

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pelanggan merupakan aspek penting bagi perusahaan penyedia produk maupun jasa karena tanpa mereka, perusahaan tidak akan bisa bertahan dan mendapatkan keuntungan. Bagi sebuah perusahaan mendapatkan pelanggan hanyalah sebuah langkah awal karena setelah itu mereka harus bisa mempertahankan atau mendapatkan loyalitas pelanggan agar tetap menggunakan produk atau jasa yang ditawarkannya. Memahami karakteristik pelanggan dan mengambil langkah-langkah yang tepat agar mereka tidak berhenti menggunakan produk atau jasa, apalagi beralih ke pesaing, sangat penting untuk menjaga stabilitas perusahaan (Panuju, 2019). Hal ini dikarenakan mempertahankan pelanggan yang sudah ada jauh lebih hemat biaya dibandingkan dengan mendapatkan pelanggan baru (Reichheld dan Sasser, 1990).

Kemampuan perusahaan untuk mempertahankan pelanggan dan aktif melakukan transaksi atau yang biasa disebut retensi pelanggan merupakan salah satu tantangan terbesar yang dihadapi perusahaan pada era ini. Fenomena ketika pelanggan memutuskan untuk berhenti berlangganan atau beralih ke produk atau penyedia jasa lain, yang dikenal dengan *customer churn*, memiliki dampak yang signifikan terhadap pendapatan dan stabilitas perusahaan (Kamil *et al.*, 2023). *Customer churn* yang tinggi dapat menjadi indikasi bahwa banyak pelanggan yang tidak tertarik dengan produk atau jasa yang ditawarkan oleh perusahaan tersebut dan bisa berdampak pada penurunan pendapatan (Aziz *et al.*, 2023).

Fenomena tersebut terjadi pada berbagai sektor industri, salah satunya industri kebugaran seperti *gym*. *Gym* atau klinik kebugaran adalah fasilitas yang menyediakan layanan olahraga untuk masyarakat, yang bertujuan untuk menghindari kecelakaan seperti salah gerak, *over training*, atau cedera akibat mengangkat beban terlalu berat (A. Jatmiko, 2015). Pada industri ini, *customer churn* terjadi ketika anggota suatu *gym* memutuskan untuk tidak memperpanjang keanggotaan (*membership*) mereka atau beralih ke tempat lain yang menawarkan fasilitas atau layanan yang lebih sesuai dengan kebutuhan mereka. Hal ini menjadi tantangan besar bagi pelaku usaha di industri kebugaran karena tingginya tingkat *churn* di sebuah *gym* bisa memengaruhi stabilitas bisnis (Sholeha *et al.*, 2024). Salah satu upaya pencegahan yang dapat dilakukan adalah dengan melakukan analisis dan membuat model dari data yang dimiliki oleh pemilik *gym* untuk memprediksi anggota yang berpeluang besar untuk *churn* sehingga mereka dapat mengambil langkah preventif (Aldosary dan Alrashdan, 2021).

Random Forest merupakan pengembangan dari salah satu algoritma *Decision Tree* yaitu *Clasification and Regression Trees* (CART) yang menerapkan konsep *bagging* (*bootstrap aggregating*) dan *random feature selection*. Metode ini bekerja dengan teknik *ensemble*, yaitu membangun beberapa model CART untuk membuat model yang masing-masing akan menghasilkan prediksi kemudian prediksi akhirnya diperoleh dari hasil terbanyak (*majority voting*). Metode ini memiliki keunggulan yaitu dapat meningkatkan akurasi ketika terdapat *missing value*, tahan terhadap *outlier*, *feature selection* yang mampu meningkatkan performa model,

serta mampu bekerja pada data yang besar dan variabel yang kompleks (Devella *et al.*, 2020).

