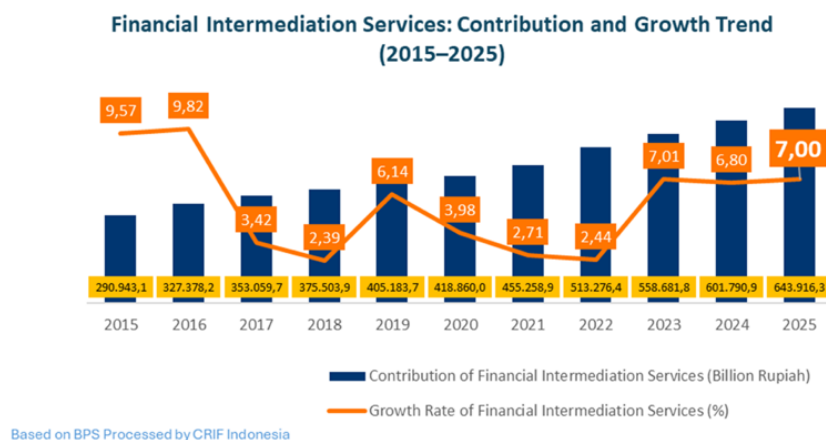


BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri pembiayaan konsumen (*multifinance*) di Indonesia mengalami pertumbuhan yang dinamis dan menjadi salah satu pilar penting dalam ekosistem finansial nasional. Sektor ini telah menunjukkan resiliensi yang kuat dengan total aset mencapai Rp 630,1 triliun pada tahun 2024, didukung oleh 200 perusahaan pembiayaan yang terdiri dari 192 perusahaan konvensional dan 8 perusahaan syariah [1]. Sektor pembiayaan (*financing companies*) mendominasi industri dengan 146 perusahaan dan total piutang mencapai Rp 530,5 triliun, dimana sekitar 80% dari penjualan mobil dan motor baru di Indonesia menggunakan skema pembiayaan [2]. Gambar 1.1 menunjukkan kontribusi dan tren pertumbuhan layanan intermediasi keuangan dari tahun 2015 hingga 2025, dimana terlihat proyeksi pertumbuhan industri multifinance mencapai 7% pada tahun 2025, didorong oleh transformasi digital, peningkatan literasi keuangan, dan permintaan yang kuat terhadap pembiayaan kendaraan dan peralatan [1]. Namun, persaingan yang intensif dalam industri ini mengharuskan perusahaan untuk terus mengoptimalkan strategi akuisisi dan retensi pelanggan agar dapat bertahan dan berkembang di pasar yang kompetitif.



Gambar 1. 1 Kontribusi dan Tren Industri Multifinance Indonesia [1]

Dalam konteks persaingan industri pembiayaan yang ketat, strategi *sales* dan *follow-up* kepada calon pelanggan menjadi faktor kritis yang menentukan tingkat konversi dan efisiensi operasional perusahaan. Studi menunjukkan bahwa implementasi sistem *lead scoring* berbasis *machine learning* dapat meningkatkan *conversion rate* hingga 35% dan mengurangi waktu pemrosesan manual hingga 80% [3]. Lebih lanjut, penelitian dalam konteks B2B menunjukkan bahwa model prediktif berbasis algoritma *Gradient Boosting* mampu mencapai akurasi hingga 98,39% dengan ROC-AUC 0,9891 dalam memprediksi probabilitas konversi *leads* [4]. Tinjauan sistematis terhadap 44 studi *lead scoring* menunjukkan bahwa *Logistic Regression* merupakan salah satu algoritma yang paling banyak diterapkan dalam penelitian prediksi konversi pelanggan, terutama sebagai baseline yang efisien dan interpretable [5]. Sementara itu, algoritma *ensemble* seperti *XGBoost* dan *Random Forest* terbukti efektif dalam menangani data yang tidak seimbang (*imbalanced data*) dan mengidentifikasi pola kompleks yang memengaruhi keputusan pembelian pelanggan [6]. Perbandingan antara algoritma klasik seperti *Logistic Regression* dan algoritma *ensemble* diperlukan untuk mengidentifikasi pendekatan yang paling sesuai dengan karakteristik data dan kebutuhan bisnis spesifik [4] [5]. Namun, sebagian besar penelitian berfokus pada prediksi *leads* yang berpotensi melakukan konversi tanpa mengaitkannya dengan strategi tindak lanjut yang optimal, padahal alokasi channel *follow-up* yang tepat dapat memberikan dampak signifikan terhadap efisiensi biaya dan tingkat keberhasilan konversi [7].

