

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian terdahulu memberikan landasan dan referensi yang relevan dalam memahami penggunaan *machine learning* untuk klasifikasi rekomendasi kunjungan pelanggan (*customer visit recommendation*), analisis perilaku pelanggan berdasarkan data historis, serta penerapan *dashboard* analisis untuk mendukung pengambilan keputusan di bidang penjualan. Penelitian-penelitian ini menjadi acuan dalam perancangan metodologi, pemilihan algoritma, dan pengembangan sistem pada penelitian ini. Berikut adalah rangkuman penelitian yang menjadi referensi utama:

Tabel 2. 1 Penelitian Terkait

| No | Penelitian Terkait (Judul, Penulis, Jurnal/Tahun) | Metode | Hasil | Gap Analisis |
|----|--|--|--|---|
| 1 | The relevance of lead prioritization: a B2B lead scoring model based on machine learning González-Flores et al. Frontiers in Artificial Intelligence, 2025 [4] | Evaluasi 15 algoritma klasifikasi pada data CRM B2B; fokus pada boosting (GBM, XGBoost, LightGBM); pembagian data 70/30; metrik evaluasi Accuracy, Precision, Recall, F1-score, ROC-AUC. | Gradient Boosting terbaik dengan Accuracy 0,9839; ROC-AUC 0,9891; Recall 0,9586; Precision 0,9106; F1-score 0,9338. | Penelitian ini berfokus pada konteks B2B software dengan pendekatan klasifikasi single-stage tanpa tahap rekomendasi kanal follow-up; tidak memanfaatkan metrik berbasis kapasitas operasional seperti Recall@K; model belum dievaluasi secara menyeluruh terhadap dampak bisnis karena baru diintegrasikan ke CRM. |
| 2 | Enhancing Conversions and Lead Scoring in Online Professional Education Yim et al. International Journal of Management, Finance and Accounting, 2024 [9] | Perbandingan algoritma LR, KNN, SVM, NB, RF, Bagging, Boosting, Stacking; seleksi fitur dengan RFE; validasi 10-fold cross-validation; penentuan threshold lead scoring. | Stacking terbaik: Accuracy 0,9233; Precision 0,9391; F1-score 0,8939. Model lead scoring berbasis LR: Accuracy 0,9019; threshold optimal 0,20. | Penelitian ini berfokus pada prediksi konversi lead di konteks pendidikan daring dengan pendekatan single-stage tanpa rekomendasi kanal follow-up; tidak menangani ketidakseimbangan kelas secara eksplisit; menggunakan hyperparameter default; terdapat potensi fitur penting yang terlewat; |

| No | Penelitian Terkait (Judul, Penulis, Jurnal/Tahun) | Metode | Hasil | Gap Analisis |
|----|--|---|---|--|
| | | | | tidak terdapat implementasi dashboard operasional maupun metrik berbasis kapasitas seperti Recall@K. |
| 3 | Predictive Modeling of Customer Response to Marketing Campaigns El-Hajj & Pavlova Electronics (MDPI), 2024 [17] | Pendekatan Decision Tree; penanganan imbalance melalui resampling; metrik evaluasi Accuracy, Precision, Recall, F1-score. | Sebelum resampling: Accuracy 0,873; Recall 0,43; F1-score 0,49. Setelah resampling: Accuracy 0,746; Recall 0,831; F1-score 0,742. | Penelitian ini menggunakan pendekatan single-stage Decision Tree pada konteks direct marketing; undersampling mengurangi ukuran data dan berpotensi menghilangkan informasi penting; tidak menyertakan rekomendasi kanal follow-up; generalisasi terbatas pada domain tertentu; penulis menyarankan penggunaan ensemble seperti Random Forest; tidak memanfaatkan metrik kapasitas operasional maupun dashboard. |
| 4 | Data-Driven Decision-Making for Bank Target Marketing Using Supervised Learning Classifiers on Imbalanced Big Data Nasir et al. Computers, Materials & Continua, 2024 [18] | Benchmark classifier terawasi (termasuk XGBoost, RF); oversampling (SMOTE, BorderlineSMOTE2); fokus pada F1-score. | XGBoost unggul pada F1-score; RF pada Precision dan AUC; BorderlineSMOTE2 meningkatkan performa secara keseluruhan. | Penelitian ini berfokus pada klasifikasi target marketing perbankan dengan pendekatan single-stage; tidak menyertakan tahap kedua maupun rekomendasi kanal follow-up; terdapat keterbatasan akses dataset perbankan; penulis menyarankan dataset lebih besar serta model baru berbasis optimisasi dan sampling; tidak memanfaatkan metrik kapasitas operasional maupun dashboard. |
| 5 | The state of lead scoring models and their impact on sales performance Wu et al. Information Technology and Management (Springer), 2023/2024 [5] | Systematic literature review terhadap 44 studi; pengelompokan model tradisional vs prediktif; pemetaan 14 metrik dampak terhadap kinerja penjualan. | Tren kuat menuju predictive lead scoring berbasis machine learning; teridentifikasi kekurangan evaluasi eksperimental dan metrik berorientasi hasil bisnis. | Sebagai tinjauan literatur, penelitian ini tidak menyajikan implementasi empiris; tidak menyediakan kerangka two-stage maupun rekomendasi kanal follow-up; penulis menekankan perlunya penggantian model tradisional dengan prediktif serta sistematisasi metode |

| No | Penelitian Terkait (Judul, Penulis, Jurnal/Tahun) | Metode | Hasil | Gap Analisis |
|----|---|---|--|--|
| | | | | supervised/unsupervised berdasarkan integritas data; tidak membahas implementasi dashboard operasional. |
| 6 | Enhancing Bank Customer Churn Prediction from an Imbalanced Dataset Using Machine Learning and SMOTE-Tomek Resampling Chakour et al. Lecture Notes in Networks and Systems (Springer), 2025 [19] | Perbandingan LR, DT, RF, KNN, XGBoost, SVM; penyeimbangan dengan SMOTE-Tomek; tuning Optuna; metrik Accuracy, Precision, Recall, F1-score. | RF dan XGBoost terbaik: Accuracy 0,86; F1-score 0,67 setelah resampling. | Penelitian ini berfokus pada prediksi churn nasabah perbankan dengan pendekatan single-stage; tidak menyertakan rekomendasi kanal follow-up; penulis menyarankan integrasi deep learning untuk meningkatkan akurasi di masa depan; tidak memanfaatkan metrik kapasitas operasional maupun dashboard. |
| 7 | A Data-Driven Approach for Direct Marketing of Banking Products Hematyar Knowledge and Data in Perspective (KDIP), 2023 [8] | Pendekatan two-stage: klasifikasi (RF, NB, NN) diikuti clustering K-means; pembagian data 70/30; analisis profit/cost. | NN mencapai accuracy 0,80; clustering 10 kelompok memberikan rasio Profit/Cost tertinggi 4,30. | Tahap kedua menggunakan clustering unsupervised sehingga rekomendasi kanal bersifat kelompok, bukan prediksi individu; tidak menangani ketidakseimbangan kelas secara eksplisit; tidak terdapat implementasi dashboard operasional; penulis menyarankan validasi pada industri lain serta eksplorasi algoritma klasifikasi tambahan. |
| 8 | The Ensemble Supervised Machine Learning for Credit Scoring Model in Digital Banking Institution Prahastiwi et al. Indonesian Journal of Artificial Intelligence and Data Mining, Vol. 8 No. 2, 2025 [20] | Perbandingan ensemble (RF, AdaBoost, LightGBM, CatBoost, XGBoost) terhadap Decision Tree; interpretabilitas SHAP; metrik ROC-AUC dan lainnya. | XGBoost terbaik dengan ROC-AUC 0,803; potensi penurunan NPL dan peningkatan approval rate. | Penelitian ini berfokus pada credit scoring untuk keputusan persetujuan kredit dengan pendekatan single-stage; tidak menyertakan rekomendasi kanal follow-up; tidak memanfaatkan metrik berbasis kapasitas operasional seperti Recall@K. |
| 9 | Cyber Security Data Science: Machine Learning Methods and Their Performance on Imbalanced Datasets Lopez-Ledezma & Velarde Digital Management and | Evaluasi single classifier (RF, LGBM, XGB dll.); teknik sampling (over/under, SMOTE, SPE); validasi 5-fold optimasi F1-score. | XGBoost dan RF robust; SMOTE sering menurunkan F1-score akibat noise; efek teknik imbalance bersifat campuran. | Penelitian ini berfokus pada deteksi fraud di cybersecurity dengan pendekatan single-stage; tidak menyertakan tahap kedua maupun rekomendasi kanal follow-up; penulis menekankan perlunya pengujian ulang |

| No | Penelitian Terkait (Judul, Penulis, Jurnal/Tahun) | Metode | Hasil | Gap Analisis |
|----|--|---|--|---|
| | Artificial Intelligence, ISPC 2024 [21] | | | teknik imbalance pada setiap dataset baru; tidak memanfaatkan metrik kapasitas operasional maupun dashboard. |
| 10 | Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers Chang et al. Risks (MDPI), 2024 [22] | Model NN, LR, AdaBoost, XGBoost, LightGBM; penanganan imbalance dengan SMOTE; pembagian data 70/30; metrik Accuracy, Precision dll. | XGBoost terbaik dengan Accuracy 0,994. | Penelitian ini berfokus pada prediksi risiko kredit dengan pendekatan single-stage; tidak menyertakan rekomendasi kanal follow-up; penulis menyatakan perlunya dataset lebih baru, beragam, dan lebih besar, serta integrasi faktor makroekonomi, validasi K-fold, dan adaptive learning; tidak memanfaatkan metrik kapasitas operasional maupun dashboard. |

Penelitian yang berjudul "The relevance of lead prioritization: a B2B lead scoring model based on machine learning" yang dipublikasikan di Frontiers in Artificial Intelligence mengembangkan model lead scoring pada konteks B2B CRM dengan mengevaluasi lima belas algoritma klasifikasi [4]. Hasil penelitian menunjukkan bahwa keluarga boosting, khususnya Gradient Boosting, memberikan kinerja paling optimal dengan nilai akurasi 0,9839 dan ROC-AUC 0,9891, serta recall tinggi sebesar 0,9586 dan F1-score 0,9338, melampaui model klasik seperti Logistic Regression. Sejalan dengan itu, studi "Enhancing Conversions and Lead Scoring in Online Professional Education" yang diterbitkan di International Journal of Management, Finance and Accounting menerapkan berbagai algoritma pembelajaran terawasi seperti Logistic Regression, KNN, SVM, Naïve Bayes, Random Forest, Bagging, Boosting, dan Stacking dalam konteks pendidikan profesional daring [9]. Dengan menggunakan Recursive Feature Elimination (RFE) dan 10-fold cross-validation, model Stacking menunjukkan kinerja terbaik dengan Accuracy 0,9233, Precision 0,9391, dan F1-score 0,8939, sementara Logistic Regression tetap kuat untuk lead scoring dengan akurasi dan recall di atas 0,90. Temuan ini mendukung pemilihan algoritma ensemble seperti

XGBoost pada penelitian ini untuk prioritas lead di PT XYZ, di mana fokus pada metrik seperti recall tinggi relevan untuk mengidentifikasi pelanggan potensial refinancing dalam data imbalanced, serta adaptasi threshold untuk membatasi lead hingga 80% kapasitas sales agar selaras dengan keterbatasan operasional.