Salah satu tantangan utama dalam membangun model *Random Forest* terutama dengan data yang besar dan variabel yang kompleks adalah menentukan parameter yang optimal untuk meningkatkan performa model. Metode *hyperparameter tuning* merupakan langkah tepat yang dapat diambil karena menentukan kombinasi parameter secara manual dapat menjadi tugas yang sulit dan membutuhkan waktu yang lama (Sitanggang dan Sitompul, 2024). Terdapat dua metode yang sering digunakan dalam *hyperparameter tuning* yaitu *RandomizedsearchCV* dan *GridsearchCV*. Kedua metode tersebut bertujuan untuk mencari kombinasi *hyperparameter* yang optimal tetapi memiliki pendekatan yang berbeda dalam mengeksplorasi ruang *hyperparameter*. *GridSearchCV* melakukan pencarian sistematis melalui ruang *hyperparameter* yang telah ditentukan, sedangkan *RandomizedSearchCV* melakukan pencarian acak yang sering kali lebih efisien dari segi waktu komputasi. Kedua metode ini memiliki kelebihan dan kekurangan masing-masing. *GridSearchCV* menyediakan pencarian yang komprehensif namun membutuhkan waktu komputasi yang lebih lama, sedangkan *RandomizedSearchCV* lebih cepat tetapi tidak menyeluruh.

Beberapa penelitian sebelumnya terkait *customer churn* pernah dilakukan oleh Damanik dan Jambak (2023), tentang klasifikasi *customer churn* pada telekomunikasi industri menggunakan algoritma C4.5 yang menghasilkan model dengan akurasi sebesar 79,53%, kemudian Suryana (2017) tentang prediksi *churn* pada pelanggan TV berbayar (studi kasus Transvision jawa barat) yang

menghasilkan model dengan akurasi 90,89%. Salah satu penelitian sebelumnya tentang *gym churn membership* menggunakan *Artificial Neural Network* (ANN) yang diperkuat dengan konsep psikologis pembentukan kebiasaan dalam industri kebugaran (Aldosary dan Alrashdan, 2021) menghasilkan model dengan akurasi 92,1%. Penelitian lain terkait *hyperparameter tuning* pada *Random Forest* untuk diagnosa Covid-19 oleh Baita *et al.* (2023), menghasilkan model dengan akurasi 96%, akurasi tersebut lebih tinggi dari model tanpa *hyperparameter tuning* yang memiliki akurasi 94%.

Penelitian ini membahas penerapan algoritma *Random Forest* dalam membangun model prediksi *gym churn membership*, dengan dukungan metode *hyperparameter tuning* menggunakan *RandomizedSearchCV* dan *GridSearchCV*. Hasil dari penelitian ini diharapkan dapat memberikan manfaat berupa model prediksi yang akurat dan dapat digunakan sebagai bahan pertimbangan oleh pelaku usaha dalam merancang strategi retensi pelanggan, seperti pengiriman notifikasi personal, program loyalitas, serta penawaran promo bagi anggota berisiko tinggi, sehingga dapat membantu menekan angka *churn* dan mempertahankan basis pelanggan yang ada (Kazim, 2025).

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diperoleh beberapa rumusan masalah sebagai berikut:

1. Bagaimana kinerja model *Random Forest* dengan *hyperparameter tuning* *RandomizedSearchCV*?

2. Bagaimana kinerja model *Random Forest* dengan *hyperparameter tuning GridSearchCV*?
3. Bagaimana perbandingan kinerja model *Random Forest* dengan *hyperparameter tuning RandomizedSearchCV* dan *GridSearchCV*?

1.3 Batasan Masalah

Berikut batasan masalah yang digunakan dalam penelitian ini:

1. Data *Gym customer feature and churn* yang diperoleh dari Kaggle.
2. Metode yang digunakan adalah *Random Forest* dengan *hyperparameter tuning RandomizedSearchCV* dan *GridSearchCV* dan evaluasi matriks yang digunakan adalah nilai *accuracy*, *precision*, *recall*, *F1-score*, dan AUC.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengidentifikasi kinerja model *Random Forest* dengan *hyperparameter tuning* dan *RandomizedSearchCV*.
2. Mengidentifikasi kinerja model *Random Forest* dengan *hyperparameter tuning GridSearchCV*.
3. Membandingkan kinerja model *Random Forest* dengan *hyperparameter tuning RandomizedSearchCV* dan *GridSearchCV*.