BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Relevan

Pada bagian ini disajikan rangkuman sejumlah penelitian sebelumnya yang memiliki keterkaitan dengan topik penelitian yang dibahas. Kajian-kajian tersebut memberikan gambaran mengenai metode yang digunakan sekaligus menunjukkan bagaimana pendekatan tersebut diterapkan dan diuji pada berbagai permasalahan yang berbeda.

1. (Nawaf, 2022) meneliti penggunaan algoritma *Random Forest* dalam menganalisis faktor-faktor yang memengaruhi profitabilitas dan risiko pada industri perbankan. Fokus utama penelitian ini adalah pada kemampuan *Random Forest* dalam mengukur *relative variable importance*, yaitu tingkat kontribusi masing-masing variabel terhadap perubahan kinerja bank. Hasil analisis menunjukkan bahwa variabel spesifik internal bank, seperti ukuran bank, kekuatan pasar, efisiensi operasional, serta indikator pendapatan bunga berupa *Net Interest Margin*, memiliki pengaruh dominan terhadap tingkat profitabilitas. Sementara itu, risiko kredit yang direpresentasikan oleh rasio *Non-Performing Loan* (NPL) lebih banyak dipengaruhi oleh kondisi makroekonomi dan faktor regulasi. Al-Maskati menegaskan bahwa *Random Forest* lebih unggul dibandingkan regresi linear dalam menangkap hubungan nonlinier dan interaksi kompleks antarvariabel keuangan, sehingga mampu menghasilkan model yang lebih stabil dan akurat dalam memodelkan kinerja serta pendapatan bank.
2. (Hartini et al., 2021) mengkaji penggunaan algoritma *Random Forest* untuk mengestimasi probabilitas terjadinya krisis perbankan dengan memanfaatkan data keuangan dari 79 negara pada periode 1981–1999. Algoritma *Random Forest* dipilih karena memiliki karakteristik yang *robust*, tidak sensitif terhadap *outlier*, serta mampu mengurangi risiko *overfitting* dibandingkan metode statistik konvensional yang bergantung pada asumsi distribusi tertentu.

   Hasil penelitian menunjukkan bahwa *Random Forest* mampu memodelkan hubungan nonlinier antarvariabel makroekonomi dan indikator keuangan bank secara efektif, serta menghasilkan tingkat akurasi, sensitivitas, presisi, dan *F1-*

score yang tinggi dalam membedakan kondisi krisis dan non-krisis perbankan. Temuan ini menegaskan bahwa pendekatan *ensemble learning* berbasis pohon keputusan memiliki keunggulan dalam menangkap interaksi kompleks antarvariabel keuangan dan memberikan kinerja prediktif yang lebih stabil dibandingkan model regresi tradisional. Oleh karena itu, penelitian ini memperkuat dasar pemilihan algoritma *Random Forest* dalam pemodelan dan prediksi variabel keuangan perbankan yang bersifat dinamis dan tidak linier, termasuk dalam konteks peramalan pendapatan bunga kredit.

3. (Jatnika et al., 2024) meneliti pengaruh metode pengumpulan data terhadap kinerja model *data mining* dengan membandingkan data hasil kuesioner dan *web mining*. Penelitian ini menggunakan algoritma *Support Vector Machine* dan *Naive Bayes Classifier* untuk mengevaluasi tingkat akurasi dan nilai *Area Under Curve* (AUC). Hasil penelitian menunjukkan bahwa data yang terstruktur dengan baik dan telah melalui tahapan *preprocessing* seperti *data cleaning*, transformasi, serta seleksi fitur mampu menghasilkan performa model yang lebih stabil dan akurat. Temuan ini menegaskan bahwa kualitas *dataset* dan tahapan persiapan data merupakan faktor krusial dalam membangun model prediksi yang andal. Prinsip tersebut relevan dengan penelitian ini, karena keberhasilan penerapan *Random Forest Regressor* dalam memprediksi pendapatan bunga kredit konsumtif juga sangat bergantung pada kualitas dan kesiapan data historis yang digunakan.
4. (Jatnika et al., 2025)menerapkan algoritma *Random Forest* untuk mengklasifikasikan kualitas layanan berdasarkan data umpan balik peserta sertifikasi internasional di ITCC ITPLN. Penelitian ini mengolah sebanyak 2.720 data yang telah melalui tahapan *text preprocessing*, meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil evaluasi menunjukkan bahwa *Random Forest* mampu mencapai tingkat akurasi sebesar 88,97%, dengan nilai presisi 0,92, *recall* 0,91, dan *F1-score* 0,92. Temuan tersebut menunjukkan bahwa algoritma *Random Forest* memiliki kemampuan yang sangat baik dalam mempelajari pola data yang kompleks serta menghasilkan prediksi yang stabil dan konsisten. Karakteristik ini relevan dengan penelitian skripsi ini, mengingat

data keuangan perbankan juga bersifat multivariat dan memiliki hubungan nonlinier antarvariabel.

5. (Abdurrasyid et al., 2021) membahas penerapan regresi linier berganda untuk memprediksi kuota pemesanan bahan bakar pada Stasiun Pengisian Bahan Bakar Umum. Variabel yang digunakan meliputi stok sisa dan stok masuk sebagai variabel independen, serta stok keluar sebagai variabel dependen. Evaluasi kinerja model dilakukan menggunakan metrik *Mean Absolute Percentage Error* (MAPE), yang menghasilkan tingkat kesalahan sebesar 11,0% untuk pertalite dan 13,2% untuk solar. Hasil tersebut menunjukkan bahwa pemodelan prediktif berbasis data historis mampu memberikan estimasi yang cukup akurat dan dapat dimanfaatkan sebagai dasar pengambilan keputusan operasional. Meskipun metode yang digunakan berbeda, penelitian ini relevan karena menekankan pentingnya evaluasi kuantitatif menggunakan metrik kesalahan, seperti MAPE, MAE, atau RMSE, dalam menilai kinerja model prediksi, sebagaimana diterapkan dalam penelitian ini pada evaluasi *Random Forest Regressor* untuk memprediksi pendapatan bunga kredit konsumtif.
6. (Asri et al., 2023)mengkaji penerapan algoritma *C4.5 decision tree* dalam mengklasifikasikan titik serta jenis gangguan pada jaringan distribusi tenaga listrik. Proses pembentukan pohon keputusan dilakukan berdasarkan perhitungan nilai *entropy* dan *information gain* untuk menentukan atribut yang paling berpengaruh pada setiap percabangan. Hasil penelitian menunjukkan bahwa metode berbasis pohon mampu memetakan hubungan antarvariabel secara sistematis dan mengidentifikasi faktor dominan penyebab gangguan. Meskipun objek penelitian berada pada sektor ketenagalistrikan, pendekatan *tree-based* yang digunakan memiliki relevansi metodologis dengan penelitian ini. *Random Forest* sebagai pengembangan dari *decision tree* melalui konsep *ensemble learning* memiliki keunggulan dalam meningkatkan stabilitas dan akurasi model, sehingga temuan ini memperkuat dasar pemilihan *Random Forest Regressor* dalam memodelkan hubungan nonlinier antarvariabel keuangan.
2. (Wulandari et al., 2024) menerapkan algoritma *C4.5* untuk mengidentifikasi faktor-faktor yang memengaruhi ketepatan waktu kelulusan mahasiswa. Variabel yang dianalisis meliputi Indeks Prestasi Kumulatif, asal sekolah, keaktifan