Penelitian "Predictive Modeling of Customer Response to Marketing Campaigns" yang dipublikasikan di Electronics (MDPI) meneliti predictive modeling terhadap respons pelanggan pada kampanye pemasaran menggunakan algoritma Decision Tree [17]. Tanpa penyeimbangan kelas, model hanya menghasilkan Accuracy 0,873, Precision 0,56, dan Recall 0,43 ($F1 = 0,49$), tetapi setelah penerapan teknik resampling, performanya meningkat signifikan menjadi Accuracy 0,746, Precision 0,67, Recall 0,831, dan F1-score 0,742. Hasil ini menegaskan pentingnya teknik resampling dalam meningkatkan deteksi pelanggan potensial. Temuan serupa juga diperoleh dari studi "Data-Driven Decision-Making for Bank Target Marketing Using Supervised Learning Classifiers on Imbalanced Big Data" yang diterbitkan di Computers, Materials & Continua yang menguji berbagai Supervised Learning Classifiers pada data pemasaran bank yang sangat tidak seimbang [18]. Kombinasi XGBoost dengan Borderline SMOTE2 memberikan hasil terbaik dengan peningkatan F1-score dari 0,50 menjadi 0,87 serta peningkatan AUC dari 0,91 menjadi 0,94, menunjukkan efektivitas kombinasi ensemble boosting dan oversampling pada data keuangan. Pendekatan ini terkait dengan penelitian ini yang menerapkan cost-sensitive learning pada XGBoost untuk menangani imbalance di dataset operasional PT XYZ (lebih dari 500.000 record dengan rasio booking ~10%), menghindari oversampling untuk efisiensi komputasi sambil meningkatkan metrik bisnis seperti Recall@K yang selaras dengan kapasitas eksekusi tim sales.

Penelitian "The state of lead scoring models and their impact on sales performance" yang dipublikasikan di Information Technology and Management (Springer) melakukan systematic literature review terhadap 44 studi lead scoring dan mengelompokkan pendekatannya menjadi dua kategori, yaitu berbasis aturan dan berbasis machine learning [5]. Model prediktif terbukti lebih efektif dalam

meningkatkan efisiensi proses penjualan dan kolaborasi antara tim pemasaran dan penjualan, meskipun sebagian besar studi belum menghubungkan hasil model dengan indikator bisnis nyata. Dalam konteks penanganan data tidak seimbang, studi "Enhancing Bank Customer Churn Prediction from an Imbalanced Dataset Using Machine Learning and SMOTE-Tomek Resampling" yang diterbitkan di Lecture Notes in Networks and Systems (Springer), COCIA 2025 mengombinasikan teknik SMOTE-Tomek dan optimisasi hiperparameter Optuna pada enam algoritma populer [19]. Hasil penelitian menunjukkan bahwa Random Forest dan XGBoost menghasilkan performa terbaik dengan Accuracy 86%, Recall 0,70, dan F1-score 0,67, memperlihatkan bahwa pendekatan ensemble boosting dan hybrid resampling efektif dalam menangani data tidak seimbang. Temuan ini memperkuat penggunaan XGBoost di penelitian ini untuk prediksi churn-like pada data multifinance PT XYZ, di mana pembobotan kelas membantu menyeimbangkan metrik tanpa resampling agresif, sehingga mendukung optimasi kinerja sales melalui dashboard Streamlit yang mudah diakses.

Penelitian "A Data-Driven Approach for Direct Marketing of Banking Products" yang dipublikasikan di Knowledge and Data in Perspective (KDIP) mengusulkan model dua tahap dalam meningkatkan efektivitas direct marketing perbankan [8]. Tahap pertama melakukan klasifikasi pelanggan menggunakan Random Forest, Naïve Bayes, dan Neural Network, sedangkan tahap kedua melakukan clustering pelanggan positif menggunakan K-Means untuk menentukan strategi kanal distribusi yang tepat. Model Neural Network memberikan akurasi tertinggi sebesar 0,80 dengan profit/cost ratio 4,30 dan silhouette coefficient 0,88. Sementara itu, studi "The Ensemble Supervised Machine Learning for Credit Scoring Model in Digital Banking Institution" yang diterbitkan di Indonesian Journal of Artificial Intelligence and Data Mining (IJAIMD) mengembangkan model credit scoring berbasis ensemble learning pada lembaga perbankan digital di Indonesia menggunakan Optuna untuk penyetelan hiperparameter [20]. XGBoost menghasilkan ROC-AUC 0,803, Precision 0,512, Recall 0,683, dan F1-score 0,586, serta mengidentifikasi tenor, plafon, pendapatan, dan riwayat tunggakan sebagai fitur paling berpengaruh. Pendekatan two-stage ini langsung menginspirasi

arsitektur model di penelitian ini, di mana tahap pertama memprediksi probabilitas booking dan tahap kedua merekomendasikan follow-up (phone/visit) pada lead prioritas, dengan adaptasi metrik profit/cost untuk mengukur efisiensi operasional di konteks multifinance PT XYZ.

Penelitian “Cyber Security Data Science: Machine Learning Methods and Their Performance on Imbalanced Datasets” yang dipublikasikan pada prosiding ISPC 2024 menganalisis berbagai algoritma machine learning dan teknik penanganan ketidakseimbangan kelas pada dua dataset fraud finansial berskala besar, yaitu Credit Card (283.726 transaksi) dan PaySim (6.362.620 transaksi) dengan rasio ketidakseimbangan ekstrem antara kelas fraud dan non-fraud [21]. Studi ini membandingkan Random Forest, LightGBM, XGBoost, Logistic Regression, Decision Tree, dan Gradient Boosting pada beberapa skenario, yakni tanpa teknik sampling, oversampling, undersampling, SMOTE, dan Self-Paced Ensembling. Hasilnya menunjukkan bahwa XGBoost dan Random Forest sudah cukup robust pada kedua dataset, sementara teknik sampling tidak selalu meningkatkan kinerja; oversampling dan SMOTE cenderung hanya menaikkan recall dengan mengorbankan precision dan dalam banyak kasus justru menurunkan nilai F1 karena potensi penambahan noise dari sampel sintetis [21]. Di sisi lain, studi “Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers” yang diterbitkan di Risks (MDPI) membandingkan enam algoritma machine learning dan deep learning untuk memprediksi risiko gagal bayar kartu kredit dengan menerapkan teknik SMOTE pada data latih [22]. Penelitian tersebut melaporkan bahwa XGBoost memberikan kinerja tertinggi dengan akurasi 99,4%, diikuti oleh LightGBM dan Deep Neural Network, sekaligus menegaskan peran variabel usia, pendapatan, lama bekerja, dan jumlah tanggungan sebagai prediktor utama risiko kredit [22]. Temuan ini relevan dengan dataset besar PT XYZ di penelitian ini, di mana kami memilih cost-sensitive learning daripada SMOTE untuk menghindari noise, sehingga model XGBoost lebih efisien dalam memprediksi lead prioritas dan rekomendasi follow-up tanpa mengorbankan performa bisnis.

Secara keseluruhan, penelitian-penelitian terdahulu menunjukkan konsistensi bahwa keluarga ensemble boosting seperti Gradient Boosting, XGBoost, dan LightGBM serta Random Forest cenderung memberikan kinerja lebih unggul dibandingkan model klasik seperti Logistic Regression atau Naïve Bayes pada berbagai kasus pemasaran dan keuangan yang tidak seimbang. Pada beberapa studi dengan skala data kecil hingga menengah, teknik resampling seperti SMOTE, Borderline-SMOTE, dan SMOTE-Tomek dilaporkan mampu meningkatkan recall dan F1-score kelas minoritas. Namun, temuan Lopez-Ledezma dan Velarde (2024) pada dua dataset fraud finansial berskala besar menunjukkan bahwa teknik imbalance learning tidak selalu lebih baik daripada model dasar; SMOTE dan oversampling dapat menurunkan F1 karena menambah noise sintetis dan memperbesar beban komputasi [21]. Dengan mempertimbangkan karakteristik data penelitian ini yang berupa dataset finansial operasional berskala besar dan sangat tidak seimbang (lebih dari setengah juta observasi), pendekatan yang lebih sesuai adalah cost-sensitive learning melalui pembobotan kelas, misalnya dengan parameter `scale_pos_weight` pada XGBoost, serta penentuan threshold berbasis kapasitas bisnis, daripada menerapkan oversampling agresif seperti SMOTE yang berpotensi memicu overfitting dan inefisiensi komputasi.

