

BAB I

PENDAHULUAN

1.1 Latar Belakang

Apartemen adalah salah satu jenis tempat tinggal vertikal yang terdiri atas banyak unit dengan berbagai tipe yang berbeda, baik dari segi ukuran, tata letak, maupun fasilitas lainnya (Vidyaprabha et al., n.d.). Industri apartemen di Indonesia terus berkembang pesat terutama di kota-kota besar seperti Surabaya. Berdasarkan data Colliers, total pasokan apartemen di Surabaya hingga semester pertama tahun 2024 mencapai 54.487 unit. Tingkat hunian mengalami peningkatan sebesar 1,5% dibandingkan semester sebelumnya, sehingga mencapai 58,1%. Hal ini mencerminkan adanya pergerakan positif di pasar properti meskipun tantangan ekonomi masih memengaruhi sektor ini (Colliers, 2024). Data ini menunjukkan adanya potensi pertumbuhan dalam permintaan apartemen di Surabaya. Dengan banyaknya pilihan apartemen, persaingan antar pengelola semakin ketat, sehingga mempertahankan penyewa menjadi tantangan besar. Untuk menjaga stabilitas pendapatan, diperlukan strategi yang efektif guna meminimalkan churn atau berhentinya penyewa. Salah satu strategi yang diterapkan adalah menggunakan metode *Gradient Boosting* dan *CatBoost*. Perbandingan antara kedua metode ini akan dilakukan untuk mengetahui metode mana yang lebih unggul dalam memprediksi kemungkinan churn penyewa.

Apartemen Trillium merupakan salah satu apartemen unggulan di Surabaya yang diminati masyarakat terutama bagi penyewa yang mencari tempat tinggal dengan harga terjangkau dan lokasi strategis yang berdekatan dengan area perkantoran, sekolah, dan kampus. Apartemen ini juga didukung dengan pelayanan setara hotel bintang lima serta keamanan maksimal dari Badan Pengelola sehingga berhasil menciptakan rasa aman dan nyaman yang meningkatkan daya tarik bagi penyewa (Donisisus et al., 2024). Di Apartemen Trillium, tingkat churn penyewa menjadi salah satu permasalahan utama yang memengaruhi tingkat hunian. Meskipun data riwayat sewa tersedia, pengelola belum memiliki pendekatan berbasis data untuk memprediksi kemungkinan churn. Akibatnya, strategi retensi masih bersifat reaktif, bukan proaktif. Oleh karena itu, diperlukan analisis mendalam untuk mengidentifikasi pola churn agar langkah-langkah preventif dapat diambil.

Gradient Boosting adalah algoritma *machine learning* yang tergolong dalam metode *ensemble learning*. Algoritma ini membangun model yang kuat dengan

menggabungkan sejumlah model sederhana (*weak learners*) secara bertahap. Prosesnya bertujuan untuk mengurangi kesalahan prediksi model sebelumnya, dengan lebih menitikberatkan pada pola-pola yang sulit dikenali dalam data. Secara khusus, algoritma ini memberikan bobot tambahan pada model, sehingga fokusnya lebih terarah pada data yang sulit dijelaskan oleh model sebelumnya (Septian, n.d.). *CatBoost* adalah algoritma *machine learning* berbasis *boosting* yang dirancang untuk menyelesaikan masalah klasifikasi dan regresi. Algoritma ini tidak memerlukan pra-pemrosesan kompleks, seperti *one-hot encoding*, yang sering digunakan pada metode lain. Sebagai bagian dari keluarga algoritma *Gradient Boosting*, *CatBoost* memiliki sejumlah peningkatan, membuatnya lebih stabil, lebih cepat, dan mampu mengolah fitur kategorikal secara langsung (Pahlevi et al., 2024).

Penelitian terdahulu telah membuktikan bahwa *CatBoost* unggul dalam menangani fitur kategorikal tanpa langkah pra-pemrosesan tambahan, sehingga menghasilkan akurasi tinggi (94%) dengan *recall* 97%, yang memastikan hampir semua pelanggan *churn* teridentifikasi. Algoritma ini mencegah *overfitting* melalui pendekatan *ordered boosting*, membuat model lebih andal saat menghadapi data baru. Dibandingkan algoritma lain seperti *Stochastic Gradient Boosting* dan *XGBoost*, *CatBoost* menawarkan performa yang lebih cepat dan optimal, menjadikannya alat yang sangat efektif untuk prediksi *churn*. Hasilnya membantu perusahaan telekomunikasi fokus pada pelanggan berisiko *churn* dan merancang intervensi berbasis data yang tepat sasaran (Hasan et al., 2024).

Penelitian lain juga menyatakan bahwa *Gradient Boosting* lebih unggul dibandingkan *random forest* dalam mengklasifikasi *churn* pelanggan Telco. *Gradient Boosting* mencapai akurasi 83% dan nilai *ROC-AUC* 0.89, lebih tinggi dibandingkan *random forest* yang masing-masing mencatat 81% dan 0.87. Dengan kemampuan mengoptimalkan kesalahan prediksi secara iteratif, *Gradient Boosting* memberikan hasil yang lebih akurat dan stabil, menjadikannya algoritma yang lebih efektif untuk menangani pola *churn* pelanggan yang kompleks. Hasil ini menunjukkan bahwa *Gradient Boosting* lebih baik dalam memberikan prediksi yang akurat sekaligus efisien dari segi waktu, sehingga dapat membantu perusahaan telekomunikasi merancang strategi retensi yang lebih efektif dan cepat (Adji Purnama et al., n.d.).

Berdasarkan penjelasan di atas maka *Gradient Boosting* dan *CatBoost* adalah metode yang cocok untuk kasus prediksi *churn* karena keduanya dapat menangani data yang kompleks dan heterogen, serta mengidentifikasi interaksi non-linear antar fitur.

Hal ini juga didukung oleh penelitian terdahulu yang menunjukkan bahwa kedua metode ini memberikan kinerja yang sangat baik dalam berbagai aplikasi prediksi *churn*, terutama yang melibatkan data dengan struktur yang kompleks dan variabel yang beragam.

Berdasarkan keunggulan masing-masing algoritma, penelitian ini dilakukan untuk membandingkan performa *Gradient Boosting* dan *CatBoost* dalam memprediksi *churn* di Apartemen Trillium. *Gradient Boosting* dikenal dengan kemampuannya yang iteratif dalam memperbaiki kesalahan model sebelumnya, sehingga menghasilkan prediksi yang akurat pada dataset dengan pola yang kompleks. Di sisi lain, *CatBoost* unggul dalam mengolah data kategorikal secara langsung tanpa memerlukan pra-pemrosesan yang rumit, seperti *one-hot encoding*, yang membuatnya lebih efisien dalam hal waktu pelatihan dan inferensi. Penelitian ini bertujuan untuk menentukan algoritma terbaik yang tidak hanya dapat memberikan akurasi prediksi yang tinggi tetapi juga mampu menangani karakteristik unik data pelanggan Apartemen Trillium, termasuk data demografis, riwayat sewa, dan preferensi layanan. Hasil dari penelitian ini diharapkan dapat memberikan wawasan bagi pengelola apartemen untuk merancang strategi retensi pelanggan yang lebih tepat sasaran dan efektif.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, rumusan masalah yang akan dibahas dalam penelitian ini adalah:

1. Bagaimana performa model *Gradient Boosting* dalam memprediksi *churn* penyewa di Apartemen Trillium?
2. Bagaimana performa model *CatBoost* dalam memprediksi *churn* penyewa di Apartemen Trillium?
3. Model mana yang lebih unggul untuk diterapkan dalam kasus *churn* di Apartemen Trillium, dan apa justifikasi pemilihannya berdasarkan hasil analisis?
4. Apa saja faktor-faktor utama yang memengaruhi *churn* penyewa di Apartemen Trillium berdasarkan data historis?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang ditetapkan, penelitian ini bertujuan untuk:

1. Menganalisis performa model *Gradient Boosting* dalam memprediksi *churn* penyewa di Apartemen Trillium.
2. Menganalisis performa model *CatBoost* dalam memprediksi *churn* penyewa di Apartemen Trillium.
3. Menentukan model yang lebih unggul untuk diterapkan dalam kasus *churn* di Apartemen Trillium.
4. Mengidentifikasi faktor-faktor utama yang memengaruhi *churn* penyewa di Apartemen Trillium berdasarkan data historis.

1.4 Manfaat Penelitian

Penelitian ini memberikan manfaat bagi sejumlah pihak, antara lain:

1. Manfaat Akademis

- a. Kontribusi Ilmiah

Penelitian ini menambah referensi ilmiah tentang penerapan metode *Gradient Boosting* dan *CatBoost* dalam prediksi *churn* pada sektor properti, yang masih jarang dibahas dalam konteks lokal Indonesia.

- b. Pengembangan Pengetahuan

Memberikan wawasan baru mengenai kelebihan dan kekurangan kedua metode dalam menangani data *churn* yang kompleks dan tidak seimbang (*imbalanced data*).

- c. Landasan untuk Penelitian Selanjutnya

Hasil penelitian dapat menjadi dasar bagi studi lanjutan dalam pengembangan model prediksi *churn* di sektor apartemen atau sektor properti lainnya.

2. Manfaat Praktis

- a. Bagi Manajemen Apartemen Trillium

- Membantu Apartemen Trillium memahami faktor-faktor yang memengaruhi *churn* penyewa sehingga dapat mengambil langkah strategis untuk meningkatkan retensi penghuni.
- Memberikan rekomendasi model prediksi *churn* yang lebih efektif untuk digunakan dalam pengambilan keputusan berbasis data.

- b. Bagi Industri Properti

Menyediakan panduan bagi pelaku industri properti dalam memanfaatkan teknologi pembelajaran mesin untuk mengelola *churn*, sehingga dapat meningkatkan efisiensi operasional dan daya saing di pasar.

3. Manfaat Sosial dan Ekonomi

a. Stabilitas Komunitas Penghuni

Mengurangi *churn* di apartemen dapat menciptakan komunitas penghuni yang lebih stabil dan mendukung lingkungan sosial yang harmonis.

b. Efisiensi dalam Penyediaan Hunian

Dengan retensi penyewa yang lebih tinggi, apartemen dapat mengalokasikan sumber daya lebih baik untuk meningkatkan fasilitas dan layanan, yang pada akhirnya meningkatkan kesejahteraan penghuni.

1.5 Batasan Masalah

Agar penelitian tetap terfokus pada isu yang dikaji dan tidak menyimpang dari topik utama, diperlukan penetapan batasan masalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini diambil dari database Apartemen Trillium selama periode dua tahun.
2. Algoritma yang dibandingkan adalah Gradient Boosting dan CatBoost untuk memprediksi pelanggan yang *churn*.
3. Proses pengolahan data sebelum pengujian mencakup langkah-langkah seperti pembersihan data (data cleaning), penanganan data yang tidak seimbang (imbalanced data), serta transformasi fitur bila diperlukan.
4. Evaluasi model dilakukan menggunakan metrik-metrik seperti akurasi, presisi, recall, F1-score, dan Area Under the Curve (AUC).
5. Analisis dilakukan untuk mengidentifikasi atribut-atribut yang paling berpengaruh terhadap perilaku *churn* pelanggan.