organisasi, keaktifan dalam Unit Kegiatan Mahasiswa, status bekerja, dan motivasi. Berdasarkan perhitungan *information gain*, diperoleh bahwa variabel motivasi dan Indeks Prestasi Kumulatif memiliki kontribusi paling besar dalam menentukan kelulusan tepat waktu. Model yang dihasilkan diuji menggunakan *confusion matrix* dan menunjukkan tingkat akurasi di atas 70%. Walaupun berada pada bidang pendidikan, penelitian ini relevan secara metodologis karena menekankan pentingnya pemilihan variabel dan analisis *feature importance*, yang sejalan dengan mekanisme *feature importance* pada *Random Forest* dalam mengidentifikasi faktor dominan yang memengaruhi variabel target.

3. (Brandão et al., 2023) melakukan peramalan pendapatan individu nasabah perbankan dengan membandingkan beberapa metode *machine learning*, yaitu regresi linear, *Random Forest*, *XGBoost*, *Neural Network*, dan *Support Vector Regressor*. Penelitian ini mengombinasikan data historis nasabah dengan variabel makroekonomi, seperti pertumbuhan Produk Domestik Bruto, inflasi, suku bunga, nilai tukar, dan tingkat pengangguran. Hasil penelitian menunjukkan bahwa metode *ensemble tree-based*, khususnya *Random Forest* dan *XGBoost*, mampu memberikan kinerja prediksi yang lebih baik dibandingkan metode statistik konvensional dalam menangkap hubungan nonlinier antarvariabel. Selain itu, *Random Forest* memiliki keunggulan dalam mengidentifikasi tingkat kepentingan variabel melalui analisis *feature importance*, sehingga faktor-faktor dominan yang memengaruhi pendapatan dapat diketahui secara lebih jelas.
4. (Aysan et al., 2024)menganalisis kemampuan prediktif algoritma *Random Forest* dalam mengevaluasi manajemen risiko pada perbankan syariah dengan menggunakan data *Global Islamic Banking Survey* periode 2016–2021 yang dipadukan dengan indikator kinerja keuangan dan variabel makroekonomi. Hasil penelitian menunjukkan bahwa *Random Forest* mampu menangani data berdimensi besar, mengatasi permasalahan *missing values*, serta memodelkan hubungan nonlinier antarvariabel dengan tingkat akurasi yang tinggi. Variabel kualitas aset, rasio kredit bermasalah (NPL), suku bunga, dan profitabilitas terbukti memiliki kontribusi signifikan dalam menjelaskan risiko dan kinerja bank. Temuan ini memperkuat bahwa algoritma *Random Forest* layak digunakan untuk memodelkan variabel keuangan perbankan yang kompleks.

5. (Liva & RENÇBER, 2023) membandingkan kinerja algoritma *Random Forest* dan *XGBoost* dalam memprediksi profitabilitas bank pada sektor perbankan simpanan di Turki dengan menggunakan indikator *Return on Assets* (ROA), *Return on Equity* (ROE), serta variabel spesifik bank dalam kerangka *CAMELS*. Hasil analisis menunjukkan bahwa metode *ensemble* berbasis pohon menghasilkan nilai kesalahan prediksi yang lebih rendah serta koefisien determinasi (R^2) yang lebih tinggi dibandingkan regresi linear. Selain itu, *Random Forest* terbukti efektif dalam melakukan pemeringkatan kepentingan variabel (*feature importance*), sehingga faktor-faktor yang paling berpengaruh terhadap profitabilitas bank dapat diidentifikasi secara lebih akurat. Relevansi penelitian ini dengan skripsi terletak pada kesamaan penggunaan *Random Forest Regressor* sebagai model prediksi untuk variabel kinerja keuangan yang dipengaruhi oleh banyak faktor dengan pola hubungan nonlinier, termasuk pendapatan bunga kredit konsumtif.

Tabel 2. 1 Penelitian Relevan

Keterangan	**Isi**
No.	**1**
Penulis	Al-Maskati, Nawaf (2022)
Judul	*Random Forest for Bank Profitability and Risk Analysis*
Tujuan Penelitian	Menganalisis faktor-faktor yang memengaruhi profitabilitas dan risiko perbankan dengan menggunakan algoritma Random Forest serta mengidentifikasi variabel yang paling berpengaruh terhadap kinerja bank.
Data Penelitian	Data keuangan bank yang meliputi Net Interest Margin (NIM), *Non-Performing Loan* (NPL), ukuran bank, efisiensi operasional, dan variabel makroekonomi.
Metode	*Random Forest*
Hasil Penelitian	Random Forest mampu menangkap hubungan nonlinier antarvariabel dan menunjukkan bahwa variabel spesifik bank seperti ukuran bank, efisiensi, dan NIM memiliki pengaruh dominan terhadap profitabilitas, serta menghasilkan model yang lebih stabil dibandingkan regresi linear.
Persamaan dan Perbedaan	Persamaan: Sama-sama menggunakan algoritma Random Forest pada data perbankan dan melibatkan variabel kualitas kredit seperti NPL serta indikator pendapatan bunga. Perbedaan: Penelitian Al-Maskati berfokus pada profitabilitas bank secara umum, sedangkan penelitian ini secara khusus memprediksi Pendapatan Bunga Kredit Konsumtif pada satu bank daerah dengan mempertimbangkan faktor waktu.
Keterangan	**Isi**