Berdasarkan tinjauan tersebut, dapat diidentifikasi beberapa kesenjangan penelitian (research gap). Pertama, konteks domain multifinance, khususnya pada kasus refinancing pelanggan yang pernah memiliki kontrak sebelumnya, masih jarang diteliti dibandingkan sektor perbankan umum atau pendidikan. Kedua, arsitektur dua tahap yang terintegrasi—yang menghubungkan prediksi keberhasilan booking (Stage-1) dengan rekomendasi channel follow-up (Stage-2)—belum menjadi praktik umum; sebagian besar studi hanya berhenti pada klasifikasi prospek tanpa keputusan kanal yang bersifat kondisional. Ketiga, metrik berbasis kapasitas operasional seperti Recall@K, Precision@K, dan Lift@K belum digunakan secara luas, padahal metrik tersebut sangat relevan untuk organisasi dengan keterbatasan sumber daya, misalnya hanya 80% prospek yang dapat dieksekusi oleh tim lapangan. Keempat, kebijakan prioritisasi dan penentuan ambang keputusan (threshold) yang eksplisit sesuai kapasitas eksekusi dan biaya

salah klasifikasi jarang diformalkan. Kelima, penelitian terdahulu umumnya hanya menyoroti siapa yang harus ditindaklanjuti tanpa memberikan rekomendasi kanal tindak lanjut yang actionable seperti phone atau visit. Selain itu, hubungan antara metrik teknis dan outcome bisnis seperti capture rate, efisiensi jam kerja, atau conversion uplift belum terukur secara sistematis. Terakhir, implementasi sistem prediktif end-to-end yang ringan, mudah digunakan oleh pengguna non-teknis, serta dilengkapi dengan interpretabilitas model, misalnya Streamlit masih jarang ditemukan dalam literatur.

Dengan demikian, penelitian ini berupaya menjembatani kesenjangan tersebut melalui pengembangan model prediksi dua tahap (two-stage prediction model) yang menggabungkan prediksi probabilitas keberhasilan booking dan rekomendasi channel follow-up optimal pada konteks industri multifinance, dengan mempertimbangkan batas kapasitas eksekusi dan metrik bisnis yang relevan.

2.2 Tinjauan Teori

2.2.1 Sistem Rekomendasi Kunjungan Pelanggan

Sistem rekomendasi kunjungan pelanggan didefinisikan sebagai pendekatan sistematis yang dirancang untuk mengoptimalkan proses analisis serta penentuan prioritas kunjungan berdasarkan data historis dan kriteria tertentu. Dalam konteks predictive sales analytics, sistem rekomendasi dijelaskan sebagai aplikasi teknologi yang mengintegrasikan algoritma machine learning untuk diidentifikasi pola-pola tersembunyi seperti nilai pelanggan dan probabilitas churn, sehingga pengambilan keputusan penjualan dapat didukung secara lebih objektif [17]. Efisiensi operasional dikatakan dapat ditingkatkan melalui pemaksimalan peluang konversi yang diperoleh dari alokasi sumber daya yang lebih terarah dan berbasis data.

Dalam konteks industri pembiayaan konsumen, berbagai variabel kompleks seperti riwayat pembayaran, nilai kontrak, dan pola interaksi pelanggan dijelaskan perlu untuk diakomodasi oleh sistem AI-CRM sebagaimana disebutkan dalam penelitian sebelumnya [18]. Variabel-variabel tersebut umumnya diperoleh dari sistem Customer Relationship Management (CRM) yang digunakan untuk

merekam seluruh interaksi pelanggan serta aktivitas follow-up. Implementasi sistem serupa telah dilaporkan memberikan dampak signifikan terhadap kinerja organisasi melalui peningkatan kapabilitas CRM [19]. Selain itu, studi lain menunjukkan bahwa penerapan machine learning mampu mencapai akurasi hingga 93.4% dalam memprediksi perilaku konsumen, di mana algoritma XGBoost dilaporkan memperoleh F1-score sebesar 0.92 [20].

Dalam konteks industri pembiayaan konsumen yang kompleks, pengembangan sistem rekomendasi dijelaskan dapat dilakukan menggunakan pendekatan bertahap (staged approach) untuk meningkatkan akurasi serta efektivitas pengambilan keputusan. Melalui pendekatan two-stage, proses prediksi dipisahkan antara estimasi probabilitas keberhasilan untuk keperluan prioritisasi dan prediksi channel follow-up yang paling optimal, sehingga alokasi sumber daya oleh tim sales dapat dilakukan berdasarkan prioritas yang lebih komprehensif. Efektivitas pendekatan bertahap tersebut telah dilaporkan dalam penelitian sebelumnya, di mana penerapan XGBoost dengan Bayesian optimization terbukti menghasilkan model yang lebih berorientasi pada profit serta efisien secara komputasional [21]. Konsistensi temuan tersebut diperkuat oleh studi lain yang menunjukkan bahwa penggunaan gradient boosting dengan optimization memberikan performa yang lebih superior dibandingkan model single-stage, sebagaimana divalidasi melalui Friedman test [22]. Dengan demikian, pendekatan staged prediction dinyatakan mampu memungkinkan setiap tahap untuk mempelajari pola yang berbeda dan spesifik, sehingga akurasi prediksi akhir dapat ditingkatkan serta fleksibilitas dalam pengambilan keputusan bisnis dapat diperoleh.

2.2.2 Customer Relationship Management (CRM)

Customer Relationship Management (CRM) dipahami sebagai strategi bisnis komprehensif yang digunakan untuk mengintegrasikan proses, teknologi, dan sumber daya manusia dalam pengelolaan serta analisis interaksi pelanggan sepanjang siklus hidupnya. Pendekatan strategis tersebut dijelaskan sebagai integrasi antara teknologi, proses, dan peran manusia untuk mengelola serta

mengoptimalkan interaksi pelanggan dengan tujuan meningkatkan kepuasan, loyalitas, dan nilai pelanggan jangka panjang sebagaimana dinyatakan dalam penelitian sebelumnya [23]. Dalam era digital, CRM juga digambarkan telah berevolusi dari sekadar sistem pencatatan menjadi platform analitik prediktif yang memanfaatkan big data, kecerdasan buatan, dan predictive analytics. Transformasi ini dilaporkan memungkinkan perilaku pelanggan untuk diantisipasi secara lebih akurat serta mendukung pengambilan keputusan yang berorientasi pada data.

Sistem CRM modern dijelaskan telah mengintegrasikan berbagai touchpoint pelanggan—meliputi data transaksi, interaksi layanan, aktivitas pemasaran, serta umpan balik pelanggan—ke dalam satu platform terpadu. Dalam penelitian sebelumnya, tiga dimensi utama dari AI-powered CRM telah diidentifikasi, yaitu data management, multi-channel integration, dan tailored service offerings, yang masing-masing mendukung integrasi kecerdasan buatan secara lebih terstruktur dan mendalam dalam praktik CRM [24]. Studi komprehensif tersebut, yang mencakup analisis terhadap 64 artikel akademik serta wawancara dengan 24 ekspert CRM, melaporkan bahwa meskipun AI memiliki potensi besar dalam meningkatkan kapabilitas CRM, berbagai organisasi masih menghadapi kendala implementasi dengan tingkat kegagalan mencapai 18–69% akibat ketidaksesuaian antara infrastruktur organisasi, integrasi teknologi, dan data governance.

Sejalan dengan temuan tersebut, kerangka sistematis untuk integrasi AI-CRM yang sukses juga telah dikembangkan, dengan penekanan pada pentingnya penerapan ethics by design, sentralisasi data pelanggan, proses retraining model secara berkelanjutan, dan keterlibatan pengguna secara kontinu agar kapabilitas AI tetap selaras dengan karakteristik CRM serta mampu memenuhi kebutuhan bisnis [18].

2.2.3 Machine Learning

Machine Learning (ML) dipahami sebagai cabang dari Artificial Intelligence (AI) yang digunakan untuk mengembangkan algoritma yang memungkinkan sistem komputer belajar dari data serta meningkatkan kinerjanya tanpa memerlukan

pemrograman eksplisit. ML didefinisikan sebagai pendekatan komputasional yang memungkinkan mesin mengekstraksi pola dari data dan melakukan prediksi secara otomatis berdasarkan pengalaman sebelumnya, sebagaimana dijelaskan dalam penelitian sebelumnya [25].

Selain itu, algoritma ML juga telah diklasifikasikan ke dalam tiga kategori utama, yaitu supervised learning, unsupervised learning, dan reinforcement learning, yang masing-masing memiliki tujuan dan karakteristik pembelajaran yang berbeda, sebagaimana diuraikan dalam studi terkait pembelajaran mesin [26].

Dalam konteks bisnis, peran Machine Learning (ML) dijelaskan sebagai elemen penting dalam mendorong pengambilan keputusan berbasis data. Integrasi ML dan AI dilaporkan telah mengubah paradigma organisasi dari pendekatan intuitif menuju pengambilan keputusan yang lebih terukur melalui analisis prediktif, sebagaimana disebutkan dalam penelitian sebelumnya [27].

Tinjauan komprehensif terhadap lebih dari 100 publikasi juga telah dilakukan, menunjukkan bahwa penerapan ML dalam sektor bisnis dan keuangan—meliputi credit scoring, deteksi risiko, dan personalisasi strategi pemasaran—secara konsisten meningkatkan akurasi prediksi serta efisiensi proses pengambilan keputusan [28]. Studi tersebut turut menekankan bahwa teknik deep learning menunjukkan performa yang sangat kompetitif dalam konteks finansial, dengan prospek penelitian yang menjanjikan pada area seperti cryptocurrencies, financial crime detection, dan marketing analytics.

2.2.4 Supervised Learning

Supervised learning dipahami sebagai paradigma utama dalam machine learning yang paling banyak diterapkan di berbagai konteks bisnis. Pada pendekatan ini, proses pembelajaran dilakukan oleh algoritma melalui dataset berlabel untuk mempelajari fungsi pemetaan dari input ke output berdasarkan pasangan data yang tersedia. Supervised learning dijelaskan bertujuan untuk membangun model yang mampu mengaproksimasi fungsi target sehingga prediksi

terhadap data baru dapat dilakukan secara akurat, sebagaimana diuraikan dalam penelitian sebelumnya [25].

