

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Dengan kemajuan teknologi informasi banyak perusahaan yang bergerak di bidang industri telekomunikasi. Industri telekomunikasi merupakan salah satu industri dengan pengguna terbanyak dan menempati urutan teratas dalam daftar industri dengan pertumbuhan tercepat, menempati 90% dari populasi (Hashmi et al., 2013). Hal ini membuat perusahaan telekomunikasi terus menjalankan bermacam - macam inovasi untuk menunjang persaingan usaha yang sungguh ketat dan semakin sulit mendapatkan pelanggan. *Churn* pelanggan yang tinggi merupakan salah satu tingkat kegagalan perusahaan, oleh karena itu *churn* harus dikurangi. Perusahaan lebih memilih mempertahankan pelanggan untuk menghindari resiko *churn* karena biayanya lebih murah daripada menambah pelanggan baru (Arina & Ulfah, 2022). Pelanggan yang menghentikan layanan dan pelanggan beralih ke layanan perusahaan telekomunikasi lainnya. Perilaku pelanggan ini dikenal sebagai *churn*. *Churn* merupakan penghentian layanan telekomunikasi oleh pelanggan atau perusahaan (Saputra, 2021). Dalam mempertahankan pelanggan, perusahaan telekomunikasi memerlukan cara untuk memprediksi *churn* pelanggan dengan klasifikasi pelanggan. Klasifikasi pelanggan dapat dilakukan dengan menggunakan *machine learning*.

Sejalan dengan perkembangan teknologi kehadiran *machine learning* di bidang komputer telah menarik banyak perhatian. *Machine Learning* menjadi sebuah pengembangan algoritma dan model data itu sendiri. *Machine learning* memiliki peran penting dalam pembangunan, terutama di pembangunan analisis data (Mutmainnah et al., 2018). Salah satu algoritma yang ada dalam *machine learning* yaitu klasifikasi.

Beberapa algoritma klasifikasi seperti *K-Nearest Neighbors* dan *Random Forest* telah diterapkan oleh penelitian terdahulu yang dilakukan oleh (Husein & Harahap, 2021) dengan judul “Pendekatan *Data Science* untuk Menemukan *Churn* Pelanggan pada Sector Perbankan dengan *Machine Learning*”. Penelitian ini menggunakan *Random Forest* dasar dan menghasilkan akurasi lebih baik sebesar 86% dan algoritma *Random Forest* termasuk dalam algoritma populer, efektif dan memberikan prediksi yang relatif baik. Pada selanjutnya yang dilakukan oleh

(Pamungkas et al., 2020) dengan judul “Perbandingan Metode Klasifikasi *Supervised Learning* pada Data Bank *Customers* Menggunakan *Python*”. Penelitian ini menggunakan klasifikasi *Supervised Learning* yaitu *Regresi Logistik*, *K-Nearest Neighbor*, *Super Vector Machine*, *Naïve bayes*, *Decision Tree*, dan *Random Forest*. Dari beberapa metode klasifikasi *Supervised* ini *Random Forest* menghasilkan nilai terbaik 86,2%.

Pada penelitian berikutnya dengan optimasi *boosting* diteliti oleh (Iwendi et al., 2020) yang berjudul “COVID-19 *Patient Health Prediction Using Boosted Random Forest Algorithm*” algoritma yang digunakan menggunakan *Random Forest* dan disempurnakan oleh teknik *boosting* yaitu *AdaBoost* menghasilkan akurasi 94% dan *f1-score* 86% pada dataset yang digunakan. Peneliti selanjutnya dilakukan oleh (Bhatele & Bhadauria, 2020) dengan judul “*Glioma Segmentation and Classification System Based on Proposed Textures Features Extraction Method and Hybrid Ensemble Learning*”. Metode yang digunakan SVM, *Naïve Bayes*, RF, KNN, ADBRF dan XGBRF. Hasil dari metode kombinasi XGB dengan RF (XGBRF) mendapatkan akurasi yang memuaskan sebesar 99,25% pada paduan urutan MRI T1C+T2+Flair dan akurasi 96.75% pada perpaduan urutan MRI T1+T1C+T2+Flair.

Pada penelitian ini akan melakukan klasifikasi *churn* pelanggan menggunakan algoritma *Random Forest*. Algoritma *Random Forest* dipilih karena memiliki kemampuan dalam mengklasifikasi data yang tidak lengkap dan serta dapat menangani data sampel yang banyak. *Random Forest* merupakan algoritma yang terdiri dari banyak *tree* dan memprediksi dengan cara *voting class* dari masing - masing *tree*, *class* dengan jumlah *vote* terbanyak akan menjadi final *class* (Rachmi, 2020).

Teknik optimasi yang akan digunakan dalam penelitian ini adalah teknik *Boosting* menggunakan *XGBoost* dan *AdaBoost*. *XGBoost* dipilih karena memiliki kemampuan beradaptasi di berbagai situasi, fitur yang berguna mempercepat sistem perhitungan dan mencegah *overfitting* (Herni Yulianti et al., 2022). *Adaboost* digunakan karena sangat mudah diterapkan, tidak perlu mengatur parameter dan fleksibel sehingga dapat digabungkan dengan berbagai algoritma (Prasetio & Susanti, 2019).

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana perbandingan hasil optimasi algoritma *Random Forest* menggunakan teknik *Boosting XGBoost* dan *AdaBoost*?
2. Bagaimana menghitung akurasi, presisi, *recall* dan *f1-score*?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui hasil perbandingan optimasi algoritma *Random Forest* menggunakan teknik *Boosting XGBoost* dan *AdaBoost*.
2. Mengetahui hasil akurasi, presisi, *recall* dan *f1-score*.

1.4 Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini sebagai berikut:

1. Penelitian ini dapat digunakan sebagai referensi untuk menyeimbangkan penelitian lebih lanjut yang berkaitan dengan klasifikasi *churn* pelanggan dengan algoritma *Random Forest* dan optimasi *Random Forest* menggunakan teknik *Boosting (XGBoost dan AdaBoost)*.
2. Diharapkan dari penelitian ini dapat memberikan kontribusi secara keilmuan berupa hasil perbandingan algoritma *Random Forest* dan optimasi *Random Forest* menggunakan teknik *Boosting (XGBoost dan AdaBoost)*.
3. Hasil dari penelitian ini dapat memotivasi seorang ilmuwan data dalam mengoptimalkan algoritma *Random Forest* dan algoritma lainnya.
4. Bagi perusahaan penelitian ini dapat membantu menangani terjadinya *churn* pelanggan.

1.5 Batasan Masalah

Batasan masalah yang dapat diambil dari penelitian ini sebagai berikut:

1. Analisis yang dilakukan perbandingan optimasi algoritma *Random Forest* dan *Random Forest* menggunakan teknik *Boosting* (*XGBoost* dan *AdaBoost*) dalam klasifikasi *churn* pelanggan.
2. Tools yang digunakan dalam menganalisis yaitu *jupyter notebook*
3. Analisis data dilakukan menggunakan algoritma *Random Forest* dengan Python 3.6.4.
4. Data yang digunakan dalam penelitian ini bersumber dari *kaggle* <https://www.kaggle.com/datasets/barun2104/telecom-churn?datasetId=567482>.

