teknik *undersampling*, Teknik *oversampling* memberikan hasil yang lebih baik dibandingkan teknik *undersampling* (Magnolia dkk., 2023). Salah satu metode *oversampling* yang banyak digunakan adalah *Synthetic Minority Oversampling Technique (SMOTE)* (Ghorbani & Ghousi, 2020). Pada penelitian (Sutoyo & Fadlurrahman, 2020), metode *SMOTE* terbukti dapat mengatasi masalah ketidakseimbangan data dan mendapatkan hasil klasifikasi yang lebih baik.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk menerapkan tiga algoritma *Boosting* yaitu algoritma *Gradient Boost*, *Extreme Gradient Boost (XGBoost)*, dan *Adaptive Boost (AdaBoost)* dalam memprediksi *churn* pelanggan telekomunikasi. Penerapan ketiga algoritma tersebut akan dilakukan dalam 2 skenario yaitu, skenario pertama terhadap dataset asli yang tidak seimbang atau *imbalanced* dan skenario kedua akan diimplementasikan teknik *SMOTE* untuk mengatasi ketidakseimbangan data tersebut sehingga dataset menjadi seimbang atau *balanced*. Setiap skenario akan dilakukan proses *tuning hyperparameter*.

Penelitian ini akan berfokus pada perbandingan performa dan akurasi dari ketiga algoritma tersebut dalam memprediksi *churn* pelanggan telekomunikasi. Penelitian ini diharapkan dapat memberikan informasi yang bermanfaat terkait penggunaan dan performa algoritma atau metode yang berbeda-beda dalam memprediksi *churn* pelanggan telekomunikasi.

## 1.2. Rumusan Masalah

Berdasarkan uraian latar belakang, maka permasalahan yang akan dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana menerapkan algoritma *Gradient Boost*, *Extreme Gradient Boost*, dan *Adaptive Boost* untuk memprediksi *churn* pelanggan telekomunikasi?
2. Algoritma apa yang memiliki performa dengan tingkat akurasi yang lebih baik untuk memprediksi *churn* pelanggan telekomunikasi?

## 1.3. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma *Gradient Boost*, *Extreme Gradient Boost*, dan *Adaptive Boost* untuk memprediksi *churn* pelanggan telekomunikasi.
2. Membandingkan performa dan akurasi antara algoritma *Gradient Boost*, *Extreme Gradient Boost*, dan *Adaptive Boost* untuk prediksi *churn* pelanggan telekomunikasi.

## 1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Manfaat bagi Universitas Siliwangi, sebagai bahan referensi sarana pengembangan wawasan keilmuan khususnya jurusan Informatika.
2. Manfaat bagi Pembaca, sebagai bahan referensi terkait algoritma *Gradient Boost*, *Extreme Gradient Boost*, dan *Adaptive Boost* untuk prediksi *churn* pelanggan telekomunikasi.
3. Manfaat dalam Bidang Ilmu Pengetahuan, sebagai kontribusi pada pengetahuan dan pemahaman penerapan algoritma *Gradient Boost*, *Extreme*

*Gradient Boost*, dan *Adaptive Boost* untuk prediksi sebuah kasus terfokus pada performa dan akurasi, sehingga dapat dijadikan sebagai rujukan, sumber informasi dan bahan referensi penelitian selanjutnya.

### 1.5. Batasan Masalah

Batasan masalah yang dibahas dalam penelitian ini adalah sebagai berikut:

1. Algoritma yang digunakan yaitu algoritma *Gradient Boost*, *Extreme Gradient Boost*, dan *Adaptive Boost*.
2. *Hyperparameter* algoritma yang digunakan yaitu *n-estimator* dan *learning rate*.
3. Nilai *n-estimator* yang digunakan yaitu 50, 100, dan 200.
4. Nilai *learning rate* yang digunakan yaitu 0.1, 0.01, dan 0.001.
5. Metrik performa yang digunakan yaitu *accuracy*, *precision*, *recall*, *F1-Score*, dan *Log-Loss*.
6. Teknik *oversampling* yang digunakan yaitu *Synthetic Minority Oversampling Technique (SMOTE)*.
7. Hasil penerapan algoritma berfokus pada perbandingan performa dan akurasi pada setiap algoritma.
8. Kasus yang digunakan yaitu *churn* pelanggan telekomunikasi.
9. Data yang digunakan berupa dataset suatu perusahaan telekomunikasi.