

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil analisis dan implementasi sistem prioritas lead serta rekomendasi tindak lanjut pada data operasional, penelitian ini menyimpulkan bahwa rancangan two-stage prediction efektif untuk memisahkan tugas pemeringkatan peluang booking dan penentuan kanal follow-up yang efisien. Model dikembangkan melalui pipeline praproses yang konsisten, pemisahan data berjenjang (train-validation-test), serta hyperparameter tuning sehingga kinerja yang dilaporkan merefleksikan kondisi operasional dengan keterbatasan kapasitas eksekusi.

Pada tahap pemeringkatan booking (Stage-1), XGBoost menunjukkan kinerja terbaik dengan Recall@80% sebesar 93,00% dan Lift@80% = 1,162 \times . Nilai lift ini menandakan bahwa pada 80% lead teratas, sistem berpotensi meningkatkan conversion rate sekitar 16,2% dibandingkan pemilihan acak, sekaligus menekan missed opportunities melalui jumlah false negative yang paling rendah di antara model pembanding. Hasil ini menegaskan bahwa pemeringkatan probabilistik yang diselaraskan dengan batas kapasitas eksekusi mampu memaksimalkan penangkapan peluang booking yang benar-benar terjadi. Pada tahap rekomendasi kanal (Stage-2), XGBoost kembali unggul dengan accuracy 80,7%, F1-Score 0,707, ROC-AUC 0,8685, dan Lift@80% 1,219 \times . Model mampu mengidentifikasi 83,9% pelanggan yang layak PHONE (kanal berbiaya lebih rendah) sekaligus mempertahankan cakupan 97,51% untuk VISIT pada kapasitas 80%. Di tingkat operasional, false positive untuk rekomendasi VISIT berkurang secara signifikan (± 11.810 kasus), sehingga pemborosan kunjungan lapangan dapat ditekan dan efisiensi biaya tindak lanjut meningkat.

Analisis feature importance memperlihatkan perbedaan logika keputusan antartahap, yang secara langsung menjawab poin 3 rumusan masalah mengenai variabel-variabel paling berpengaruh terhadap keberhasilan booking pelanggan dan

pemilihan channel follow-up optimal. Pada Stage-1 (prediksi booking), variabel buss_unit mendominasi dengan nilai gain 1369.71 (kontribusi 84,5%), menunjukkan kuatnya peran segmentasi lini bisnis (Refinancing vs New Motorcycle) terhadap probabilitas booking, karena mencerminkan historical relationship dan trust customer dari engagement sebelumnya. Variabel kedua adalah contract_stage dengan gain 113.34 (7,0%), yang merepresentasikan segmentasi tenor pinjaman (\leq 12 bulan, 12-24 bulan, 24-36 bulan, $>$ 36 bulan) dan mempengaruhi commitment level customer. Variabel lain yang berpengaruh termasuk top (gain 58.57), obj_tahun (30.45), principal (23.05), marital_stat (17.22), cust_prov (16.73), bpkb_same_name (16.54), dp_principal_ratio (16.42), dan month_inst (15.80), yang secara kolektif mengukur spesifikasi produk, stabilitas finansial, kredibilitas, dan affordability customer. Secara keseluruhan, 10 variabel teratas berkontribusi 97,8% dari total importance, menandakan prediksi booking sangat bergantung pada faktor dominan seperti customer origin dan produk. Sebaliknya, pemilihan kanal pada Stage-2 bersifat lebih multifaktor dengan distribusi importance yang merata. Variabel teratas adalah cust_prov dengan gain 29.15 (24,25%), diikuti buss_unit (11.53; 9,59%) dan cust_kec (10.92; 9,08%), menekankan peran geografi (provinsi, kecamatan, kelurahan) yang secara kolektif berkontribusi 38,8% karena mempengaruhi preferensi channel berdasarkan konteks urban vs rural (misalnya, customer di urban centers lebih nyaman dengan PHONE). Variabel finansial seperti dp_principal_ratio, salary, month_inst, principal, inst_salary_ratio, grs_dp, dan salary_group berkontribusi 27,1%, mencerminkan affordability dan financial capability yang memengaruhi efisiensi channel. Variabel demografis termasuk house_stat, ocpt_code, edu_type, marital_stat, dan no_of_depend (kontribusi 16,4%), yang berkorelasi dengan digital literacy dan kebutuhan personal touch. Perbedaan pola ini mengukuhkan keunggulan arsitektur dua tahap: keputusan “siapa yang ditindak” dipandu faktor dominan seperti buss_unit dan contract_stage, sedangkan keputusan “melalui kanal apa” menuntut kombinasi atribut geografis, finansial, dan demografis yang lebih beragam.

Sistem telah diimplementasikan dalam dashboard Streamlit yang mendukung alur unggah data, validasi, penetapan kapasitas, prediksi dua tahap, dan ekspor hasil

(CSV/XLSX). Paket artefak (model, scaler, encoder) dan pipeline yang konsisten memastikan keluaran yang interpretable dan actionable bagi analis non-programmer. Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan two-stage berbasis XGBoost mampu memaksimalkan penangkapan booking pada kondisi capacity-constrained—dengan potensi peningkatan conversion rate ~16% pada tahap prioritisasi—serta mengarahkan tindak lanjut ke kanal yang lebih tepat dan hemat biaya, sehingga keputusan penjualan menjadi lebih efektif dan efisien.

5.2 Saran

Untuk pengembangan lebih lanjut, penelitian ini memiliki beberapa keterbatasan yang dapat ditingkatkan pada pekerjaan mendatang. Rekomendasi berikut disusun sebagai narasi per poin agar sejalan dengan gaya penulisan referensi.

1. Peningkatan kualitas dan kelengkapan data internal perlu dilakukan melalui penataan ulang basis data serta penambahan variabel perilaku pelanggan, meliputi payment behavior (pola janji bayar, riwayat keterlambatan, status tunggakan), interaction history (respons terhadap tindak lanjut sebelumnya, durasi komunikasi, preferensi kanal), dan riwayat transaksi per pelanggan yang lebih komprehensif (frekuensi pembelian, average transaction value, customer lifetime value). Dengan basis data yang lebih terstruktur dan lengkap, model diharapkan semakin akurat dalam menangkap pola perilaku sehingga kinerja prediksi, termasuk Lift@80% yang saat ini sekitar 1,16, berpotensi meningkat.
2. Pemanfaatan variabel eksternal yang relevan perlu diintegrasikan melalui pengayaan fitur temporal dan kontekstual, antara lain pola musiman, indikator ekonomi, kalender hari libur dan periode gajian, riwayat kampanye promosi, serta informasi kompetitor. Integrasi faktor-faktor ini diharapkan membuat model lebih adaptif terhadap dinamika pasar dan perubahan perilaku pelanggan, sehingga proyeksi booking dan efektivitas kanal tindak lanjut menjadi lebih stabil lintas periode.
3. Eksplorasi metodologi lanjutan seiring peningkatan data dan kapasitas komputasi dapat dipertimbangkan dengan menggunakan pendekatan deep

learning, seperti neural networks atau model berbasis transformer, untuk menangkap pola nonlinier dan interaksi antarfaktor yang lebih kompleks. Pada kondisi data dan sumber daya saat ini, XGBoost tetap layak dipertahankan sebagai baseline produksi pada kedua tahap, sementara penerapan metode lanjutan dapat diarahkan bertahap disertai evaluasi terukur, misalnya melalui uji A/B di lingkungan operasional.

4. Implementasi operasional secara penuh terhadap two-stage prediction system yang telah dikembangkan dalam penelitian ini disarankan kepada PT XYZ. Implementasi dapat dilakukan melalui prototipe dashboard berbasis Streamlit yang memungkinkan tim analisis atau data scientist melakukan prediksi batch secara rutin, menghasilkan daftar prioritas leads beserta rekomendasi kanal follow-up (visit atau phone), serta mengekspor hasil prediksi untuk distribusi kepada tim sales, sehingga meningkatkan conversion rate booking dan efisiensi kinerja sales secara keseluruhan.

Dengan saran-saran tersebut, pengembangan sistem lead scoring ke depan diharapkan menjadi lebih akurat, lebih komprehensif dalam mempertimbangkan berbagai faktor, serta lebih strategis dan efektif dalam mendukung optimalisasi kinerja penjualan di PT XYZ guna meningkatkan conversion rate dan revenue generation.