Selain itu, supervised learning juga telah dijelaskan memiliki dua fase utama, yaitu training phase yang digunakan untuk mempelajari pola dari data historis, serta testing phase yang digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data baru. Penegasan mengenai struktur dua fase ini disampaikan dalam studi yang membahas kerangka supervised learning [29].

Keberhasilan supervised learning dijelaskan sangat dipengaruhi oleh kualitas dan kuantitas data pelatihan yang digunakan, sebagaimana dinyatakan dalam penelitian sebelumnya [30]. Algoritma supervised learning secara umum telah diklasifikasikan ke dalam dua kategori berdasarkan tipe variabel target, yaitu: (1) regression untuk prediksi nilai kontinu, dan (2) classification untuk prediksi kategori diskrit. Pendekatan ini dilaporkan telah menjadi fondasi bagi berbagai teknik modern seperti ensemble methods dan gradient boosting, serta menjadi dasar bagi sebagian besar aplikasi machine learning di sektor bisnis dan keuangan.

Dalam domain customer analytics, aplikasi supervised learning juga telah dilaporkan menunjukkan hasil yang signifikan. Studi lain membuktikan bahwa model supervised learning menggunakan XGBoost dan CatBoost mampu mencapai nilai F1-score masing-masing sebesar 0.92 dan 0.93 dalam penelitian precision marketing berbasis UCI Online Shoppers Dataset, sehingga menegaskan efektivitas pendekatan berbasis data dalam memahami perilaku pelanggan [31].

2.2.5 Klasifikasi

Klasifikasi dipahami sebagai salah satu pendekatan utama dalam supervised learning yang digunakan untuk mengelompokkan data ke dalam kategori atau label tertentu berdasarkan pola yang dipelajari dari data historis. Klasifikasi dijelaskan berfungsi untuk memetakan sekumpulan fitur (predictor variables) terhadap variabel target yang bersifat kategorikal sehingga kelas dari suatu observasi baru dapat diidentifikasi dengan tingkat akurasi tertentu, sebagaimana dinyatakan dalam penelitian sebelumnya [30]. Berbeda dengan regresi yang menghasilkan prediksi

berupa nilai kontinu, klasifikasi menghasilkan keputusan diskrit seperti “ya/tidak” atau “tinggi/sedang/rendah”.

Dalam konteks bisnis, penggunaan klasifikasi dijelaskan banyak diterapkan untuk mendukung keputusan strategis seperti credit approval, fraud detection, dan customer segmentation. Model klasifikasi berbasis ensemble learning, seperti Random Forest dan XGBoost, dilaporkan memiliki kemampuan yang lebih unggul dalam mengenali pola kompleks pada perilaku pelanggan dibandingkan dengan model konvensional, sebagaimana ditunjukkan dalam penelitian sebelumnya [32]. Hasil penelitian tersebut memperlihatkan bahwa tuned XGBoost secara konsisten mencapai nilai F1-score tertinggi pada berbagai tingkat class imbalance dalam prediksi customer churn, dengan validasi statistik melalui Friedman test ($p < 0.05$). Temuan ini menyatakan bahwa performa XGBoost lebih superior dibandingkan Random Forest maupun metode tradisional lainnya.

Lebih lanjut, Logistic Regression juga dijelaskan tetap menjadi model dasar yang efektif untuk klasifikasi biner karena tingkat interpretabilitasnya yang tinggi, kemampuannya dalam menjelaskan pengaruh setiap variabel terhadap probabilitas suatu kejadian, serta performanya yang kompetitif pada dataset berskala besar, sebagaimana dinyatakan dalam penelitian sebelumnya [33]. Kombinasi antara model yang bersifat interpretable dan model berbasis boosting dilaporkan memungkinkan dilakukan perbandingan yang lebih komprehensif terkait performa serta kompleksitas algoritma yang digunakan.

2.2.6 Lead Scoring

Lead scoring merupakan metodologi yang digunakan untuk memberikan nilai atau skor kepada setiap lead berdasarkan atribut dan perilaku yang menunjukkan potensi konversi menjadi pelanggan. Lead scoring model diklasifikasikan oleh González-Flores et al. [34] menjadi dua kategori yaitu traditional lead scoring dan predictive lead scoring. Traditional lead scoring didasarkan pada pengalaman dan penilaian subjektif tim marketing dan sales dengan menggunakan metode berbasis aturan, poin, dan scorecard yang umumnya tidak memiliki validasi statistik formal. Sebaliknya, predictive lead scoring dikembangkan dengan memanfaatkan teknik

data mining dan machine learning untuk menganalisis data historis dan memprediksi probabilitas konversi secara lebih akurat.

Lead prioritization dianggap sebagai tugas kritis bagi organisasi karena memungkinkan alokasi sumber daya dilakukan secara lebih efektif, tenaga sales difokuskan pada peluang yang paling potensial, dan waktu yang dihabiskan untuk kualifikasi lead dapat dioptimalkan. Algoritma klasifikasi yang umum digunakan dalam predictive lead scoring meliputi decision tree, logistic regression, random forest, support vector machines (SVM), dan gradient boosting [34].

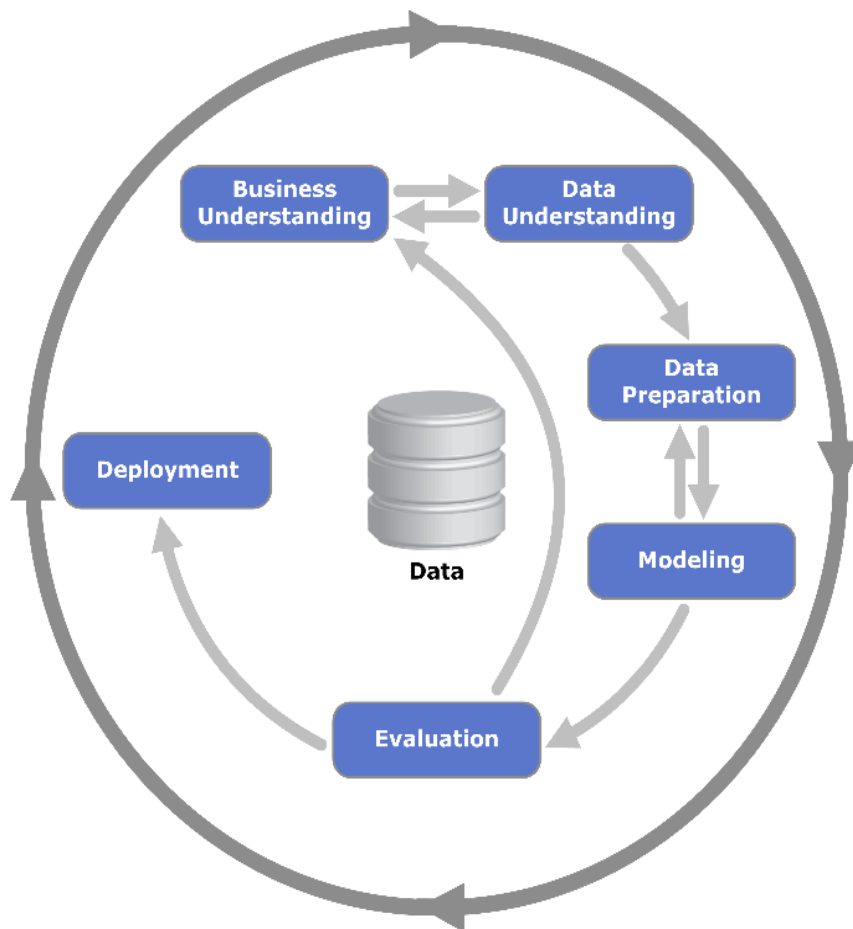
2.2.7 Strategi Follow-up dan Pemilihan Channel

Follow-up didefinisikan sebagai aktivitas tindak lanjut yang dilakukan oleh tim sales setelah kontak awal dengan lead untuk membangun hubungan dan mendorong konversi. Dalam konteks multichannel customer management, pemilihan channel yang tepat dianggap sebagai faktor krusial untuk keberhasilan follow-up. Berbagai faktor yang mempengaruhi pemilihan channel oleh pelanggan dalam lingkungan omnichannel diidentifikasi oleh Wolf dan Steul-Fischer [35] melalui systematic literature review, termasuk karakteristik channel, karakteristik pelanggan, dan tahapan customer journey.

Setiap channel memiliki karakteristik berbeda yang mempengaruhi efektivitasnya. Phone call memungkinkan interaksi langsung dan penanganan pertanyaan secara real-time dengan biaya relatif rendah, sedangkan visit atau kunjungan langsung memberikan pengalaman personal yang lebih mendalam dan memungkinkan demonstrasi produk dilakukan secara fisik namun dengan biaya operasional lebih tinggi. Pemilihan channel yang optimal ditentukan oleh karakteristik lead, kompleksitas produk, dan tahap dalam proses penjualan [35].

2.3 Framework dan Algoritma yang Digunakan

2.3.1 CRISP-DM



Gambar 2. 1 CRISP- DM [36]

Gambar 2.1 memperlihatkan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM) yang dipahami sebagai kerangka kerja standar industri dan telah menjadi *de facto* framework dalam pengembangan solusi analitik berbasis data. Dalam penelitian sebelumnya, CRISP-DM dilaporkan tetap menjadi metodologi yang paling banyak digunakan dan diakui secara luas untuk proyek artificial intelligence dan data mining [37]. Keunggulan metodologi ini dijelaskan terletak pada orientasi praktisnya—meliputi fokus eksplisit pada perumusan masalah bisnis, kemudahan implementasi, serta cakupan proses yang komprehensif—disertai tingkat agility dan fleksibilitas yang tinggi dalam penerapannya.

