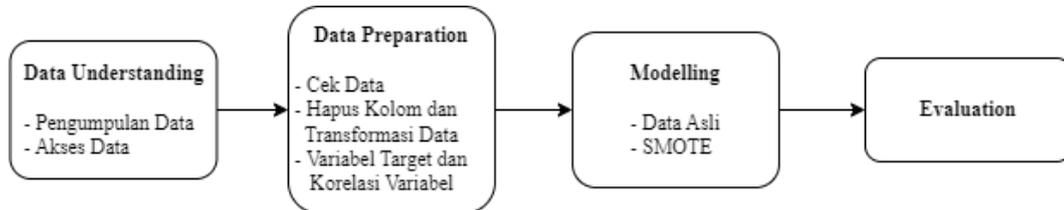


## BAB III

### METODE PENELITIAN

Alur metode penelitian yang dilakukan dari awal hingga akhir terdiri dari empat tahapan, tertera pada Gambar 3.1 berikut:



Gambar 3.1 Metode Penelitian

#### 3.1. *Data Understanding*

*Data Understanding* atau Pemahaman Data, pada tahapan ini dimulai dengan mengumpulkan data yang akan digunakan, data tersebut merupakan dataset suatu perusahaan telekomunikasi dengan format *CSV (Comma Separated Values)* yang diunduh dari situs *Kaggle*, kemudian melakukan pengaksesan data dengan *jupyter notebook* pada *google colab* dimulai dari meng-*import library* yang akan digunakan, lalu memasukan atau membuka dataset, menghitung jumlah baris dan kolom, dan mengecek tipe data dari setiap kolom atau variabel.

#### 3.2. *Data Preparation*

*Data Preparation* atau Persiapan Data, pada tahap ini dilakukan beberapa hal diantaranya:

1. Pengecekan data apakah terdapat *missing value* dan ataupun duplikasi data.

2. Penghapusan kolom atau variabel yang tidak akan digunakan pada penelitian ini dan men-transformasi atau mengubah data dengan *Data Dummy* pada dataset.
3. Analisis kolom atau variabel target dan korelasi setiap kolom atau variabel dengan kolom atau variabel target.

### 3.3. *Modelling*

*Modelling* atau Pemodelan dilakukan dengan memilih dan menerapkan teknik pemodelan. Algoritma *Gradient Boost*, *Extreme Gradient Boost*, dan *Adaptive Boost* pada pemodelan penelitian ini akan dilakukan perbandingan performa dan akurasi dalam memprediksi *churn* pelanggan telekomunikasi. Perbandingan algoritma dilakukan dengan 2 skenario yaitu skenario pertama dengan dataset asli (*imbalanced*) dan skenario kedua dengan mengimplementasikan *SMOTE* sehingga jumlah dataset bertambah dan menjadi *balanced*.

Pada setiap skenario menggunakan perbandingan *split* data 70:30 untuk data latih dan data uji nya, rasio ini banyak digunakan dalam literatur ini (Kaharudin dkk., 2019; Pham & Ho, 2021; Iqbal dkk., 2022; Adhitya dkk., 2023; Mauludin Nur Aziz dkk., 2023). Kemudian dilakukan uji coba beberapa *hyperparameter* untuk ketiga algoritma yaitu *n-estimator* dengan nilai 50, 100 dan 200 dan juga *learning rate* dengan nilai 0.1, 0.01 adalah standar nilai *Stochastic Gradient Descent (SGD)* (Ketkar, 2017), dan 0.001 adalah standar nilai *Adaptive Moment Estimation (Adam)* (Kingma & Ba, 2014).

Semakin banyak jumlah *n-estimator*, semakin baik pula hasilnya. Namun, pada suatu titik tertentu kinerja algoritma dalam prediksi akan menurun dikarenakan tingginya kebutuhan komputasi(KURNIAWATI dkk., 2022).

Jika nilai *learning rate* terlalu kecil maka waktu pembelajaran yang dibutuhkan akan semakin lama sebaliknya jika terlalu besar maka pembelajaran akan menjadi kurang optimal karena proses terlalu cepat dan pembelajaran menjadi tidak stabil(Rochmawati dkk., 2021).

#### **3.4. Evaluation**

Pada tahap ini akan dilakukan evaluasi menyeluruh terhadap seluruh model yang dihasilkan dari kedua skenario yang digunakan pada penelitian ini. Hasil evaluasi akan dianalisis untuk menentukan algoritma mana yang memiliki performa terbaik dalam memprediksi *churn* pelanggan telekomunikasi.