No.	**2**
Penulis	Sri Hartini, Zuherman Rustam, Glori Stephani Saragih, dan María Jesús Segovia Vargas (2021)
Judul	*Estimating probability of banking crises using random forest*
Tujuan Penelitian	Mengestimasi probabilitas terjadinya krisis perbankan menggunakan algoritma Random Forest.
Data Penelitian	Data keuangan 79 negara periode 1981–1999.
Metode	*Random Forest*
Hasil Penelitian	*Random Forest* menghasilkan akurasi, sensitivitas, presisi, dan F1-Score yang lebih tinggi dibandingkan metode statistik konvensional dalam memprediksi krisis perbankan.
Persamaan dan Perbedaan	Persamaan: Sama-sama menggunakan Random Forest untuk memodelkan data keuangan perbankan yang bersifat kompleks dan nonlinier. Perbedaan: Sri Hartini, Zuherman Rustam, Glori Stephani Saragih, dan María Jesús Segovia Vargas (2021), memprediksi kondisi krisis bank secara global, sedangkan penelitian ini memprediksi besaran pendapatan bunga kredit konsumtif pada tingkat cabang bank.
Keterangan	**Isi**
No.	**3**
Penulis	Hendra Jatnika, Ari Waluyo, dan Abdul Azis (2024)
Judul	*A Comparative Study on Data Collection Methods Investigating Optimal Datasets for Data Mining Analysis*
Tujuan Penelitian	Menganalisis pengaruh kualitas data dan tahapan *preprocessing* terhadap kinerja model prediksi.
Data Penelitian	Dataset klasifikasi umum.
Metode	SVM dan *Naive Bayes*
Hasil Penelitian	*Preprocessing*, pembersihan data, dan seleksi fitur terbukti meningkatkan akurasi model secara signifikan.
Persamaan dan Perbedaan	Persamaan: Sama-sama menekankan pentingnya preprocessing, pembagian data latih dan uji, serta evaluasi kinerja model. Perbedaan: Penelitian ini menggunakan *Random Forest* pada data perbankan, sedangkan Jatnika et al. menggunakan SVM dan *Naive Bayes* pada data umum.
Keterangan	**Isi**
No.	**4**
Penulis	Hendra Jatnika, Luqman, Mochamad Farid Rifai, dan Nasya Miranda Umar (2025)
Judul	*Application of Random Forest Classification Method in Determining the Best Quality Service in the Implementation of International Certification at ITCC ITPLN*
Tujuan Penelitian	Mengklasifikasikan kualitas layanan berdasarkan data umpan balik peserta.

Data Penelitian	2.720 data teks umpan balik.
Metode	*Random Forest*
Hasil Penelitian	*Random Forest* menghasilkan akurasi, presisi, recall, dan F1-Score yang tinggi dalam klasifikasi kualitas layanan.
Persamaan dan Perbedaan	Persamaan: Sama-sama menggunakan algoritma Random Forest dan mengevaluasi kinerja model dengan metrik akurasi. Perbedaan: Penelitian ini melakukan regresi untuk memprediksi pendapatan bunga, sedangkan Jatnika et al. melakukan klasifikasi kualitas layanan berbasis teks.
Keterangan	**Isi**
No.	**5**
Penulis	Abdurrasyid, Indrianto, dan Meilia Nur Susanti (2021)
Judul	Prediksi Kuota Pemesanan Bahan Bakar Pada SPBU dengan Metode Regresi Linear Berganda
Tujuan Penelitian	Memprediksi kebutuhan stok BBM untuk mendukung pengambilan keputusan operasional.
Data Penelitian	Data historis stok SPBU.
Metode	Regresi Linier Berganda
Hasil Penelitian	Model menghasilkan nilai MAPE yang rendah sehingga mampu memberikan estimasi kebutuhan stok yang cukup akurat.
Persamaan dan Perbedaan	Persamaan: Sama-sama melakukan peramalan berbasis data historis dan mengevaluasi kinerja model menggunakan metrik kesalahan. Perbedaan: Metode yang digunakan adalah regresi linier, sedangkan penelitian ini menggunakan *Random Forest Regressor* untuk data perbankan.
Keterangan	**Isi**
No.	**6**
Penulis	Yessy Asri, Dwina Kuswardani, Widya Nita Suliyanti, dan Chrystyna Monica Tambunan (2023)
Judul	Klasifikasi Gangguan Jaringan Listrik dengan C4.5
Tujuan Penelitian	Mengidentifikasi titik dan jenis gangguan jaringan distribusi listrik.
Data Penelitian	Data teknis jaringan distribusi.
Metode	C4.5
Hasil Penelitian	Algoritma C4.5 mampu membentuk pohon keputusan yang mengidentifikasi variabel paling berpengaruh.
Persamaan dan Perbedaan	Persamaan: Sama-sama menggunakan pendekatan berbasis pohon keputusan. Perbedaan: Yessy Asri, Dwina Kuswardani, Widya Nita Suliyanti, dan Chrystyna Monica Tambunan (2023). menggunakan C4.5 untuk klasifikasi gangguan listrik,

	sedangkan penelitian ini menggunakan Random Forest sebagai metode ensemble untuk regresi pendapatan bunga.
Keterangan	**Isi**
No.	**7**
Penulis	Dewi Arianti Wulandari, Herman Bedi, Luqman, dan Ahmad Arfan Mashuda (2024)
Judul	Prediksi Kelulusan Mahasiswa dengan C4.5
Tujuan Penelitian	Menentukan faktor dominan yang memengaruhi kelulusan tepat waktu.
Data Penelitian	Data akademik mahasiswa.
Metode	C4.5
Hasil Penelitian	IPK dan motivasi merupakan variabel yang paling berpengaruh terhadap kelulusan tepat waktu.
Persamaan dan Perbedaan	Persamaan: Sama-sama menganalisis pengaruh variabel dan konsep *feature importance*. Perbedaan: Objek penelitian pendidikan, sedangkan penelitian ini fokus pada pendapatan bunga kredit perbankan.
Keterangan	**Isi**
No.	**8**
Penulis	Ricardo Brandão dan Simona Šulžickytė (2023)
Judul	*Individual revenue forecasting in the banking sector*
Tujuan Penelitian	Memprediksi pendapatan individu nasabah bank.
Data Penelitian	Data nasabah dan variabel makroekonomi.
Metode	*Random Forest, XGBoost,* SVR*, Neural Network*
Hasil Penelitian	Metode ensemble berbasis pohon memberikan performa prediksi terbaik.
Persamaan dan Perbedaan	Persamaan: Sama-sama memprediksi variabel pendapatan di sektor perbankan menggunakan Random Forest. Perbedaan: Skala prediksi pada individu nasabah, sedangkan penelitian ini pada pendapatan bunga kredit tingkat cabang.
Keterangan	**Isi**
No.	**9**
Penulis	Ahmet Faruk Aysan, Bekir Sait Ciftler, dan Ibrahim Musa Unal (2024)
Judul	*Predictive Power of Random Forests in Analyzing Risk Management in Islamic Banking*
Tujuan Penelitian	Menganalisis risiko dan kinerja bank syariah.
Data Penelitian	Data keuangan dan makro 2016–2021.
Metode	Random Forest
Hasil Penelitian	NPL, suku bunga, dan NIM memiliki kontribusi signifikan terhadap risiko dan kinerja bank.