Secara umum, CRISP-DM terdiri atas enam fase utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment, yang saling terhubung dalam siklus iteratif untuk mendukung proses

analisis end-to-end mulai dari perumusan masalah bisnis hingga penerapan model pada lingkungan operasional. Keenam fase tersebut membentuk alur kerja yang memungkinkan dilakukannya umpan balik secara berkelanjutan sehingga hasil analisis dapat terus disesuaikan dengan kebutuhan bisnis yang dinamis.

Lebih lanjut, hasil studi longitudinal selama enam tahun melaporkan bahwa CRISP-DM dapat diperluas menjadi suatu life cycle yang berkelanjutan dan aktif untuk mendukung pengembangan solusi AI [37]. Pengembangan tersebut dijelaskan dilakukan melalui penambahan fase “Operation and Maintenance” serta integrasi kerangka kerja berbasis tugas yang menghubungkan kompetensi domain knowledge, data science, dan data engineering. Metodologi ini juga disebut memfasilitasi komunikasi yang lebih efektif antara tim teknis dan pemangku kepentingan bisnis melalui penggunaan bahasa serta terminologi yang konsisten, sehingga keselarasan antara tujuan bisnis dan solusi teknis yang dikembangkan dapat dipastikan.

2.3.1.1 Business Understanding

Fase Business Understanding merupakan tahap paling krusial dalam siklus CRISP-DM karena menentukan arah dan keberhasilan proyek analitik. Tahapan ini berfokus pada penyelarasan antara tujuan bisnis dan rancangan solusi analitik, dengan penekanan pada pemahaman mendalam terhadap permasalahan bisnis serta keterlibatan pemangku kepentingan sejak awal proses pengembangan.

Pada tahap ini, aktivitas utama meliputi analisis proses bisnis yang sedang berjalan, identifikasi hambatan operasional, serta penetapan metrik keberhasilan yang terukur sebagai dasar pengembangan solusi. Pemahaman yang baik terhadap kebutuhan dan tantangan bisnis menjadi fondasi penting agar model analitik yang dibangun relevan, aplikatif, dan selaras dengan tujuan organisasi.

2.3.1.2 Data Understanding

Fase Data Understanding dipahami sebagai tahap eksplorasi awal terhadap data yang bertujuan untuk memperoleh pemahaman mengenai karakteristik, kualitas, serta potensi wawasan yang dapat diekstrak dari dataset. Penelitian sebelumnya menekankan bahwa fase ini berperan penting dalam mengidentifikasi berbagai data quality issues serta menilai kesiapan data sebelum proses pemodelan dilakukan, khususnya dalam penerapan kerangka CRISP-DM pada lingkungan organisasi yang kompleks [38]. Selain itu, pemahaman mendalam mengenai data availability dan data quality juga dilaporkan sebagai salah satu tantangan utama dalam implementasi solusi berbasis AI, karena struktur data, tipe variabel, dan hubungan antarfitur memiliki pengaruh signifikan terhadap kinerja model prediktif yang dihasilkan [37].

Pada tahap ini, sejumlah aktivitas eksplorasi dilakukan untuk memperoleh pemahaman mengenai sumber data, konteks bisnis, dan karakteristik setiap variabel. Aktivitas umum dalam fase Data Understanding dijelaskan meliputi statistical profiling, visualisasi distribusi data, serta analisis korelasi untuk mengidentifikasi pola, tren, maupun anomali dalam dataset. Pemahaman terhadap keberadaan missing values, outliers, dan inkonsistensi data menjadi aspek krusial yang menentukan strategi preprocessing pada fase Data Preparation berikutnya. Selain itu, fase ini juga mencakup penilaian terhadap kecukupan serta relevansi data untuk memastikan bahwa informasi yang tersedia mampu mendukung tujuan pemodelan yang telah didefinisikan pada fase Business Understanding.

2.3.1.3 Data Preparation

Fase Data Preparation dipahami sebagai tahap yang paling memerlukan waktu dalam siklus CRISP-DM karena mencakup proses transformasi dan pembersihan data sebelum pemodelan dilakukan. Tahap ini dilaporkan berperan penting dalam memastikan kualitas, konsistensi, dan representativitas data yang akan digunakan untuk analisis, sebab kesalahan pada fase ini dapat secara langsung memengaruhi performa model prediktif [38]. Proses Data Preparation mencakup kegiatan data cleaning, data integration, dan feature engineering yang bertujuan

menghasilkan dataset yang bersih, konsisten, serta siap digunakan pada tahap Modeling. Aktivitas umum pada fase ini dijelaskan meliputi penanganan missing values melalui teknik imputasi, normalisasi variabel numerik, serta encoding variabel kategorikal menggunakan metode seperti one-hot encoding atau label encoding.

Selain itu, feature engineering dijelaskan berperan penting dalam menciptakan variabel turunan yang lebih informatif terhadap perilaku pelanggan, seperti recency of last transaction, frequency of interaction, dan monetary value indicators, yang berpotensi meningkatkan kemampuan prediksi model. Fase ini juga mencakup integrasi data dari berbagai sumber, seperti sistem Customer Relationship Management (CRM), histori transaksi, dan data operasional perusahaan, yang dinyatakan sebagai komponen krusial dalam pengembangan solusi AI berkelanjutan [37]. Dengan demikian, Data Preparation berfungsi sebagai penghubung antara eksplorasi data dan pembangunan model, memastikan bahwa dataset yang digunakan memiliki kualitas tinggi sehingga mampu mendukung proses analisis serta menghasilkan model yang akurat dan reliabel..

2.3.1.4 Modelling

Fase Modeling dipahami sebagai tahap yang melibatkan penerapan berbagai algoritma machine learning untuk membangun model prediktif berdasarkan pola dan hubungan antarvariabel dalam data. Tahap ini dijelaskan sebagai inti dari proses analitik yang memerlukan eksplorasi sistematis terhadap beragam algoritma guna mengidentifikasi pendekatan yang paling sesuai dengan karakteristik data serta tujuan bisnis [37]. Secara umum, algoritma klasifikasi seperti XGBoost, Random Forest, dan Logistic Regression banyak digunakan dalam pengembangan model prediksi pelanggan karena kemampuannya menyeimbangkan akurasi, stabilitas, dan interpretabilitas hasil. Kombinasi antara pendekatan berbasis ensemble dan model linier dilaporkan memungkinkan terciptanya model yang lebih robust sekaligus tetap mudah dijelaskan dari sisi keputusan analitiknya.

Proses Modeling juga mencakup pembagian data menjadi training, validation, dan testing sets dengan memperhatikan prinsip non-leakage agar tidak

terjadi kebocoran informasi antar subset data. Selanjutnya, penyetelan hiperparameter dilakukan menggunakan metode seperti grid search atau Bayesian optimization untuk memperoleh kinerja model yang optimal. Validasi model dilakukan melalui stratified k-fold cross-validation guna memastikan kemampuan generalisasi terhadap data baru serta menghindari risiko overfitting. Hasil yang diperoleh dari tahap ini menjadi dasar bagi fase Evaluation, yang berfokus pada pengukuran performa model secara kuantitatif menggunakan berbagai metrik evaluasi yang relevan.

2.3.1.5 Evaluation

Fase Evaluation dipahami sebagai tahap krusial untuk memastikan bahwa model yang dikembangkan tidak hanya unggul secara teknis, tetapi juga mampu memberikan nilai bisnis yang nyata bagi organisasi. Proses evaluasi dijelaskan perlu dilakukan secara komprehensif dengan mempertimbangkan metrik teknis, relevansi bisnis, serta kemampuan model untuk diintegrasikan dalam proses operasional, sehingga keselarasan antara performa model dan tujuan strategis perusahaan dapat dipastikan [37]. Evaluasi teknis umumnya dilakukan menggunakan metrik kuantitatif seperti accuracy, precision, recall, F1-score, dan AUC-ROC untuk menilai kemampuan model dalam melakukan klasifikasi secara objektif.

Selain aspek teknis, pentingnya evaluasi bisnis juga telah ditegaskan, khususnya dalam menilai sejauh mana model mampu meningkatkan efisiensi operasional dan memberikan dampak nyata terhadap kinerja organisasi, terutama pada sektor industri keuangan yang sensitif terhadap risiko [38]. Pendekatan seperti cost-sensitive learning sering digunakan untuk menyeimbangkan biaya kesalahan klasifikasi berdasarkan nilai ekonomis tiap pelanggan, sementara hasil evaluasi juga dapat dibandingkan dengan baseline methods maupun praktik operasional eksisting. Melalui pendekatan yang menyeluruh, fase Evaluation berfungsi sebagai tahap verifikasi akhir untuk memastikan bahwa model analitik tidak hanya unggul secara statistik, tetapi juga relevan serta siap diimplementasikan dalam lingkungan bisnis yang dinamis.

2.3.1.6 Deployment

Fase Deployment dipahami sebagai tahap akhir dalam siklus CRISP-DM yang berfokus pada penerapan model ke lingkungan produksi serta integrasinya dengan sistem bisnis yang telah ada. Tahap ini dilaporkan memerlukan kesiapan infrastruktur teknis, manajemen perubahan organisasi, dan mekanisme pemantauan berkelanjutan agar model dapat beroperasi secara stabil dalam konteks bisnis nyata [38]. Sejalan dengan temuan sebelumnya, aktivitas pada fase ini dijelaskan meliputi penerapan model drift detection, scheduled retraining, dan feedback loops untuk menjaga kinerja model terhadap dinamika data maupun kebutuhan bisnis yang terus berubah [39], [37].

Selain itu, dokumentasi yang komprehensif, pelatihan pengguna, serta sistem monitoring and maintenance juga disebut sebagai elemen penting untuk memastikan keberlanjutan serta realisasi nilai bisnis dari solusi analitik yang diterapkan. Dengan demikian, fase Deployment berfungsi sebagai tahap implementasi yang memastikan bahwa model tidak hanya berhasil dibangun secara teknis, tetapi juga memberikan dampak operasional yang konsisten dan terukur bagi organisasi.