Penelitian terkini menunjukkan bahwa [8] telah mengembangkan model two-stage untuk direct marketing perbankan dengan Neural Network (akurasi 80%, sensitivitas 67%) dan K-means clustering (profit-to-cost ratio 4,30), namun belum menangani class imbalance dan tidak mengintegrasikan rekomendasi channel berbasis biaya operasional; [9] berhasil meningkatkan konversi lead pendidikan online melalui Stacking (akurasi 92,33%, F1-score 89,39%) dan Logistic Regression dengan threshold optimal 0,20 (akurasi 90,19%), tetapi hanya berfokus pada prediksi konversi tanpa rekomendasi channel visit/phone; serta [4] mencapai performa tertinggi pada prioritasasi lead B2B dengan Gradient Boosting (akurasi

98,39%, ROC-AUC 0,9891, efisiensi +40%), namun tidak menyertakan tahap rekomendasi channel cost-sensitive maupun implementasi dashboard. Berbeda dengan ketiga penelitian tersebut, penelitian ini mengusulkan sistem two-stage berbasis XGBoost yang tidak hanya memprediksi probabilitas booking lead, melainkan juga merekomendasikan channel follow-up optimal (kunjungan langsung atau telepon) dengan mempertimbangkan biaya operasional dan kapasitas tim sales di PT XYZ, dilengkapi penanganan class imbalance, hyperparameter tuning, serta dashboard Streamlit interaktif — sehingga mengisi kesenjangan penerapan predictive lead management yang praktis dan kontekstual pada industri pembiayaan konsumen di Indonesia.

Proses *follow-up* pelanggan dalam industri pembiayaan konsumen umumnya dilakukan melalui dua *channel* utama: kunjungan langsung (*visit*) oleh tim *sales* dan komunikasi melalui telepon (*phone*). Kedua *channel* ini memiliki karakteristik yang berbeda dalam hal alokasi sumber daya, biaya operasional, dan tingkat konversi. Kunjungan langsung memerlukan investasi yang lebih besar meliputi waktu, biaya transportasi, dan tenaga kerja, namun berpotensi menghasilkan tingkat konversi yang lebih tinggi karena kemampuan membangun hubungan personal yang lebih kuat dengan calon pelanggan [10]. Di sisi lain, *follow-up* melalui telepon lebih efisien dari segi biaya dan waktu, namun memiliki keterbatasan dalam membangun *engagement* yang mendalam [11]. Perbedaan karakteristik kedua *channel* ini mengindikasikan pentingnya strategi alokasi yang tepat untuk memaksimalkan efisiensi operasional dan *return on investment* dari aktivitas *sales*. PT XYZ, sebagai salah satu perusahaan pembiayaan konsumen yang beroperasi di Indonesia dengan basis pelanggan yang luas, menghadapi tantangan serupa dalam mengoptimalkan strategi *follow-up* pelanggan. Data historis perusahaan menunjukkan bahwa kunjungan langsung menghasilkan *conversion rate* sebesar 12,98%, sementara *follow-up* melalui telepon mencapai 11,12%. Meskipun perbedaan *conversion rate* tampak relatif kecil (1,86%), dampaknya terhadap efisiensi biaya sangat signifikan mengingat biaya kunjungan langsung jauh lebih mahal dibandingkan *follow-up* via telepon. Sebagai perusahaan dengan data historis

yang komprehensif, PT XYZ memiliki potensi besar untuk memanfaatkan pendekatan *data-driven* dalam mengoptimalkan strategi *sales*.

Permasalahan yang dihadapi PT XYZ saat ini adalah tim *sales* menghadapi keterbatasan kapasitas dalam melakukan *follow-up* terhadap seluruh *leads* yang tersedia. Dengan *resource* yang terbatas, tidak semua calon pelanggan dapat di-*follow-up*, sehingga diperlukan prioritisasi yang tepat untuk memaksimalkan peluang keberhasilan *booking*. Proses penentuan prioritas dan pemilihan *channel follow-up* saat ini masih dilakukan secara manual berdasarkan pengalaman dan intuisi subjektif dari tim *sales*. Pendekatan ini menyebabkan beberapa kendala operasional: pertama, tidak adanya sistem prioritas yang objektif menyebabkan hilangnya peluang bisnis dari pelanggan berpotensi tinggi yang tidak sempat di-*follow-up* karena kapasitas terbatas; kedua, alokasi *resource* yang tidak optimal dimana kunjungan langsung dilakukan kepada pelanggan yang sebenarnya cukup di-*follow-up* via telepon atau bahkan tidak perlu diprioritaskan sama sekali; ketiga, inkonsistensi keputusan antar anggota tim *sales* yang dapat mempengaruhi performa keseluruhan; dan keempat, tingginya biaya operasional akibat kunjungan yang tidak efektif.