Persamaan dan Perbedaan	Persamaan: Sama-sama menggunakan Random Forest dan variabel NPL serta suku bunga. Perbedaan: Fokus pada manajemen risiko perbankan syariah, sedangkan penelitian ini fokus pada prediksi pendapatan bunga kredit konsumtif bank daerah.
Keterangan	**Isi**
No.	**10**
Penulis	Liva Oflazoğlu dan Ömer Faruk Rençber (2023)
Judul	*RANDOM FOREST AND XGBOOST IMPLEMENTATIONS TO PREDICT BANK PROFITABILITY: EVIDENCE FROM TURKISH DEPOSIT BANKS*
Tujuan Penelitian	Membandingkan kinerja Random Forest dan XGBoost dalam memprediksi ROA dan ROE.
Data Penelitian	Data bank simpanan Turki (CAMELS).
Metode	Random Forest, XGBoost
Hasil Penelitian	Metode ensemble menghasilkan RMSE lebih rendah dan R^2 lebih tinggi dibanding regresi linear.
Persamaan dan Perbedaan	Persamaan: Sama-sama memodelkan kinerja keuangan bank dengan Random Forest. Perbedaan: Variabel target ROA/ROE, sedangkan penelitian ini memprediksi Pendapatan Bunga Kredit Konsumtif.

Berdasarkan telaah terhadap penelitian terdahulu, penerapan algoritma *Random Forest Regressor* dalam sektor perbankan umumnya difokuskan pada klasifikasi risiko kredit, penilaian kelayakan nasabah, atau analisis profitabilitas pada skala institusi secara agregat. Namun, kajian yang secara khusus mengembangkan model prediksi pendapatan bunga kredit konsumtif pada tingkat cabang bank daerah masih sangat terbatas. Padahal, keputusan operasional pada level cabang sangat dipengaruhi oleh karakteristik lokal yang tidak selalu tercermin dalam data perbankan skala nasional. Ketiadaan model prediksi yang dirancang secara spesifik untuk konteks cabang menyebabkan proses perencanaan pendapatan bunga kredit konsumtif masih bergantung pada pendekatan konvensional yang kurang adaptif terhadap dinamika data. Oleh karena itu, penelitian ini memiliki nilai kebaruan dengan memfokuskan pengembangan algoritma *Random Forest Regressor* pada prediksi pendapatan bunga kredit konsumtif berbasis data historis cabang bank daerah, sehingga tidak hanya memperluas penerapan algoritma dalam konteks yang lebih spesifik, tetapi juga memberikan kontribusi praktis bagi pengambilan keputusan berbasis data di tingkat operasional perbankan.

2.2 Landasan Teori

2.2.1 Pendapatan Bunga Kredit Konsumtif

Pendapatan bunga kredit konsumtif merupakan salah satu komponen utama pendapatan operasional bank yang berasal dari selisih antara bunga yang dibayarkan debitur atas fasilitas kredit dengan biaya dana yang dikeluarkan bank. Kredit konsumtif diberikan kepada individu untuk memenuhi kebutuhan non-produktif seperti pembelian rumah, kendaraan, pendidikan, dan konsumsi rumah tangga, sehingga karakteristiknya sangat dipengaruhi oleh tingkat suku bunga, jumlah debitur aktif, serta kemampuan bayar nasabah. Menurut penelitian (Rahmawati & Handayani, 2025), pendapatan bunga kredit konsumtif berperan penting dalam menjaga stabilitas kinerja keuangan bank karena bersifat relatif berkelanjutan dan menjadi sumber arus kas utama, namun tetap sensitif terhadap perubahan kondisi ekonomi dan risiko kredit.

Selain faktor suku bunga dan volume penyaluran kredit, kualitas kredit yang tercermin dari rasio *Non-Performing Loan* (NPL) juga berpengaruh langsung terhadap realisasi pendapatan bunga. Peningkatan NPL dapat menurunkan efektivitas pendapatan bunga karena sebagian bunga tidak dapat diakui sebagai pendapatan akibat terjadinya tunggakan atau restrukturisasi kredit. (Brandão et al., 2023) menjelaskan bahwa hubungan antara variabel-variabel keuangan perbankan, termasuk suku bunga, jumlah debitur, dan risiko kredit, bersifat nonlinier dan saling berinteraksi, sehingga diperlukan pendekatan analitis yang mampu menangkap pola hubungan tersebut secara lebih komprehensif. Dalam konteks ini, pemodelan pendapatan bunga kredit konsumtif tidak hanya berfungsi sebagai alat peramalan, tetapi juga sebagai dasar evaluasi kinerja portofolio kredit dan perencanaan strategi penyaluran kredit berbasis data historis.

2.2.2 Prediksi Pendapatan

Prediksi pendapatan merupakan proses peramalan nilai pendapatan di masa mendatang berdasarkan pola historis dan hubungan antarvariabel yang memengaruhi kinerja keuangan suatu entitas. Dalam konteks perbankan, prediksi pendapatan bunga kredit konsumtif menjadi instrumen penting dalam perencanaan strategis, pengelolaan likuiditas, dan evaluasi portofolio kredit, karena hasil prediksi ini dapat membantu manajemen dalam mengantisipasi perubahan tren ekonomi serta

menetapkan kebijakan penyaluran kredit yang lebih responsif terhadap dinamika pasar. Penelitian (Brandão et al., 2023) menunjukkan bahwa penggunaan model prediktif seperti *Random Forest*, *XGBoost*, dan metode *ensemble* lainnya mampu menangkap hubungan nonlinier antara variabel internal bank dan indikator makroekonomi seperti suku bunga atau inflasi, sehingga menghasilkan estimasi pendapatan yang lebih akurat dibandingkan model statistik tradisional yang berbasis asumsi linear. Pada penelitian tersebut, model *Random Forest* terbukti efektif dalam memprediksi nilai pendapatan nasabah dengan mempertimbangkan faktor-faktor yang kompleks dan berinteraksi.

Selanjutnya, prediksi pendapatan tidak hanya berfokus pada peramalan angka tunggal, tetapi juga berperan sebagai alat evaluatif untuk menilai kontributor utama dari setiap variabel input terhadap hasil akhir model. Misalnya, pendekatan *Random Forest* dapat menghitung *feature importance* yang menunjukkan seberapa besar setiap variabel memengaruhi prediksi pendapatan, sehingga manajemen bank dapat mengidentifikasi variabel mana yang memberikan dampak signifikan terhadap hasil prediksi. Hal ini penting karena pemahaman terhadap variabel dominan dapat meningkatkan akurasi model sekaligus memberikan wawasan strategis bagi perumusan kebijakan kredit dan mitigasi risiko. Peran prediksi pendapatan melalui teknik *machine learning* seperti ini semakin menguat terutama ketika data yang tersedia bersifat kompleks dan tidak linier, sehingga pendekatan prediktif menjadi lebih relevan dalam pengambilan keputusan berbasis data dibandingkan pendekatan konvensional.