2.3.2 Algoritma

2.3.2.1 XGBoost (Extreme Gradient Boosting)

XGBoost dipahami sebagai pengembangan dari kerangka gradient boosting yang dirancang untuk meningkatkan efisiensi, skalabilitas, dan akurasi model machine learning, khususnya pada data tabular. Algoritma ini dijelaskan bekerja dengan membangun serangkaian decision trees secara berurutan, di mana setiap pohon baru digunakan untuk memperbaiki kesalahan residual dari pohon sebelumnya melalui proses optimisasi iteratif. Sebagaimana dinyatakan dalam penelitian sebelumnya, XGBoost dibangun sebagai sistem scalable tree-boosting yang mengoptimalkan objective function melalui kombinasi fungsi loss dan regularisasi, serta dirancang untuk mendukung pemrosesan data berukuran besar dengan kemampuan paralelisasi yang tinggi [40].

Selain itu, studi komparatif juga melaporkan bahwa tuned XGBoost secara konsisten menghasilkan performa tertinggi dibandingkan model lain pada berbagai tingkat ketidakseimbangan data, sehingga menegaskan kekuatan algoritma ini dalam konteks praktis aplikasi bisnis [41].

Secara matematis, struktur model XGBoost sebagai ansambel pohon keputusan dapat direpresentasikan melalui formulasi aditif, di mana prediksi akhir dibentuk dari penjumlahan kontribusi setiap pohon yang dilatih secara bertahap. Representasi dasar dari model ansambel XGBoost tersebut ditunjukkan pada Formula 2.1.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F$$

Rumus 2. 1 Rumus Extreme Gradient Boost [42]

Keterangan:

- \hat{y} = Jumlah prediksi akhir dari setiap pohon
- \sum = Penjumlahan (kontribusi setiap pohon)
- k = Indeks pohon ke- k (adanya penambahan untuk setiap pohon)
- f_k = Fungsi prediksi hasil dari pohon ke- k (individual)
- F = Ruang fungsi (mengandung semua pohon yang sudah dilatih)
- x_i = Vektor fitur dari sampel data ke- i (data yang diprediksi)

Rumus 2.1 memperlihatkan bahwa prediksi untuk suatu sampel data x_i diperoleh dengan menjumlahkan keluaran dari seluruh fungsi pohon f_k yang berada dalam himpunan fungsi F . Melalui struktur aditif ini, setiap pohon digunakan untuk memperbaiki kesalahan prediksi tahap sebelumnya sehingga model secara keseluruhan mampu mencapai akurasi dan kemampuan generalisasi yang lebih baik.

Prinsip kerja XGBoost berfokus pada minimisasi *objective function* yang terdiri atas dua komponen utama: *loss function* untuk mengukur kesalahan prediksi dan *regularization term* untuk mengontrol kompleksitas model. Kombinasi

keduanya memungkinkan XGBoost mencapai keseimbangan antara akurasi dan kemampuan generalisasi. Selain itu, algoritma ini memiliki fitur *built-in regularization* (L1 dan L2), *parallel processing*, serta kemampuan menangani *missing values* secara otomatis. Keunggulan-keunggulan tersebut menjadikan XGBoost salah satu algoritma yang banyak digunakan dalam berbagai domain, seperti keuangan, pemasaran, dan sistem rekomendasi, terutama untuk permasalahan klasifikasi dan prediksi pada data berskala besar maupun tidak seimbang (*imbalanced data*).

2.3.2.2 Random Forest

Random Forest dipahami sebagai metode ensemble learning yang dikembangkan berdasarkan konsep bootstrap aggregating (bagging) serta random feature selection. Dalam algoritma ini, sejumlah decision trees dibangun dengan memanfaatkan sampel acak (bootstrap samples) dari data pelatihan, sementara prediksi dihasilkan melalui proses majority voting untuk kasus klasifikasi atau averaging untuk kasus regresi. Pada setiap pohon, data pelatihan yang berbeda dan subset acak fitur digunakan untuk mengurangi korelasi antar-pohon serta meningkatkan kemampuan generalisasi model, sebagaimana dijelaskan dalam penelitian sebelumnya [43].

Prinsip kerja Random Forest juga dikaitkan dengan konsep wisdom of crowds, di mana kombinasi dari banyak weak learners menghasilkan strong predictor yang lebih robust terhadap noise maupun outliers. Efektivitas metode ini dalam berbagai aplikasi machine learning diperkuat oleh temuan yang menunjukkan bahwa Random Forest tetap menjadi salah satu algoritma yang paling banyak digunakan karena kemampuannya menjaga keseimbangan antara akurasi prediktif, ketahanan terhadap gangguan data, serta kemudahan implementasi [44].

Secara matematis, proses agregasi prediksi pada Random Forest direpresentasikan sebagai rata-rata keluaran dari seluruh pohon keputusan yang membentuk ansambel. Formulasi umum dari mekanisme agregasi tersebut ditunjukkan pada rumus 2.2.

$$H(x) = \frac{1}{M} \sum_{m=1}^M h_m(x)$$

Rumus 2. 2 Rumus Random Forest [45]

Keterangan :

- $H(x)$ = Prediksi akhir dari model Random Forest
- M = Jumlah total pohon (decision trees) dalam ensemble
- \sum = Simbol penjumlahan dari semua pohon
- fk = Fungsi prediksi hasil dari pohon ke- k (individual)
- m = Indeks pohon ke- m (dari 1 sampai M)
- $h_m(x)$ = Fungsi prediksi dari pohon ke- m terhadap input x

Formula 2.2 memperlihatkan bahwa prediksi akhir $H(x)$ diperoleh dengan menghitung rata-rata prediksi dari setiap pohon keputusan $h_m(x)$ untuk $m=1, \dots, M$. Melalui proses perataan ini, variasi dan kesalahan individual pada masing-masing pohon dapat dikompensasi sehingga model menjadi lebih stabil dan robust terhadap noise maupun risiko overfitting.

Keunggulan Random Forest meliputi kemampuan menangani pola non-linear dan interaksi antarvariabel secara alami, ketahanan terhadap *outliers* dan *noise*, penyediaan ukuran *feature importance* melalui *mean decrease impurity* maupun *permutation importance*, kebutuhan penyetelan hiperparameter yang minimal, proses pelatihan yang dapat dijalankan secara paralel, serta adanya *out-of-bag error estimation* yang memberikan estimasi tak bias terhadap *generalization error*. Sementara itu, keterbatasan algoritma ini antara lain potensi *overfitting* pada data yang sangat bising, kesulitan dalam menangkap hubungan linear sederhana, serta rendahnya interpretabilitas untuk prediksi individu dibandingkan dengan *single decision tree*.

2.3.2.3 Logistic Regression

Logistic Regression dipahami sebagai metode statistik fundamental untuk klasifikasi biner maupun multikelas yang digunakan untuk memodelkan probabilitas keanggotaan suatu kelas melalui logistic function. Meskipun termasuk dalam kategori algoritma klasik, metode ini tetap dianggap sebagai salah satu teknik yang paling banyak digunakan dalam machine learning karena tingkat interpretabilitasnya yang tinggi, efisiensi komputasinya, serta dasar teorinya yang kuat. Estimasi parameter pada Logistic Regression dilakukan dengan memaksimalkan likelihood dari data teramati melalui prosedur optimisasi iteratif seperti gradient descent atau Newton–Raphson.

Penjelasan sebelumnya menyatakan bahwa Logistic Regression bekerja dengan memetakan kombinasi linier variabel prediktor ke dalam ruang log-odds sehingga probabilitas kelas yang dihasilkan menjadi stabil dan dapat diinterpretasikan secara langsung [46]. Selain itu, relevansi metode ini sebagai baseline classifier juga ditegaskan dalam penelitian lain, terutama karena kemampuannya menghasilkan prediksi probabilistik yang dapat dimanfaatkan untuk analisis risiko, penentuan threshold keputusan, serta interpretasi hubungan antarvariabel secara transparan [47]. Kombinasi antara interpretabilitas, konsistensi performa, dan fleksibilitas menjadikan Logistic Regression tetap diposisikan sebagai komponen penting dalam studi empiris maupun aplikasi praktis modern.

Logistic Regression memodelkan hubungan antara variabel prediktor dan probabilitas kejadian suatu kelas biner. Model ini menggunakan *logistic function* untuk mengubah kombinasi linier fitur menjadi probabilitas antara 0 dan 1. Secara matematis, rumus utama Logistic Regression didefinisikan sebagai berikut:

$$p(y = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}}$$

Rumus 2. 3 Rumus Logistic Regression [48]

Keterangan :

- $p(y = 1 | x)$: probabilitas bahwa sampel termasuk kelas 1 (kelas positif).

- e : bilangan eksponensial (≈ 2.71828).
- β_0 : intercept atau konstanta model.
- β_i : koefisien regresi untuk variabel prediktor ke-i.
- x_i : nilai variabel prediktor ke-i.
- k : jumlah total variabel prediktor.

Model Logistic Regression pada dasarnya memetakan hubungan antara variabel prediktor dan probabilitas terjadinya suatu kelas dengan melakukan transformasi logistik pada kombinasi linier dari fitur. Transformasi ini memastikan bahwa nilai output selalu berada pada rentang 0 hingga 1, sehingga dapat diinterpretasikan sebagai probabilitas. Nilai koefisien β_i menggambarkan besarnya pengaruh setiap variabel prediktor terhadap perubahan *log-odds* dari kelas positif. Apabila β_i bernilai positif, maka peningkatan nilai x_i akan meningkatkan probabilitas kejadian kelas 1; sebaliknya, jika β_i bernilai negatif, maka peningkatan x_i akan menurunkan probabilitas tersebut.

Dengan demikian, Logistic Regression tidak hanya menghasilkan prediksi klasifikasi, tetapi juga menyediakan interpretasi yang jelas terkait hubungan antara variabel bebas dan variabel dependen. Kemampuan ini menjadikan Logistic Regression banyak digunakan dalam pengambilan keputusan berbasis risiko dan analisis inferensial, terutama pada konteks yang memerlukan transparansi model dan pemahaman mengenai kontribusi setiap variabel prediktor.

