

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini bertujuan untuk membangun model *Machine Learning* untuk memprediksi lama menginap tamu (*length of stay*) pada Hotel XYZ dengan memanfaatkan data transaksi yang tersimpan dalam sistem internal hotel. Proses penelitian dilakukan secara bertahap dan sistematis menggunakan *framework* KDD, dimulai dari data preprocessing untuk memastikan kualitas dan konsistensi data, dilanjutkan dengan analisis eksploratif guna memahami karakteristik transaksi dan pola perilaku tamu. Selanjutnya, data digunakan untuk melatih dua algoritma regresi, yaitu *Random Forest* dan *Gradient Boosting*, yang kemudian dievaluasi menggunakan metrik regresi yang sesuai. Pendekatan ini memungkinkan penelitian menghasilkan model prediktif yang tidak hanya berbasis data aktual, tetapi juga dapat direplikasi dan diterapkan dalam konteks operasional hotel.

Berdasarkan hasil evaluasi kinerja model pada data tahun 2024, model *Gradient Boosting* menunjukkan kinerja prediksi yang lebih stabil dengan tingkat kesalahan yang lebih rendah dibandingkan model *Random Forest*. Nilai koefisien determinasi yang tinggi serta nilai kesalahan yang relatif kecil menunjukkan bahwa model mampu menangkap pola hubungan antara variabel input dan lama menginap tamu secara konsisten. Model ini dapat menjelaskan sebagian besar variasi pada data target tanpa menunjukkan fluktuasi kesalahan yang signifikan antar segmen. Sementara itu, model *Random Forest* tetap memberikan hasil yang relevan, khususnya dalam membantu memahami struktur data dan hubungan antar variabel, meskipun performa prediksinya tidak sebaik *Gradient Boosting*.

Hasil analisis *feature importance* memberikan gambaran mengenai faktor-faktor utama yang berkontribusi terhadap variasi lama menginap tamu. Variabel jumlah tamu, jumlah pemesanan, dan pendapatan kamar muncul sebagai faktor yang paling berpengaruh dalam model. Temuan ini menunjukkan bahwa karakteristik skala pemesanan dan nilai transaksi memiliki keterkaitan yang kuat

dengan durasi menginap tamu. Identifikasi faktor-faktor tersebut diperoleh langsung dari proses pemodelan dan analisis data transaksi, sehingga memberikan dasar analitis yang objektif dalam memahami perilaku tamu Hotel XYZ tanpa bergantung pada asumsi subjektif.

Hasil prediksi yang dihasilkan oleh model selanjutnya diterjemahkan ke dalam alur keputusan operasional yang dirancang untuk mendukung kebutuhan manajemen Hotel XYZ. Alur keputusan ini memanfaatkan informasi prediksi lama menginap, kontribusi *room-nights* per segmen, serta evaluasi kesalahan prediksi sebagai dasar dalam perencanaan operasional. Dengan pendekatan ini, hasil model tidak berhenti pada tahap analisis statistik, tetapi dapat digunakan sebagai pendukung pengambilan keputusan yang lebih terstruktur, khususnya dalam pengelolaan segmen pelanggan, perencanaan layanan operasional, dan pengendalian kualitas prediksi secara berkelanjutan.

Sebagai tahap pengujian penggunaan model, penelitian ini juga menerapkan model yang telah dibangun pada data transaksi tahun 2025 yang tersedia hingga bulan September, dengan total 1.770 data setelah proses pembersihan. Hasil pengujian menunjukkan bahwa model *Gradient Boosting* tetap menghasilkan kinerja prediksi yang konsisten dengan nilai R^2 sebesar 0.9977, RMSE sebesar 0.4084, MAE sebesar 0.1708, dan MAPE sebesar 2.28%, serta lebih baik dibandingkan *Random Forest* pada periode yang sama. Hasil ini menunjukkan bahwa model dapat digunakan secara konsisten pada data periode berbeda dalam konteks operasional Hotel XYZ. Penelitian ini menyimpulkan bahwa model *Gradient Boosting* merupakan pendekatan yang paling sesuai untuk memprediksi lama menginap tamu berdasarkan data transaksi yang tersedia. Model ini dapat dimanfaatkan sebagai pendukung analisis dan pengambilan keputusan operasional berbasis data, serta menjadi dasar bagi penerapan sistem prediksi lama menginap secara berkelanjutan di Hotel XYZ.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilaksanakan, penulis mengajukan beberapa saran untuk pengembangan penelitian selanjutnya dan penyempurnaan

penerapan model prediksi dalam konteks industri perhotelan. Saran-saran ini disusun dengan mempertimbangkan keterbatasan penelitian, ruang lingkup data, serta potensi penerapan teknologi *Machine Learning* yang lebih luas di masa mendatang. Adapun saran-saran tersebut adalah sebagai berikut:

- 1) Penelitian ini menggunakan dataset transaksi tamu dari Hotel XYZ dengan cakupan waktu tahun 2023 hingga tahun 2025 dan atribut tertentu. Meskipun data tersebut telah mencukupi untuk membangun model prediksi yang akurat, namun representasi data masih terbatas pada satu sumber institusional yaitu data internal tahun 2024. Penelitian berikutnya disarankan untuk memperluas cakupan dataset dengan menambahkan data dari berbagai cabang hotel, segmentasi pelanggan, maupun periode waktu yang berbeda. Selain itu, integrasi data eksternal seperti tren wisata, musim liburan, atau acara lokal juga dapat memberikan konteks tambahan yang memperkaya model, sehingga meningkatkan kemampuan generalisasi prediksi pada berbagai kondisi bisnis.
- 2) Pada penelitian ini, analisis difokuskan pada dua algoritma utama, yaitu *Random Forest* dan *Gradient Boosting*. Hasil menunjukkan bahwa *Gradient Boosting* memberikan performa terbaik dalam akurasi prediksi, sedangkan *Random Forest* unggul dalam interpretabilitas. Untuk penelitian lanjutan, disarankan agar dilakukan eksplorasi terhadap algoritma lain seperti *XGBoost*, *LightGBM*, atau *CatBoost* yang memiliki efisiensi komputasi lebih tinggi dan kemampuan menangani dataset besar dengan lebih optimal. Penggunaan metode *Deep Learning* juga dapat dipertimbangkan untuk melihat sejauh mana model berbasis jaringan saraf mampu meningkatkan presisi hasil prediksi.
- 3) Sistem prediksi yang dikembangkan dalam penelitian ini masih bersifat *offline*, di mana hasil prediksi diperoleh setelah proses komputasi selesai dijalankan. Untuk tahap pengembangan selanjutnya, disarankan agar sistem prediksi diintegrasikan secara *real-time* ke dalam sistem manajemen hotel atau *Property Management System* (PMS). Dengan integrasi ini, manajemen hotel dapat memantau perubahan perilaku pelanggan dan tingkat okupansi secara dinamis, sehingga strategi harga (*dynamic pricing*) dan perencanaan sumber daya dapat disesuaikan dengan kondisi aktual secara langsung.

- 4) Walaupun model *Gradient Boosting* menunjukkan akurasi yang sangat tinggi, interpretasi hasil prediksi masih terbatas karena sifat algoritmanya yang kompleks. Oleh karena itu, penelitian mendatang disarankan untuk menggabungkan metode interpretabilitas seperti *SHAP* (*SHapley Additive Explanations*) atau *LIME* (*Local Interpretable Model-Agnostic Explanations*). Pendekatan ini dapat membantu menjelaskan secara kuantitatif kontribusi setiap variabel terhadap hasil prediksi, sehingga pihak manajemen hotel dapat memahami faktor-faktor yang paling berpengaruh dalam meningkatkan lama menginap tamu.
- 5) Model yang dibangun pada penelitian ini telah menunjukkan performa yang unggul berdasarkan data historis tahun penelitian, namun belum diuji untuk kestabilan jangka panjang. Penelitian berikutnya diharapkan dapat melakukan evaluasi berkelanjutan terhadap performa model dengan memasukkan data baru secara periodik. Langkah ini penting untuk memastikan model tetap relevan terhadap perubahan tren tamu, pergeseran pasar, dan kondisi ekonomi yang dinamis di industri perhotelan.

Saran-saran di atas diharapkan dapat menjadi acuan bagi peneliti selanjutnya dalam mengembangkan penelitian serupa maupun memperdalam kajian mengenai penerapan *Machine Learning* di sektor perhotelan. Melalui pendekatan yang lebih holistik dan adaptif, diharapkan model prediksi yang dihasilkan pada penelitian mendatang tidak hanya unggul secara teknis, tetapi juga mampu memberikan kontribusi nyata terhadap peningkatan kualitas pelayanan, efisiensi operasional, dan pengambilan keputusan strategis yang berbasis data di industri perhotelan.