2.2.3 Random Forest Regressor

Random Forest Regressor adalah metode *machine learning* berbasis *ensemble learning* yang membangun sejumlah pohon keputusan (*decision tree*) dari *subset* data acak dan mengombinasikan hasil prediksi setiap pohon untuk menghasilkan estimasi yang lebih akurat dan stabil. Pendekatan ini mengurangi varians model dan potensi *overfitting* yang sering terjadi pada pohon tunggal karena setiap pohon dilatih dari sampel *bootstrap* yang berbeda, sehingga prediksi akhir mencerminkan gambaran umum pola data yang kompleks. Berbagai penelitian menunjukkan bahwa *Random Forest Regressor* mampu memberikan prediksi yang lebih baik dibandingkan metode statistik tradisional seperti regresi linear, terutama ketika hubungan antarvariabel

bersifat non linier dan data memiliki dimensi yang tinggi. Studi oleh (Muchtar & Afiyati, 2024) menunjukkan bahwa *Random Forest Regressor* memberikan kinerja prediksi harga komoditas (premium rice) lebih efektif dibandingkan *Linear Regressor* pada dataset yang sama, yang mengindikasikan keunggulan *Random Forest* dalam konteks prediksi regresi data nyata.

Prediksi akhir yang dihasilkan oleh *Random Forest Regressor* diperoleh dengan mengambil rata-rata dari prediksi semua pohon dalam hutan (*forest*), yaitu:

$$\hat{Y} = \frac{1}{N}\sum_{i=1}^{N} \hat{Y}_i \qquad (2.1)$$

di mana:

$\hat{Y}$= prediksi akhir yang dihasilkan oleh *Random Forest Regressor*

N= jumlah pohon dalam hutan

$\hat{Y}_i$= prediksi yang dihasilkan oleh pohon keputusan ke-i.

Rumus ini menunjukkan bahwa setiap pohon memberikan kontribusi terhadap prediksi akhir sehingga model dapat menangkap pola hubungan yang kompleks antarvariabel. Selain itu, *Random Forest Regressor* juga mampu menyediakan informasi *feature importance*, yaitu ukuran kontribusi relatif masing-masing variabel terhadap hasil prediksi, yang membantu peneliti memahami variabel mana yang paling berpengaruh terhadap target prediksi dalam konteks regresi.

2.2.4 *Mean Absolute Error* (MAE)

Mean Absolute Error (MAE) adalah metrik evaluasi yang umum digunakan untuk mengukur akurasi model regresi dengan cara menghitung rata-rata dari selisih absolut antara nilai prediksi dan nilai aktual. MAE memberikan ukuran kesalahan dalam satuan yang sama dengan variabel yang diprediksi, sehingga mudah diinterpretasikan secara praktis dalam konteks performa model. Berbeda dengan metrik berbasis kuadrat seperti *Mean Squared Error* (MSE), MAE memberikan kontribusi yang seimbang untuk setiap selisih prediksi tanpa memfokuskan penalti lebih besar pada kesalahan ekstrem, sehingga dapat memberikan gambaran kesalahan rata-rata secara linear. Analisis MAE juga sering dipakai dalam literatur *machine learning* dan *data science* karena kemudahan dalam perhitungan serta stabilitasnya

terhadap *outlier*, terutama ketika tujuan evaluasi adalah menilai kesalahan prediksi rata-rata secara langsung (Jierula et al., 2021).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \qquad (2.2)$$

di mana:

MAE= *Mean Absolute Error*;

n= jumlah pengamatan;

y_i= nilai aktual variabel dependen;

$\hat{y}_i$= nilai prediksi model untuk observasi ke-i.

Rumus tersebut menunjukkan bahwa MAE merupakan aritmetika rata-rata dari kesalahan absolut antara nilai yang diprediksi dan nilai aktual, sehingga nilai MAE yang lebih kecil mencerminkan akurasi prediksi yang lebih tinggi. MAE sering digunakan bersama metrik lain seperti *Root Mean Squared Error* untuk memberikan evaluasi performa model yang lebih komprehensif.

2.2.5 *Root Mean Squared Error* (RMSE)

Root Mean Squared Error (RMSE) adalah metrik evaluasi yang banyak digunakan untuk menilai akurasi prediksi dalam model regresi. RMSE mengukur rata-rata kuadrat dari selisih antara nilai prediksi dan nilai aktual, kemudian mengambil akar kuadrat dari nilai rata-rata tersebut. Karena error dikuadratkan sebelum dirata-ratakan, RMSE memberikan bobot lebih besar pada deviasi besar antara prediksi dan observasi nyata, sehingga cocok untuk menyoroti model yang menghasilkan prediksi yang jauh dari realitas. Metrik ini sering dipakai dalam penelitian yang mengevaluasi *forecast accuracy* pada berbagai domain, termasuk ekonomi, *machine learning*, dan sistem informasi, karena kesederhanaannya dan interpretabelnya yang langsung pada satuan variabel target. (Jierula et al., 2021)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (2.3)$$

di mana:

RMSE= *Root Mean Squared Error*;

n= jumlah data pengamatan;

y_i= nilai aktual pada observasi ke-i;

$\hat{y}_i$= nilai prediksi model untuk observasi ke-i.

Rumus ini menunjukkan bahwa RMSE merupakan akar dari *Mean Squared Error* (MSE), yang merupakan rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual. Nilai RMSE yang lebih rendah menandakan bahwa prediksi model lebih dekat dengan nilai aktual secara keseluruhan, sehingga model dinilai lebih akurat. RMSE sering digunakan bersama dengan metrik lain seperti MAE untuk mengevaluasi kinerja model secara komprehensif karena memberikan gambaran yang berbeda terkait distribusi error prediksi.

2.2.6 R-Squared (R^2)

Koefisien determinasi, yang biasa disebut *R-Squared* (R^2), adalah metrik evaluasi yang digunakan untuk mengukur seberapa baik sebuah model regresi menjelaskan variabilitas (keragaman) data aktual melalui variabel-variabel prediktor yang digunakan. Secara sederhana, R^2 merepresentasikan proporsi variasi variabel dependen yang dapat dijelaskan oleh variabel independen dalam model; nilai R^2 berkisar antara kurang dari nol hingga 1, di mana nilai yang lebih dekat ke 1 menunjukkan model yang lebih baik dalam menangkap variasi data. Studi oleh (Chicco et al., 2021)menegaskan bahwa *R-Squared* merupakan metrik yang informatif dalam evaluasi regresi karena memberikan ukuran langsung mengenai kemampuan model dalam menjelaskan variasi hasil prediksi, dan dalam banyak kasus lebih informatif dibandingkan metrik lain seperti *Mean Absolute Error* atau *Root Mean Squared Error.* Nilai R^2 yang tinggi secara empiris menunjukkan bahwa model regresi mampu menangkap pola utama dalam data yang berbeda secara signifikan.
Secara matematis, *R-Squared* dihitung sebagai perbandingan antara varians yang dijelaskan oleh model dengan total varians dari variabel dependen:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (2.4)$$

di mana:

R^2= koefisien determinasi

y_i= nilai aktual variabel dependen

$\hat{y}_i$= nilai prediksi model

$\bar{y}$= nilai rata-rata aktual variabel dependen

n= jumlah observasi.

Rumus tersebut menunjukkan bahwa R^2 mengukur proporsi penurunan total variasi yang tidak dapat dijelaskan model setelah memasukkan variabel prediktor dibandingkan dengan variasi awal data. Semakin kecil sisa variasi yang tidak dijelaskan (*residual variance*), semakin besar nilai R^2, sehingga model dapat dianggap semakin baik dalam menjelaskan variasi data.

2.2.7 *Machine Learning*

Machine learning merupakan disiplin ilmu yang menjadi fondasi penting dalam analisis data dan pembuatan model prediktif tanpa memerlukan instruksi eksplisit untuk setiap langkahnya. Secara umum, *machine learning* didefinisikan sebagai suatu pendekatan komputasi di mana sistem komputer *belajar* dari data yang tersedia dengan tujuan meningkatkan kinerjanya dalam melakukan tugas tertentu seiring bertambahnya pengalaman atau informasi dari data tersebut. Konsep ini berbeda dengan pemrograman tradisional yang bergantung pada aturan eksplisit; dalam *machine learning*, model secara otomatis membangun hubungan atau pola dari data yang diberikan untuk kemudian digunakan dalam proses prediksi atau klasifikasi pada data baru. Definisi ini selaras dengan pemahaman bahwa *machine learning* adalah cabang penting dari *artificial intelligence* yang memanfaatkan algoritma dan teknik statistik untuk mengekstraksi wawasan dari data (Trisal & Mandloi, 2021)

Dalam konteks penelitian kuantitatif berbasis data finansial seperti perbankan, *machine learning* memungkinkan pemodelan hubungan kompleks antarvariabel yang tidak mudah ditangkap oleh metode statistik konvensional, terutama ketika hubungan variabel bersifat non-linear dan dipengaruhi oleh interaksi multivariat. Penggunaan *machine learning* dalam prediksi finansial telah banyak diteliti karena kemampuannya untuk secara otomatis menemukan pola dari data historis serta memprediksi hasil di masa depan dengan akurasi yang lebih tinggi dibanding pendekatan tradisional. Implementasi tersebut meliputi pemodelan risiko kredit, prediksi pendapatan, serta klasifikasi nasabah menurut karakteristik finansial mereka, yang semuanya menunjukkan pentingnya *machine learning* dalam mendukung pengambilan keputusan berbasis data (Sahu et al., 2023).

2.2.8 Python

Python merupakan bahasa pemrograman tingkat tinggi yang banyak digunakan dalam *machine learning* karena sintaksnya yang sederhana, ekspresif, dan mudah dipelajari oleh peneliti maupun praktisi data. Bahasa ini menyediakan ekosistem pustaka (*libraries*) yang kuat untuk pengolahan data, visualisasi, dan pemodelan prediktif, seperti *NumPy*, *Pandas*, *Scikit-learn*, *TensorFlow*, dan *PyTorch*, sehingga mempercepat pengembangan model *machine learning* tanpa perlu menulis kode secara detail dari awal (Raschka et al., 2020). Python mendukung berbagai paradigma pemrograman dan dapat digunakan di berbagai platform sistem operasi, sehingga menjadi pilihan utama dalam penelitian dan implementasi solusi berbasis data di berbagai bidang, termasuk finansial dan perbankan.

Dalam konteks *machine learning*, Python memfasilitasi seluruh proses analisis data, mulai dari tahap *data preprocessing*, ekstraksi fitur, pembentukan model, hingga evaluasi hasil prediksi melalui metrik yang telah ditentukan. Keunggulan Python dalam hal *readability*, komunitas pengguna yang besar, serta ketersediaan dokumentasi dan tutorial yang luas menjadikannya sangat sesuai untuk tugas prediksi seperti yang dilakukan dalam penelitian ini, yakni membangun model *Random Forest Regressor* untuk memprediksi variabel keuangan perbankan.

2.2.9 Google Colab

Google Colab merupakan lingkungan *cloud based Jupyter notebook* yang memungkinkan peneliti dan praktisi untuk menulis, menjalankan, serta berbagi kode Python secara langsung melalui peramban tanpa memerlukan konfigurasi perangkat lunak secara lokal. Karena berjalan sepenuhnya di *cloud*, Google Colab menyediakan akses terhadap sumber daya komputasi seperti *Graphics Processing Units* (GPU) dan *Tensor Processing Units* (TPU) yang sangat mendukung proses komputasi intensif, khususnya dalam pelatihan model *machine learning* dan *deep learning*. Penelitian yang dilakukan oleh (Carneiro et al., 2016) menunjukkan bahwa Google Colab mampu meningkatkan efisiensi waktu komputasi dan performa pelatihan model *deep learning* dibandingkan lingkungan komputasi konvensional. Selain itu, studi oleh (Gabriel Do Breviário et al., 2025) menegaskan bahwa Google Colab memiliki potensi besar dalam analisis *machine learning* dan *big data* karena kemudahan akses, integrasi dengan pustaka Python populer, serta dukungan kolaborasi secara *real-time*.

Oleh karena itu, Google Colab dinilai sebagai platform yang efektif dan efisien untuk mendukung kegiatan analisis data dan pengembangan model *machine learning* dalam penelitian.

2.2.10 *Data Visualization*

Data Visualization atau visualisasi data merupakan proses mempresentasikan data dalam bentuk grafik, diagram, atau representasi visual lainnya untuk memudahkan pemahaman terhadap pola, tren, dan hubungan antarvariabel dalam suatu himpunan data. Teknik ini sangat penting dalam *machine learning* karena membantu peneliti mengeksplorasi karakteristik dataset sebelum, selama, maupun setelah pemodelan dilakukan, seperti mengenali *outlier*, tren musiman, distribusi variabel, serta hubungan antarfitur yang dapat memengaruhi kinerja model prediktif. Visualisasi data tidak hanya meningkatkan keterbacaan, tetapi juga mendukung pembuatan keputusan yang lebih cepat dan tepat berbasis informasi yang tersaji secara visual, sehingga memperkuat proses *exploratory data analysis* (EDA) sebagai dasar langkah pemodelan berikutnya.

Dalam konteks pengembangan model *machine learning*, visualisasi data dipandang sebagai tahapan penting yang memperkuat *data preprocessing*, *feature selection*, serta evaluasi hasil model. Penelitian oleh (Shinde & Shivthare, 2024) menunjukkan bahwa visualisasi data memberikan *insights* yang intuitif terhadap pola data yang kompleks, sehingga mempercepat identifikasi hubungan, anomali, serta tren yang relevan dalam persiapan data untuk pelatihan model. Dengan pemanfaatan teknik visualisasi seperti *scatter plot*, *heatmap*, dan grafik distribusi lainnya, analisis data menjadi lebih informatif dan mendukung model *machine learning* yang lebih efektif serta efisien.

2.2.11 *Heatmap Korelasi*

Heatmap korelasi merupakan alat visualisasi yang digunakan untuk menunjukkan hubungan linear antar variabel dalam suatu kumpulan data secara dua dimensi, di mana setiap sel mewakili koefisien korelasi antara dua variabel yang dibandingkan. Warna pada *heatmap* disusun menurut intensitas hubungan misalnya semakin cerah atau semakin gelap menunjukkan korelasi positif atau negatif yang semakin kuat sehingga memudahkan peneliti melihat pola hubungan (positif, negatif, atau tidak berkorelasi) antar variabel sekaligus dalam satu tampilan. *Heatmap* ini

sangat berguna dalam tahap *exploratory data analysis* (*EDA*), karena dapat mengarahkan pemilihan fitur atau pemahaman awal tentang interaksi antar variabel sebelum dilakukan pemodelan lanjutan ketika menganalisis dataset yang memiliki banyak fitur.

Dalam konteks penelitian prediksi menggunakan *machine learning*, *heatmap* korelasi sering digunakan untuk menilai keterkaitan awal antara variabel independen dan variabel target sehingga dapat digunakan untuk menentukan variabel yang relevan dan mengurangi redundansi fitur sebelum pemodelan. (Singh & Nagahara, 2024) menggunakan *matrix-based heatmap* untuk merangkum dan mengevaluasi keterkaitan antar banyak variabel dalam dataset energi, yang membantu mereka memahami pola hubungan dan memutuskan langkah analisis selanjutnya secara sistematis.

2.2.12 *Flowchart*

Flowchart merupakan representasi grafik dari urutan langkah atau alur kerja dalam sebuah proses, yang digunakan untuk memetakan tahapan dan keputusan dalam pengembangan suatu sistem *machine learning*. Dalam konteks *machine learning*,*flowchart* menggambarkan urutan mulai dari pemahaman masalah (*problem understanding*), pengumpulan dan pembersihan data (*data preparation*), pemodelan (*model development*), evaluasi (*model evaluation*), hingga penerapan (*deployment*), sehingga setiap langkah proses dapat dipahami secara sistematik dan konsisten. Pendekatan semacam ini membantu peneliti dan praktisi memastikan proses pengolahan data dan pembangunan model dilakukan secara terstruktur, meminimalkan kesalahan, dan mendukung pengulangan eksperimen (*reproducibility*) yang lebih baik dalam studi prediktif dan eksperimental.

No.	Simbol Flowchart	Nama	Arti Simbol Flowchart
1		*Terminator*	Awal atau akhir konsep (prosedur)
2		*Process*	Proses operasional
3		*Document*	Dokumen atau laporan berupa *print out*
4		*Decision*	Keputusan atau sub-point. Garis yang terhubung dengan bentuk decision menujuk pada situasi-situasi yang berbeda sesuai dengan keputusan yang digambarkan
5		Data	Input dan Output (Contohnya, Input: feedback dari pelanggan, Output: desain produk baru)
6		*On-Page Reference/ Connector*	Penghubung alur dalam halaman yang sama
7		*Off-Page Reference/ Off-Page Connector*	Penghubung alur dalam halaman yang berbeda
8		*Flow*	Arah alur dalam konsep (prosedur)

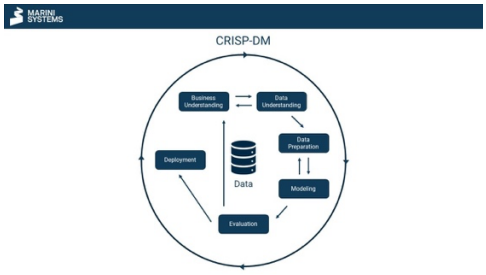
Gambar 2. 1 Flowchart

Dalam studi tinjauan tentang peran *machine learning* dalam peningkatan *workflow* ilmu data (*data science workflows*) (Sharma, 2025), dijelaskan bahwa struktur alur kerja *machine learning* mencakup serangkaian langkah berulang yang saling terkait mulai dari tahap awal eksplorasi data hingga evaluasi model akhir, sehingga *flowchart* tidak hanya menggambarkan tahapan teknis, tetapi juga fungsi kontrol kualitas yang memastikan output model dapat dikendalikan dan dioptimalkan. Dengan representasi *flowchart* ini, kompleksitas alur *machine learning* dapat divisualisasikan secara jelas, mendukung pemahaman yang sistematis atas proses pembelajaran mesin yang bersifat iteratif dan data centred.

2.2.13 CRISP-DM (*Cross-Industry Standard Process for Data Mining*)

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah kerangka kerja metodologis yang telah menjadi *standar de facto* dalam proses proyek *data mining* dan *machine learning*, karena memberikan struktur yang sistematis untuk menjalankan serangkaian tahap pengolahan data mulai dari pemahaman tujuan bisnis hingga implementasi solusi. Metodologi ini terdiri dari enam fase utama *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* yang saling terkait dan dapat dilakukan secara iteratif sehingga setiap tahapan dapat ditinjau ulang berdasarkan kebutuhan proyek yang kompleks. Penelitian tinjauan literatur menunjukkan bahwa *CRISP-DM* tetap menjadi model yang dominan dalam praktik *data mining* lintas industri, membantu peneliti memastikan proses pemodelan dilakukan secara terstruktur dan dapat diulang

(*reproducibility*) dengan tetap menyesuaikan kebutuhan aplikasi nyata (Schröer et al., 2021).



Gambar 2. 2 Tahapan CRISP-DM

Sumber : https://marini.systems/en/glossary/crisp-dm/

Pada Gambar 2.2 tahapan *CRISP-DM* meliputi:

1. *Business Understanding*

   Tahap *Business Understanding* berfokus pada perumusan tujuan penelitian yang diturunkan dari sasaran dan strategi organisasi. Pada tahap ini ditetapkan permasalahan utama, kebutuhan analisis, serta ruang lingkup kegiatan *data mining* atau *machine learning* yang akan dilakukan, sehingga diperoleh kerangka awal yang jelas sebagai dasar pelaksanaan penelitian.
2. *Data Understanding*

   Tahap *Data Understanding* mencakup proses pengumpulan dan eksplorasi data. Data dapat diperoleh dari berbagai sumber, seperti basis data, lembar kerja, atau sistem informasi lainnya, baik dalam bentuk terstruktur maupun tidak terstruktur. Pada tahap eksplorasi, data dianalisis secara deskriptif untuk memahami karakteristiknya, meliputi kualitas, jumlah, struktur, format, distribusi, serta hubungan awal antarvariabel tanpa menggunakan model prediktif. Hasil analisis ini dapat digunakan untuk meninjau kembali tujuan dan ruang lingkup penelitian pada tahap *Business Understanding* apabila ditemukan ketidaksesuaian antara data yang tersedia dan kebutuhan analisis.
3. *Data Preparation*

   Tahap *Data Preparation* bertujuan untuk menyiapkan data agar sesuai dengan kebutuhan pemodelan. Proses ini meliputi pembersihan data, transformasi, pemilihan variabel, serta penyusunan data dalam format yang dapat diolah oleh

algoritma, umumnya berupa tabel yang terdiri atas baris sebagai observasi dan kolom sebagai fitur atau prediktor. Pada tahap ini juga dilakukan pembagian data menjadi data latih (*training set*) dan data uji (*testing set*).