Kelebihan Logistic Regression meliputi interpretabilitas yang tinggi, di mana arah dan besaran koefisien secara langsung menunjukkan pengaruh fitur terhadap target; kerangka probabilistik yang mendukung keputusan di bawah ketidakpastian; efisiensi komputasi; kebutuhan penyetelan hiperparameter yang minimal; serta teori statistik yang mapan untuk inferensi dan pengujian hipotesis. Dengan penerapan regularisasi (L1 dan L2), model ini juga dapat menjaga keseimbangan antara *bias* dan *variance* serta meningkatkan ketahanan terhadap *overfitting*. Adapun keterbatasan Logistic Regression mencakup asumsi linearitas pada ruang *log-odds*, ketidakmampuannya menangkap pola non-linear yang kompleks tanpa

feature engineering eksplisit, serta sensitivitas terhadap *multicollinearity* antar fitur.

2.4 Teori tentang Metrik Evaluasi Model

2.4.1 Confusion Matrix

Confusion Matrix atau matriks kebingungan dipahami sebagai tabel evaluasi yang digunakan untuk memetakan hasil prediksi model klasifikasi terhadap label aktual pada data uji. Untuk kasus klasifikasi biner, matriks ini dijelaskan berukuran 2×2 dan terdiri atas empat komponen utama, yaitu True Positive (TP) yang menunjukkan jumlah observasi yang benar diprediksi sebagai kelas positif, True Negative (TN) yang menunjukkan jumlah observasi yang benar diprediksi sebagai kelas negatif, False Positive (FP) atau Type I Error yang menunjukkan jumlah observasi yang salah diprediksi sebagai positif, serta False Negative (FN) atau Type II Error yang menunjukkan jumlah observasi yang salah diprediksi sebagai negatif [49].

Melalui pemahaman terhadap keempat komponen tersebut, berbagai metrik evaluasi seperti Accuracy, Precision, Recall (Sensitivity), dan F1-Score dapat dihitung untuk menilai kekuatan maupun kelemahan spesifik dari model klasifikasi. Lebih lanjut, penelitian tersebut menekankan bahwa interpretasi terhadap komponen luar-diagonal (FP dan FN) menjadi sangat krusial dalam konteks data tidak seimbang (imbalanced), karena penggunaan metrik seperti Accuracy saja dapat menghasilkan penilaian performa yang menyesatkan. Dengan demikian, Confusion Matrix diposisikan sebagai alat evaluasi dasar yang penting untuk mendukung analisis performa model secara komprehensif.

2.4.2 Accuracy

Akurasi dipahami sebagai salah satu metrik evaluasi yang paling intuitif dan paling banyak digunakan dalam pengukuran performa model klasifikasi. Metrik ini dihitung sebagai proporsi total prediksi yang benar—baik True Positive (TP) maupun True Negative (TN)—terhadap keseluruhan jumlah observasi pada dataset. Dalam penelitian sebelumnya dijelaskan bahwa akurasi kerap digunakan sebagai

ukuran dasar performa klasifikasi, namun interpretasinya perlu dilakukan secara berhati-hati karena nilainya dapat dipengaruhi oleh ketidakseimbangan kelas, prevalensi data, serta kualitas referensi yang digunakan [50]. Secara matematis, akurasi dapat dirumuskan sebagai berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP}$$

Rumus 2. 4 Rumus Accuracy [51]

Keterangan :

- $TP = \text{True Positive}$ (prediksi benar untuk kelas positif)
- $TN = \text{True Negative}$ (prediksi benar untuk kelas negatif)
- $FP = \text{False Positive}$ (prediksi salah untuk kelas positif)
- $FN = \text{False Negative}$ (prediksi salah untuk kelas negatif)

Pada rumus di atas, TP adalah jumlah prediksi benar untuk kelas positif, TN adalah jumlah prediksi benar untuk kelas negatif, FP adalah jumlah prediksi salah untuk kelas positif, dan FN adalah jumlah prediksi salah untuk kelas negatif. Nilai akurasi yang mendekati 1,0 menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan data secara keseluruhan.

Namun, dalam penelitian sebelumnya ditegaskan bahwa akurasi memiliki kelemahan signifikan ketika digunakan pada dataset yang sangat tidak seimbang [52]. Pada kondisi di mana distribusi kelas sangat timpang, sebuah model dapat menghasilkan nilai akurasi yang tinggi meskipun kegagalan dalam mengenali kelas minoritas tetap terjadi. Oleh karena itu, penggunaan akurasi perlu dilakukan dengan hati-hati dan sebaiknya dilengkapi dengan metrik lain seperti precision, recall, dan F1-score agar kinerja model dapat dievaluasi secara lebih komprehensif.

2.4.3 Precision

Precision, atau presisi, dipahami sebagai metrik yang mengukur proporsi prediksi positif yang benar dari seluruh observasi yang diklasifikasikan sebagai

positif. Dengan kata lain, precision digunakan untuk menjawab pertanyaan mengenai berapa banyak observasi yang benar-benar positif dari seluruh prediksi yang telah ditetapkan sebagai kelas positif. Dalam penelitian sebelumnya dijelaskan bahwa precision menjadi sangat relevan pada situasi di mana kesalahan berupa false positives (FP) dapat menimbulkan konsekuensi besar, seperti kerugian sumber daya, peningkatan biaya operasional, ataupun risiko terhadap reputasi organisasi [50].

Secara matematis, precision dirumuskan sebagai berikut:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Rumus 2. 5 Rumus Precision [51]

Keterangan :

- TP = *True Positive* (prediksi benar untuk kelas positif)
- FP = *False Positive* (prediksi salah untuk kelas positif)

Pada rumus di atas, *TP* adalah jumlah prediksi benar untuk kelas positif dan *FP* adalah jumlah prediksi salah untuk kelas positif. Nilai precision yang tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang keliru, sehingga hasil prediksinya dapat dipercaya. Dengan demikian, precision menekankan pada kualitas prediksi positif, bukan jumlah keseluruhan prediksi yang benar.

2.4.5 Recall

Recall, atau sensitivitas, dipahami sebagai metrik yang digunakan untuk mengukur kemampuan model dalam mengenali seluruh sampel yang sebenarnya termasuk ke dalam kelas positif. Dengan kata lain, recall digunakan untuk menjawab pertanyaan mengenai berapa banyak observasi yang benar-benar positif yang berhasil diidentifikasi dengan benar oleh model. Dalam penelitian sebelumnya dijelaskan bahwa recall menjadi sangat krusial pada situasi di mana kesalahan berupa False Negative (FN) menimbulkan risiko besar—baik dalam konteks bisnis,

kesehatan, maupun deteksi risiko—karena kegagalan dalam mengenali kasus positif dapat menghasilkan konsekuensi operasional yang signifikan [50]. Oleh karena itu, recall sering digunakan bersama precision untuk memberikan evaluasi yang lebih seimbang terhadap performa model, terutama pada kondisi ketidakseimbangan kelas.

Secara matematis, recall dapat dirumuskan sebagai:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Rumus 2. 6 Rumus Recall [50]

Keterangan :

- $TP = \text{True Positive}$ (prediksi benar untuk kelas positif)
- $FN = \text{False Negative}$ (prediksi salah untuk kelas negatif)

di mana TP adalah jumlah prediksi benar untuk kelas positif dan FN adalah jumlah prediksi salah untuk kelas negatif. Nilai recall yang tinggi menunjukkan bahwa model mampu menangkap sebagian besar sampel positif yang relevan. Dengan demikian, recall berfokus pada **kelengkapan deteksi (completeness)**, bukan ketepatan prediksi, sehingga ideal untuk digunakan ketika kegagalan mendeteksi kelas positif lebih berisiko dibanding memberikan prediksi positif yang keliru.

2.4.6 F-Measure

F-Measure atau F1-Score dipahami sebagai metrik evaluasi yang digunakan untuk menggabungkan Precision dan Recall guna memberikan penilaian yang lebih seimbang terhadap kinerja model klasifikasi. Metrik ini dihitung sebagai rata-rata harmonik dari kedua metrik tersebut, sehingga nilai F1 hanya akan menjadi tinggi apabila Precision dan Recall sama-sama berada pada tingkat yang baik. Dalam penelitian sebelumnya dijelaskan bahwa penggunaan rata-rata harmonik membuat F1-Score menjadi lebih sensitif terhadap nilai ekstrem serta lebih efektif pada kondisi imbalanced dataset, karena metrik ini mampu memberikan gambaran

performa model yang lebih stabil dibandingkan akurasi semata . Oleh karena itu, F1-Score banyak digunakan dalam aplikasi prediktif yang menekankan pentingnya keseimbangan antara kemampuan mendeteksi kasus positif dan ketepatan prediksi positif.

Secara matematis, F1-Score dirumuskan sebagai:

$$F\text{-Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Rumus 2. 7 Rumus Precision [53]

Pada rumus di atas, *precision* menyatakan proporsi prediksi positif yang benar terhadap seluruh prediksi positif yang dihasilkan model, sedangkan *recall* menyatakan proporsi kasus positif yang berhasil dideteksi model dari seluruh kasus positif yang ada. F1-Score merepresentasikan rata-rata harmonik antara *precision* dan *recall*, sehingga nilai F1-Score yang tinggi menunjukkan bahwa model mampu menjaga keseimbangan antara ketepatan dan kelengkapan dalam mengidentifikasi kelas positif.