Secara lebih spesifik, sistem pengelolaan leads yang berjalan di PT XYZ saat ini mengikuti alur dimana data calon pelanggan potensial diperoleh dari database CRM melalui proses query menggunakan platform Toad for Oracle dengan kriteria yang telah ditetapkan perusahaan, seperti pelanggan dengan kontrak aktif yang mendekati masa jatuh tempo untuk penawaran *refinancing*. Data hasil query tersebut kemudian langsung diunggah ke CRM System untuk didistribusikan kepada tim sales di lapangan tanpa melalui proses penyaringan atau prioritas berbasis analitik. Akibatnya, seluruh leads diperlakukan setara tanpa mempertimbangkan probabilitas keberhasilan *booking* maupun kesesuaian *channel follow-up*, sehingga tim sales harus menentukan sendiri leads mana yang akan dihubungi terlebih dahulu dan melalui kanal apa berdasarkan pertimbangan subjektif masing-masing.

Berbeda dengan kondisi tersebut, sistem yang dikembangkan dalam penelitian ini menambahkan lapisan analitik berbasis *machine learning* di antara proses query data dan distribusi ke tim sales. Pada sistem yang diusulkan, hasil output query berupa data leads potensial akan diproses terlebih dahulu melalui *two-stage prediction model* sebelum didistribusikan ke tim sales, **dimana** Stage 1 memprediksi probabilitas keberhasilan *booking* untuk menghasilkan peringkat prioritas leads, kemudian Stage 2 merekomendasikan *channel follow-up* optimal (kunjungan langsung atau telepon) berdasarkan karakteristik masing-masing pelanggan. Dengan demikian, tim sales menerima daftar leads yang telah diurutkan berdasarkan potensi konversi tertinggi beserta rekomendasi kanal yang paling efektif, sehingga alokasi waktu dan sumber daya dapat difokuskan pada leads dengan peluang keberhasilan tinggi melalui kanal yang tepat.

Lebih lanjut, tanpa sistem prioritisasi berbasis data, tim *sales* berisiko mengalokasikan waktu dan tenaga pada *leads* dengan probabilitas *booking* rendah, sementara *leads* berpotensi tinggi justru terlewatkan. Kondisi ini mengakibatkan *opportunity cost* yang signifikan karena *resource* yang seharusnya difokuskan pada pelanggan dengan peluang konversi tertinggi malah tersebar tidak merata. Bahkan setelah prioritisasi dilakukan, masih diperlukan keputusan optimal mengenai *channel follow-up* yang paling efektif untuk setiap pelanggan—apakah melalui kunjungan langsung yang lebih personal namun membutuhkan biaya tinggi, atau melalui telepon yang lebih efisien namun memiliki tingkat konversi lebih rendah. Oleh karena itu, diperlukan pendekatan bertahap (*two-stage approach*) [12] yang dapat (1) memprediksi probabilitas keberhasilan *booking* untuk melakukan prioritisasi leads secara objektif, dan (2) merekomendasikan *channel follow-up* optimal bagi leads yang telah diprioritaskan, sehingga dapat memaksimalkan efektivitas alokasi sumber daya tim sales dan meminimalkan *opportunity cost* dari keputusan yang kurang tepat [13].

Perkembangan teknologi *machine learning* dan *data analytics* menawarkan solusi untuk mengatasi permasalahan tersebut. Data historis pelanggan yang mencakup informasi demografis, riwayat finansial, karakteristik aset, dan hasil