4. *Modeling*

   Pada tahap *Modeling*, algoritma *machine learning* dilatih menggunakan data latih untuk membentuk model prediktif. Proses ini melibatkan penentuan parameter model sesuai dengan karakteristik data. Beberapa algoritma yang umum digunakan antara lain regresi logistik, *Random Forest*, *Decision Tree*, dan *Gradient Boosting*, yang masing-masing dapat diuji untuk memperoleh model dengan kinerja terbaik.

5. *Evaluation*

   Tahap *Evaluation* dilakukan untuk menilai kinerja model yang telah dibangun menggunakan data yang tidak dilibatkan dalam proses pelatihan, yaitu data uji. Evaluasi bertujuan untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik. Apabila kinerja model belum memenuhi kriteria yang ditetapkan, maka proses dapat diulang, baik dengan memperbaiki kualitas data, menambahkan variabel, maupun mengganti algoritma yang digunakan. Tahap ini bersifat iteratif hingga diperoleh model dengan performa yang memadai.

6. *Deployment*

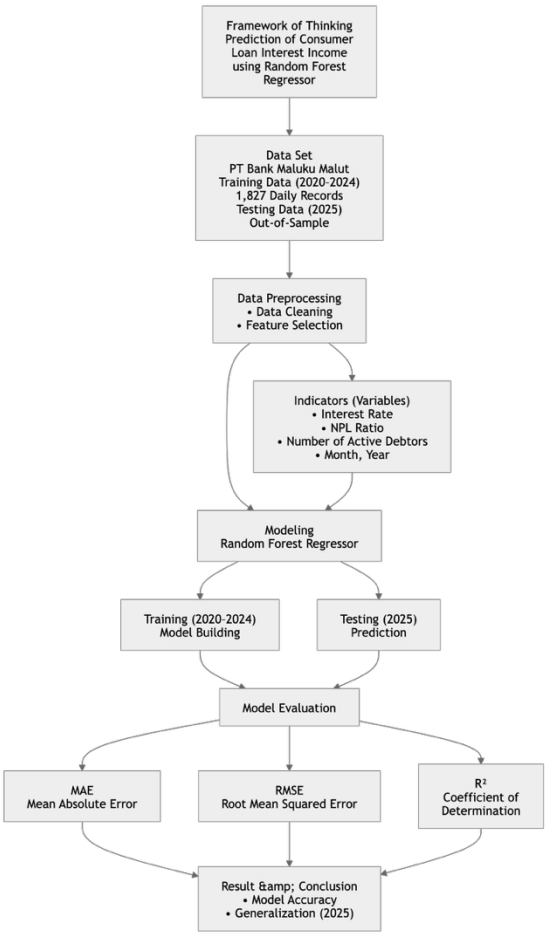
   Tahap *Deployment* merupakan fase penerapan model ke dalam lingkungan operasional. Pada tahap ini, model yang telah dievaluasi dan dinyatakan layak diimplementasikan untuk digunakan dalam proses pengambilan keputusan atau sistem pendukung keputusan sesuai dengan tujuan penelitian.

2.2.14 *Fishbone*

Fishbone Diagram, yang dikenal sebagai *cause and effect diagram* atau Ishikawa Diagram, merupakan alat visual yang digunakan untuk mengidentifikasi serta mengelompokkan berbagai faktor penyebab yang berkontribusi terhadap suatu permasalahan tertentu. Diagram ini digambarkan menyerupai kerangka ikan, di mana permasalahan utama ditempatkan sebagai bagian kepala, sedangkan faktor-faktor penyebab potensial disusun pada tulang-tulang utama yang bercabang dari tulang punggung. Melalui pengelompokan penyebab ke dalam kategori tertentu, Fishbone Diagram membantu memetakan hubungan sebab

akibat secara sistematis sehingga akar permasalahan dapat dianalisis secara lebih mendalam dan terstruktur. Pendekatan ini memudahkan peneliti dalam memahami keterkaitan antar faktor yang bersifat kompleks, khususnya pada tahap *Business Understanding* sebelum dilakukan pemodelan prediktif. Penggunaan Fishbone Diagram dalam analisis sebab akibat telah banyak diterapkan untuk mendukung proses pengambilan keputusan dan pemecahan masalah secara efektif dalam berbagai bidang penelitian dan manajemen kualitas (Kumah et al., 2023).

2.3 Kerangka Pemikiran



Gambar 2. 3 Kerangka Pemikiran

Kerangka pemikiran penelitian ini menggambarkan alur sistematis dalam membangun dan mengevaluasi model prediksi pendapatan bunga kredit konsumtif menggunakan algoritma *Random Forest Regressor*. Penelitian dimulai dari identifikasi permasalahan, yaitu fluktuasi pendapatan bunga kredit konsumtif yang dipengaruhi oleh berbagai faktor internal perbankan dan faktor waktu. Untuk menjawab permasalahan

tersebut, digunakan pendekatan *machine learning* berbasis metode ensemble yang mampu menangkap hubungan nonlinier antarvariabel.

Tahap pertama adalah penentuan dan pengolahan data, yaitu data historis kredit PT Bank Pembangunan Daerah Maluku dan Maluku Utara Cabang Masohi periode 2020–2024 sebagai data pelatihan (*training set*). Data ini merepresentasikan pola historis aktivitas kredit dan digunakan untuk membangun model prediksi. Selanjutnya, data tahun 2025 digunakan sebagai data pengujian (*testing set*) untuk mengevaluasi kemampuan generalisasi model terhadap periode yang belum pernah dilatih sebelumnya.

Tahap berikutnya adalah proses pemodelan menggunakan algoritma *Random Forest Regressor*. Metode ini dipilih karena mampu mengurangi risiko *overfitting* melalui kombinasi banyak pohon keputusan (*decision tree*) yang dibangun dari sampel data berbeda melalui teknik *bootstrap*. Hasil prediksi dari setiap pohon kemudian dirata-ratakan untuk menghasilkan estimasi akhir yang lebih stabil.

Variabel yang digunakan dalam penelitian ini meliputi suku bunga, rasio *Non-Performing Loan* (NPL), jumlah debitur aktif, bulan, dan tahun. Variabel-variabel tersebut secara teoritis memiliki pengaruh terhadap pendapatan bunga, baik dari sisi volume kredit, risiko pembiayaan, maupun pola musiman.

Kinerja model dievaluasi menggunakan tiga metrik, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). Ketiga metrik tersebut digunakan untuk menilai tingkat akurasi, stabilitas, serta kemampuan model dalam menjelaskan variasi pendapatan bunga kredit konsumtif secara komprehensif.