2.4.7 Area Under Curve (AUC)

Area Under the Curve (AUC) dipahami sebagai metrik evaluasi yang digunakan untuk mengukur kemampuan model klasifikasi dalam membedakan antara kelas positif dan kelas negatif. AUC dihitung sebagai luas area di bawah kurva Receiver Operating Characteristic (ROC), yaitu grafik yang menampilkan hubungan antara True Positive Rate (TPR atau Recall) pada sumbu Y dan False Positive Rate (FPR) pada sumbu X. Dalam penelitian sebelumnya dijelaskan bahwa ROC–AUC memberikan ukuran yang stabil terhadap performa model pada berbagai nilai threshold, sehingga metrik ini sangat berguna dalam skenario klasifikasi biner, khususnya ketika distribusi kelas tidak seimbang [54]. Nilai AUC yang lebih tinggi menunjukkan bahwa model memiliki kemampuan diskriminatif yang lebih baik dalam memisahkan kelas positif dan negatif.

$$FPR = \frac{FP}{FP + TN}$$

Rumus 2. 8 Rumus AUC [53]

Keterangan :

- $TN = \text{True Negative}$ (prediksi benar untuk kelas negatif)
- $FP = \text{False Positive}$ (prediksi salah untuk kelas positif)

di mana FP adalah jumlah prediksi salah untuk kelas positif dan TN adalah jumlah prediksi benar untuk kelas negatif.

Nilai AUC berada pada rentang antara 0 hingga 1, dengan interpretasi sebagai berikut:

1. $AUC = 1.0$ menunjukkan model sempurna dalam membedakan kelas.
2. $AUC = 0.5$ menunjukkan model tidak lebih baik dari tebakan acak.
3. $AUC < 0.5$ menunjukkan model memiliki kemampuan diskriminatif yang buruk.

Keunggulan utama AUC dijelaskan terletak pada ketahanannya terhadap ketidakseimbangan kelas serta sifatnya yang tidak bergantung pada ambang batas klasifikasi (threshold-independent). Dalam penelitian sebelumnya ditegaskan bahwa AUC mampu memberikan evaluasi yang lebih stabil dibandingkan metrik lainnya karena penilaiannya didasarkan pada kemampuan model untuk secara konsisten memberikan skor probabilitas yang lebih tinggi kepada sampel kelas positif dibandingkan sampel kelas negatif [54]. Dengan demikian, AUC direpresentasikan sebagai ukuran kualitas pemeringkatan (ranking ability) model secara menyeluruh, dan tidak terbatas pada performa pada satu titik threshold tertentu saja.

2.4.8 Marketing-Specific Metrics

Dalam konteks aplikasi bisnis, khususnya pada sistem prioritas lead dan rekomendasi tindak lanjut (*follow-up*), organisasi seringkali menghadapi keterbatasan kapasitas operasional. Sebagai contoh, tim sales atau telemarketing hanya mampu menindaklanjuti sebagian kecil dari seluruh calon nasabah dalam suatu periode (misalnya hanya 10%–20% dari total populasi). Kondisi ini menyebabkan perlunya metrik evaluasi tambahan yang mampu menilai performa

model pada subset teratas dari hasil pemeringkatan probabilitas. Metrik tersebut dikenal sebagai *capacity-based metrics* atau *top-K ranking metrics* [55].

Beberapa metrik yang digunakan antara lain:

a. Precision@K

Precision@K mengukur presisi model pada K% teratas dari hasil pemeringkatan probabilitas. Metrik ini menghitung seberapa banyak *lead* yang benar-benar melakukan booking (positif) di antara K% *lead* dengan skor tertinggi.

$$\text{Precision@K} = \frac{TP_K}{TP_K + FP_K}$$

Rumus 2. 9 Rumus Precision@k [55]

b. Recall@K (Capture@K)

Recall@K, sering disebut juga Capture Rate@K, menunjukkan proporsi total kelas positif (booking) yang berhasil ditangkap oleh model pada K% teratas dari ranking[55].

$$\text{Recall@K} = \frac{TP_K}{TP + FN}$$

Rumus 2. 10 Rumus Recall@k [55]

Metrik ini penting karena menunjukkan seberapa banyak potensi booking yang dapat diamankan jika tim hanya melakukan follow-up pada K% *lead* teratas.

c. Lift@K

Lift@K adalah rasio antara Precision@K dan tingkat dasar (*baseline rate*) pada seluruh populasi. Lift mengukur seberapa besar peningkatan kinerja model dibandingkan pemilihan acak[56].

$$\text{LiftK} = \frac{\frac{TP_K}{TP_K + FP_K}}{\frac{TP}{TP + FP + FN + TN}}$$

Rumus 2. 11 Rumus Lift @k [56]

Lift@K > 1 menunjukkan bahwa model secara signifikan lebih baik daripada pemilihan acak.

2.5 Teori tentang Tools/Software yang digunakan

2.5.1 Bahasa Pemrograman Python

Python merupakan bahasa pemrograman tingkat tinggi yang bersifat interpretatif dan serbaguna, dirancang dengan filosofi yang menekankan pada keterbacaan kode [57]. Dengan sintaks yang sederhana dan ekspresif, Python memungkinkan pengembang menuliskan konsep komputasi dengan jumlah baris kode yang lebih sedikit dibandingkan bahasa seperti C++ atau Java.

Sebagai bahasa pemrograman yang bersifat open-source, Python tidak memerlukan biaya lisensi dan didukung oleh ekosistem pustaka yang sangat luas, mencakup bidang ilmiah, statistik, serta machine learning. Pustaka populer seperti NumPy, Pandas, scikit-learn, TensorFlow, dan Matplotlib telah digunakan secara luas sehingga Python diposisikan sebagai salah satu bahasa utama dalam analisis data dan pengembangan model kecerdasan buatan. Selain itu, dukungan komunitas yang besar, dokumentasi yang komprehensif, serta sifatnya yang lintas platform dijelaskan sebagai faktor yang menjadikan Python unggul dalam mendukung seluruh tahapan proyek analitik—mulai dari proses ekstraksi data hingga visualisasi hasil [58].

2.5.3 Lingkungan Pengembangan

1. Visual Studio Code (VS Code)

Visual Studio Code (VS Code) dipahami sebagai source code editor yang ringan, gratis, dan bersifat lintas platform, yang dikembangkan oleh Microsoft.

Berdasarkan dokumentasi resmi Microsoft, berbagai fitur inti seperti debugging, syntax highlighting, intelligent code completion, integrated terminal, serta dukungan bawaan untuk version control Git dijelaskan telah disediakan sehingga proses penulisan, pengujian, dan pengelolaan kode dapat dilakukan secara lebih efisien [59].

Selain itu, ekosistem ekstensi yang sangat luas pada Visual Studio Code memungkinkan penambahan dukungan bahasa pemrograman, alat pengembangan web, data, maupun machine learning sesuai kebutuhan proyek. Dalam dokumentasi yang sama ditegaskan bahwa fleksibilitas, performa yang ringan, serta kemampuan kustomisasi yang tinggi menjadikan VS Code diposisikan sebagai salah satu editor kode yang paling banyak digunakan dalam pengembangan aplikasi modern [59].

2. Google Collaboratory

Google Collaboratory (Google Colab) dipahami sebagai lingkungan komputasi berbasis cloud yang disediakan oleh Google untuk menjalankan kode Python secara interaktif melalui notebook. Platform ini dijelaskan dirancang untuk mendukung kebutuhan analisis data, machine learning, dan eksperimen komputasi tanpa memerlukan instalasi perangkat lunak pada komputer lokal. Berdasarkan dokumentasi resmi Google, akses gratis ke GPU dan TPU telah disediakan, beserta lingkungan eksekusi yang dilengkapi dengan pustaka populer seperti NumPy, Pandas, scikit-learn, TensorFlow, dan PyTorch, sehingga pengembangan serta pengujian model dapat dilakukan secara efisien [60].

Sebagai layanan kolaboratif, Google Colab memungkinkan notebook untuk disimpan, dibagikan, dan diedit secara real-time melalui integrasi langsung dengan Google Drive. Fitur ini dijelaskan mendukung kerja tim dalam proyek analisis data dan machine learning, terutama ketika proses eksplorasi data, pemodelan, dan dokumentasi eksperimen dilakukan. Selain itu, Google Colab menyediakan kemampuan untuk memasang pustaka tambahan, menjalankan visualisasi interaktif, serta menghubungkan notebook dengan dataset eksternal. Dengan kemudahan akses, dukungan komputasi yang fleksibel, serta integrasi dengan

ekosistem Google Cloud, platform ini diposisikan sebagai salah satu layanan yang banyak digunakan dalam penelitian, pembelajaran, maupun pengembangan aplikasi berbasis data.

3. Streamlit

Streamlit dipahami sebagai framework open-source berbasis Python yang dirancang untuk memudahkan pembuatan aplikasi web interaktif secara cepat, khususnya pada bidang data science, analitik, dan machine learning. Berdasarkan dokumentasi resmi Streamlit, pengembang dapat mengubah skrip Python menjadi aplikasi web interaktif hanya dengan beberapa baris kode, tanpa memerlukan pengetahuan mengenai bahasa front-end seperti HTML, CSS, ataupun JavaScript [61]. Framework ini menyediakan arsitektur deklaratif yang memungkinkan komponen antarmuka dihasilkan langsung dari kode Python, sehingga proses pembuatan dashboard, visualisasi data, dan demonstrasi model machine learning dapat dilakukan secara lebih ringkas dan efisien.

Keunggulan utama Streamlit dijelaskan terletak pada kesederhanaan sintaks, performa yang cepat, serta dukungan komponen bawaan seperti tombol, slider, tabel data, grafik interaktif, dan file uploader. Dokumentasi resmi juga menyebutkan bahwa Streamlit dapat diintegrasikan dengan berbagai pustaka visualisasi dan machine learning, termasuk Matplotlib, Plotly, Altair, TensorFlow, dan PyTorch, sehingga analisis data dapat divisualisasikan secara real-time dalam aplikasi yang responsif. Selain itu, Streamlit mendukung deployment melalui Streamlit Community Cloud maupun platform pihak ketiga seperti GitHub, yang memungkinkan pembaruan aplikasi dilakukan secara otomatis dan kolaboratif. Dengan kemudahan penggunaan dan fleksibilitas yang tinggi, Streamlit diposisikan sebagai salah satu framework populer dalam pengembangan aplikasi analitik modern secara cepat dan efisien.