interaksi dapat dianalisis untuk mengidentifikasi pola dan memprediksi probabilitas *booking* serta *channel follow-up* yang optimal. Penelitian terdahulu menunjukkan bahwa algoritma *ensemble* seperti *XGBoost* dan *Random Forest* telah berhasil diterapkan dalam berbagai kasus klasifikasi *customer analytics* dengan tingkat akurasi tinggi, khususnya dalam menangani data *imbalanced* yang umum ditemukan pada konteks bisnis [14]. Pendekatan bertahap (*staged approach*) dalam *predictive modeling* juga telah terbukti meningkatkan akurasi prediksi dengan memisahkan proses prediksi menjadi tahapan-tahapan yang lebih spesifik [15][16]. PT XYZ sebagai perusahaan pembiayaan konsumen memiliki *database customer* yang komprehensif dengan data historis *leads refinancing* yang dapat dimanfaatkan untuk mengembangkan model prediktif. Pengembangan sistem rekomendasi berbasis *machine learning* diharapkan dapat meningkatkan efisiensi operasional tim *sales*, mengurangi biaya kunjungan yang tidak produktif, dan meningkatkan *conversion rate* secara keseluruhan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan *two-stage prediction model* berbasis *machine learning* yang dapat: (1) memprediksi probabilitas keberhasilan *booking* untuk melakukan prioritisasi *leads* secara objektif berdasarkan karakteristik pelanggan, dan (2) merekomendasikan *channel follow-up* optimal (*visit* atau *phone*) bagi *leads* yang telah diprioritaskan. Penelitian ini membandingkan tiga algoritma klasifikasi—*XGBoost*, *Random Forest*, dan *Logistic Regression*—untuk mengidentifikasi model dengan performa terbaik pada setiap tahap prediksi. Pemilihan ketiga algoritma ini didasarkan pada keunggulan masing-masing: *XGBoost* dan *Random Forest* untuk menangani kompleksitas data dan interaksi non-linear, serta *Logistic Regression* sebagai baseline yang interpretable dan efisien komputasi. Implementasi sistem dilakukan melalui *dashboard* interaktif berbasis *Streamlit* yang memungkinkan tim analis menghasilkan daftar prioritas pelanggan beserta rekomendasi *channel follow-up* untuk didistribusikan kepada tim *sales*, meningkatkan pengambilan keputusan yang lebih objektif, terukur, dan efisien.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengembangkan two-stage prediction model untuk (1) memprediksi probabilitas keberhasilan booking guna melakukan prioritisasi leads secara objektif, dan (2) merekomendasikan channel follow-up optimal (visit atau phone) untuk leads yang telah diprioritaskan?
2. Algoritma machine learning manakah (XGBoost, Random Forest, atau Logistic Regression) yang memberikan performa terbaik untuk masing-masing stage prediksi pada konteks prioritisasi booking dan rekomendasi channel follow-up?
3. Variabel-variabel apa saja yang paling berpengaruh terhadap keberhasilan booking pelanggan dan pemilihan channel follow-up optimal?
4. Bagaimana mengintegrasikan two-stage prediction system ke dalam dashboard interaktif yang dapat digunakan oleh tim analis untuk menghasilkan daftar prioritas pelanggan beserta rekomendasi channel follow-up dan mengeksponnya untuk didistribusikan kepada tim sales?

1.3 Batasan Masalah

Untuk memfokuskan penelitian dan memastikan hasil yang optimal, batasan masalah yang ditetapkan adalah:

1. Data yang digunakan adalah data historis pelanggan PT XYZ periode Q2 2025 (April-Juni 2025).
2. Penelitian fokus pada klasifikasi binary: channel follow-up VISIT atau PHONE, tidak mencakup channel lain seperti WhatsApp, email, atau media sosial.
3. Dataset yang digunakan untuk modeling pada two-stage prediction mencakup:
 - a. Stage 1 - Prediksi Booking (Prioritisasi): Seluruh customer dengan informasi is_book, termasuk yang tidak jadi booking

(NO_BOOKING: 86%) dan yang berhasil booking (BOOKING: 14%)

- b. Stage 2 - Rekomendasi Channel Follow-Up: Seluruh customer yang memiliki data last_type_fu dengan distribusi PHONE (68%) dan VISIT (32%)
4. Algoritma machine learning yang dibandingkan terbatas pada tiga metode: XGBoost, Random Forest, dan Logistic Regression pada masing-masing stage prediksi.
5. Framework metodologi yang digunakan adalah CRISP-DM (Cross Industry Standard Process for Data Mining).
6. Metrik evaluasi model mencakup Accuracy, Precision, Recall, F1-Score, AUC-ROC, serta metrik evaluasi dari segi bisnis untuk kedua stage prediksi.
7. Deployment sistem menggunakan dashboard Streamlit untuk interface dan tidak mencakup integrasi dengan sistem CRM existing perusahaan.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Adapun tujuan penelitian yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Mengembangkan two-stage prediction model untuk (1) memprediksi probabilitas keberhasilan booking guna melakukan prioritas leads secara objektif, dan (2) merekomendasikan channel follow-up optimal (visit atau phone) untuk leads yang telah diprioritaskan berdasarkan karakteristik pelanggan di PT XYZ.
2. Membandingkan performa tiga algoritma machine learning (XGBoost, Random Forest, dan Logistic Regression) pada masing-masing stage prediksi untuk mengidentifikasi kombinasi model terbaik berdasarkan metrik Accuracy, Precision, Recall, F1-Score, dan AUC-ROC pada konteks prioritas booking dan rekomendasi channel follow-up.

3. Mengidentifikasi variabel-variabel yang paling berpengaruh terhadap keberhasilan booking pelanggan dan pemilihan channel follow-up optimal melalui feature importance analysis pada masing-masing stage prediksi.
4. Mengembangkan dashboard berbasis Streamlit untuk mendukung evaluasi two-stage prediction model, yang memungkinkan prediksi batch, visualisasi prioritas leads beserta rekomendasi kanal follow-up, serta simulasi dampak terhadap optimasi kinerja sales di PT XYZ.

1.4.2 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi akademis melalui studi empiris penerapan two-stage prediction model menggunakan algoritma machine learning (XGBoost, Random Forest, Logistic Regression) serta implementasi framework CRISP-DM dalam optimasi strategi sales di industri pembiayaan konsumen, khususnya dalam konteks prioritas leads dan optimasi channel follow-up.
2. Menyediakan sistem pendukung keputusan berbasis data bagi PT XYZ yang mampu memprediksi probabilitas keberhasilan booking untuk prioritas leads secara objektif dan merekomendasikan channel follow-up optimal untuk leads yang telah diprioritaskan, sehingga dapat meningkatkan efisiensi alokasi sumber daya tim sales, mengurangi opportunity cost dari leads yang terlewatkan, menekan biaya operasional kunjungan yang tidak produktif, serta meningkatkan conversion rate.
3. Memberikan framework praktis berbasis pendekatan bertahap (staged approach) yang dapat diadaptasi oleh industri pembiayaan lain dalam mengoptimalkan strategi customer relationship management dan mendemonstrasikan nilai bisnis dari penerapan predictive modeling untuk meningkatkan efektivitas tim sales.
4. Mempermudah tim analis dalam menghasilkan daftar prioritas leads berbasis data dan membantu tim sales dalam menentukan channel follow-

up yang tepat untuk setiap pelanggan, meningkatkan produktivitas dengan fokus pada aktivitas berpotensi tinggi, serta memberikan insight mengenai karakteristik pelanggan yang memengaruhi keberhasilan booking dan efektivitas channel follow-up.

1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran yang terstruktur mengenai isi dari tiap bab yang dibahas. Adapun sistematika penulisan skripsi ini adalah sebagai berikut:

A. BAB I PENDAHULUAN

Bab ini menguraikan latar belakang penelitian mengenai permasalahan optimasi channel follow-up pada tim sales PT XYZ, merumuskan masalah penelitian, menjelaskan batasan penelitian yang digunakan, serta memaparkan tujuan dan manfaat penelitian.

B. BAB II LANDASAN TEORI

Bab ini membahas teori dan konsep yang mendasari penelitian, seperti Customer Relationship Management (CRM), strategi sales follow-up, konsep dasar machine learning untuk klasifikasi, algoritma yang digunakan (XGBoost, Random Forest, Logistic Regression), framework CRISP-DM, metrik evaluasi model, serta tinjauan penelitian terdahulu yang relevan.

C. BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi penelitian menggunakan pendekatan CRISP-DM, mulai dari tahap pemahaman data, persiapan data, pemodelan, evaluasi, hingga perencanaan deployment sistem rekomendasi channel follow-up.

D. IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian berupa analisis data, hasil preprocessing, implementasi dan evaluasi model klasifikasi, perbandingan

performa antar algoritma, analisis feature importance, serta implementasi sistem rekomendasi melalui dashboard berbasis Streamlit.

E. BAB V KESIMPULAN DAN SARAN

Bab ini menyimpulkan hasil penelitian terkait pemilihan channel follow-up yang optimal, serta memberikan saran untuk pengembangan penelitian dan implementasi lebih lanjut di PT XYZ.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA