

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian ini disusun dengan merujuk pada berbagai referensi dan informasi dari penelitian-penelitian sebelumnya yang membahas mengenai metode *clustering*, prediksi menggunakan Random, topik perbankan, serta aspek lain yang memiliki keterkaitan dengan penelitian ini. Adapun Tabel 2.1 berikut menyajikan beberapa penelitian terdahulu yang relevan :

Tabel 2. 1 Penelitian Terdahulu

Jurnal 1	
Nama Penulis	Wendi Hu, Chujian Shao, Wenyu Zhang
Tahun Penelitian	2025
Judul Penelitian	<i>Predicting U.S. bank failures and stress testing with machine learning algorithms [19]</i>
Penerbit	<i>Elsevier</i>
Permasalahan	Penelitian ini mengidentifikasi kebutuhan untuk memprediksi kegagalan bank di Amerika Serikat, terutama dalam konteks krisis keuangan yang terjadi baru-baru ini. Kegagalan bank besar, seperti Silicon Valley Bank dan Signature Bank, menyoroti pentingnya deteksi dini untuk mengurangi risiko finansial dan menjaga stabilitas ekonomi.
Metode	Penelitian ini menggunakan berbagai model machine learning, termasuk Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), dan Convolutional Neural Network (CNN). Data yang dianalisis berasal dari Federal Deposit Insurance Corporation (FDIC) dan mencakup variabel keuangan yang relevan. Pendekatan exponentially weighted moving averages (EWMA) digunakan untuk meningkatkan akurasi prediksi.
Hasil dan Kesimpulan	Hasil penelitian menunjukkan bahwa model Random Forest mencapai akurasi tertinggi, yaitu 98.4% dalam memprediksi kegagalan bank. Variabel keuangan kunci, seperti rasio leverage dan kecukupan modal, memiliki dampak signifikan pada prediksi. Analisis sensitivitas

	menunjukkan bahwa akurasi model sangat bergantung pada fitur-fitur kunci, dan penurunan performa terjadi ketika fitur tersebut terpengaruh oleh kondisi stres. Kesimpulannya, integrasi metode pemodelan canggih dan fitur sensitif waktu dapat memberikan prediksi yang lebih robust, memberikan wawasan berharga untuk pengelolaan risiko keuangan.
--	---

Jurnal 2

Nama Penulis	Dhea Aryona Putri, Gusmanely Z
Tahun Penelitian	2023
Penerbit	Inovasi Pembangunan : Jurnal Kelitbangan
Judul Penelitian	Perbandingan Algoritma <i>K-Means</i> dan <i>K-Medoids</i> dalam Pengelompokan Bank Umum Syariah di Indonesia dengan <i>Davies-Bouldin Index</i> , <i>Calinski-Harabasz Index</i> , dan <i>Silhouette Coefficient</i> . [21]
Permasalahan	Terdapat kesenjangan dalam kinerja keuangan bank umum syariah di Indonesia, yang terlihat dari rasio <i>Return On Assets</i> (ROA) yang menunjukkan hasil yang kurang optimal.
Metode	Penelitian ini menggunakan analisis klaster dengan dua pendekatan, yaitu <i>K-Means</i> dan <i>K-Medoids</i> . Validasi hasil pengelompokan dilakukan dengan tiga indeks yaitu <i>Davies-Bouldin Index</i> , <i>Calinski-Harabasz Index</i> , dan <i>Silhouette Coefficient</i> .
Hasil dan Kesimpulan	Algoritma <i>K-Means</i> memberikan performa terbaik dalam pengelompokan bank umum syariah berdasarkan rasio keuangan dibandingkan dengan <i>K-Medoids</i> . Nilai <i>Davies-Bouldin Index</i> untuk <i>K-Means</i> (0,1109) lebih baik dibandingkan <i>K-Medoids</i> (0,7558). <i>Calinski-Harabasz Index</i> menunjukkan nilai yang lebih baik untuk <i>K-Medoids</i> (0,1328) dibandingkan <i>K-Means</i> (0,0937). Namun, untuk <i>Silhouette Coefficient</i> , <i>K-Means</i> (0,8516) lebih unggul dibandingkan <i>K-Medoids</i> (0,2193). Kesimpulan utama adalah <i>K-Means</i> merupakan algoritma yang lebih efektif untuk pengelompokan bank umum syariah dalam penelitian ini.

Jurnal 3

Nama Penulis	P. K. Viswanathan, Suresh Srinivasan, N. Hariharan
Tahun Penelitian	2020
Judul Penelitian	<i>Predicting Financial Health of Banks for Investor Guidance Using Machine Learning Algorithms</i> [22]
Penerbit Jurnal	<i>Journal of Emerging Market Finance</i>
Permasalahan	Penelitian ini mengidentifikasi kurangnya mekanisme klasifikasi yang komprehensif untuk membedakan

	tingkat kesehatan finansial bank di India. Selain itu, terdapat indikasi bahwa penilaian pasar dan pemeringkatan kredit sering kali tidak konsisten dengan kondisi keuangan aktual bank, sehingga diperlukan model yang dapat memberikan gambaran yang lebih akurat bagi pengambilan keputusan.
Metode	Penelitian menggunakan pendekatan <i>hybrid machine learning</i> , dimulai dengan <i>K-Means clustering</i> untuk mengelompokkan bank ke dalam tiga tingkat kesehatan finansial yaitu tinggi, sedang, rendah. Hasil <i>cluster</i> kemudian diprediksi menggunakan <i>Linear Discriminant Analysis</i> (LDA), CART, dan <i>Random Forest</i> untuk mengevaluasi akurasi klasifikasi. Model-model tersebut diuji menggunakan indikator kinerja utama seperti ROA, NITA, GNPATA, serta rasio kecukupan modal. Studi ini juga membandingkan hasil klasifikasi dengan valuasi pasar (PB ratio) dan peringkat kredit lembaga rating.
Hasil dan Kesimpulan	Model berhasil mengklasifikasikan 44 bank ke dalam tiga tingkat kesehatan finansial secara akurat. Bank dengan kategori “Low” menunjukkan GNPATA dua digit dan modal yang lebih rendah dibanding kategori lain, menunjukkan konsistensi klasifikasi model. Sejumlah bank besar justru tidak berada pada kategori “High”, dan ditemukan anomali antara hasil model dengan penilaian pasar serta rating kredit AAA yang diberikan lembaga pemeringkat. LDA dan <i>Random Forest</i> menunjukkan akurasi hampir sama yaitu di angka 95,36% dan 95,93%, namun LDA dinilai lebih unggul karena memiliki daya jelaskan yang kuat terhadap variabel input. Penelitian menyimpulkan bahwa teknik <i>machine learning</i> dapat menjadi alat evaluasi risiko yang lebih akurat bagi deposan maupun regulator, serta berpotensi mendukung kerangka pengawasan berbasis risiko yang lebih responsif.
Jurnal 4	
Nama Penulis	Dwi Bagus Saputra, Vihi Atina, Faulinda Ely Nastiti
Tahun Penelitian	2024
Judul Penelitian	Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit Menggunakan Algoritma <i>Random Forest</i> [23]
Penerbit	<i>Idealis: Indonesia Journal Information System</i>
Permasalahan	Koperasi Simpan Pinjam Baitut Tamwil Tazakka menghadapi tingkat kredit macet (non-performing loans) yang cukup tinggi, yang mengancam stabilitas finansial dan keberlanjutan operasional lembaga.

	Diperlukan model prediktif yang lebih akurat untuk mengidentifikasi potensi kredit macet lebih awal, sehingga risiko dapat diminimalkan.
Metode	Penelitian menerapkan algoritma <i>Random Forest</i> sebagai metode utama untuk memprediksi kemungkinan kredit macet. Data yang dianalisis dibagi menggunakan skema <i>train-test</i> split 90:10, kemudian dievaluasi menggunakan berbagai metrik performa seperti akurasi dan metrik pendukung lainnya yang relevan dalam pemodelan klasifikasi.
Hasil dan Kesimpulan	Model <i>Random Forest</i> menunjukkan performa sangat baik dengan akurasi prediksi mencapai 94,8%. Studi ini menegaskan bahwa Random Forest mampu menjadi alat yang efektif dalam mendukung manajemen risiko kredit pada lembaga keuangan mikro karena sifatnya yang stabil dan akurat. Namun, penelitian ini masih terbatas pada satu jenis algoritma dan ruang lingkup data yang relatif sempit. Penelitian selanjutnya disarankan memperluas variasi algoritma serta cakupan data untuk menghasilkan analisis yang lebih komprehensif. Temuan ini berkontribusi pada pengembangan strategi mitigasi risiko yang lebih proaktif dan berdampak langsung pada penguatan stabilitas lembaga keuangan mikro.
Jurnal 5	
Nama Penulis	Sadiya Firdose RZ, Laxmana Rao Goranta
Tahun Penelitian	2025
Judul Penelitian	<i>Applying Machine Learning to Economic Analysis: Predicting Bank Failures in India Using Decision Tree Models [24]</i>
Penerbit	IEEE
Permasalahan	Penelitian ini mengidentifikasi tantangan dalam menilai stabilitas keuangan institusi perbankan di India, khususnya dalam memprediksi potensi kebangkrutan bank.
Metode	Penelitian ini menggunakan algoritma pohon keputusan (C5.0) untuk menganalisis data keuangan bank. Model dibangun dan diuji menggunakan 4,850 variabel, dengan 20% data digunakan sebagai sampel uji.
Hasil dan Kesimpulan	Hasil penelitian menunjukkan bahwa model C5.0 mencapai akurasi 89.7% dalam mengklasifikasikan bank sebagai sukses atau gagal. Indikator kunci seperti margin finansial dan profitabilitas aset diidentifikasi sebagai yang paling berpengaruh. Meskipun ada beberapa keterbatasan, seperti ketersediaan data dan

	ketidakmampuan untuk belajar secara bertahap, model pohon keputusan terbukti sebagai alat yang berharga untuk analisis keuangan. Penelitian ini menyarankan eksplorasi lebih lanjut dengan algoritma ML alternatif untuk meningkatkan akurasi prediksi.
Jurnal 6	
Nama Penulis	Karim Farag, Loubna Ali, Noah Cheruiyot Mutai, Rabia Luqman, Ahmed Mahmoud, Nol Krasniqi
Tahun Penelitian	2025
Judul Penelitian	<i>Machine Learning for Predicting Bank Stability: The Role of Income Diversification in European Banking [25]</i>
Penerbit	<i>MDPI</i>
Permasalahan	Studi ini menyoroti perdebatan yang berkelanjutan mengenai efektivitas diversifikasi pendapatan (income diversification) dalam meningkatkan stabilitas bank di Eropa. Hasil penelitian terdahulu menunjukkan temuan yang saling bertentangan, sehingga belum ada bukti jelas apakah diversifikasi pendapatan berperan memperkuat atau justru melemahkan stabilitas bank. Selain itu, model prediksi stabilitas bank yang ada masih lemah dalam mendeteksi bank yang berisiko tinggi (distressed banks), terutama akibat masalah ketidakseimbangan data (class imbalance) pada sektor perbankan Eropa.
Metode	Penelitian menggunakan pendekatan hybrid yang menggabungkan dua teknik ekonometrika generalized Method of Moments (GMM) dan fixed-effect model untuk menguji pengaruh diversifikasi pendapatan terhadap stabilitas bank (diukur dengan Z-score, NPL, dan CAR). Selain itu, penelitian menerapkan algoritma machine learning Random Forest dan Support Vector Machine (SVM) untuk membangun model prediksi stabilitas bank. Data berasal dari 26 negara Eropa pada periode 2000–2021. Validasi model dilakukan menggunakan 10-fold cross-validation.
Hasil dan Kesimpulan	Hasil ekonometrika menunjukkan bahwa diversifikasi pendapatan dapat meningkatkan stabilitas bank, tercermin dari penurunan NPL, meningkatnya CAR, dan peningkatan Z-score. Pada aspek prediktif, algoritma SVM menunjukkan akurasi lebih tinggi dibandingkan Random Forest pada data awal. Namun, kedua model gagal mendeteksi bank bermasalah karena ketidakseimbangan kelas. Setelah menerapkan teknik oversampling, performa kedua

	<p>model meningkat signifikan, dengan tingkat deteksi distress melebihi 95%. Secara keseluruhan, penelitian menegaskan bahwa diversifikasi pendapatan merupakan indikator penting stabilitas bank di Eropa, dan bahwa model prediktif berbasis SVM dan Random Forest yang telah diseimbangkan dapat berfungsi sebagai early warning system yang efektif.</p>
Jurnal 7	
Nama Penulis	Emmanuel Awoin, Peter Appiahene, Frank Gyasi, Abdulai Sabtiwu
Tahun Penelitian	2020
Judul Penelitian	<i>Predicting the Performance of Rural Banks in Ghana Using Machine Learning Approach [26]</i>
Penerbit	<i>Advances in Fuzzy Systems</i>
Permasalahan	Penelitian ini menyoroti permasalahan rendahnya kinerja dan stabilitas bank-bank pedesaan di Ghana, yang banyak dipengaruhi oleh praktik ilegal, manajemen yang tidak efisien, serta lemahnya pengelolaan risiko. Kondisi tersebut menimbulkan kebutuhan mendesak akan model prediksi yang dapat menilai kesehatan bank secara akurat berbasis rasio keuangan.
Metode	Penelitian ini menggunakan data rasio keuangan yang dikumpulkan dari berbagai rural banks anggota ARB sebagai dasar analisis. Dua pendekatan utama diterapkan dalam proses pemodelan. Pertama, <i>metode Random Forest Variable Importance</i> digunakan untuk menyeleksi variabel yang paling berpengaruh dari total 16 prediktor, dan hasilnya menunjukkan bahwa 13 di antaranya memiliki nilai <i>Mean Decrease Gini</i> (MDG) $\geq 0,10$ sehingga dianggap relevan untuk digunakan dalam model. Kedua, penelitian membangun model prediksi menggunakan tiga algoritma pohon keputusan, yaitu C5.0, CART, dan C4.5, yang masing-masing diuji untuk mengukur kemampuan klasifikasi kinerja bank. Evaluasi model dilakukan menggunakan metrik akurasi dan <i>confusion matrix</i> , dengan membandingkan performa pada data latih dan data uji guna memastikan konsistensi serta keandalan prediksi yang dihasilkan.
Hasil dan Kesimpulan	Hasil penelitian menunjukkan bahwa algoritma <i>Decision Tree</i> mampu memberikan prediksi yang sangat akurat terhadap kinerja bank pedesaan. Model C5.0 tampil sebagai metode terbaik dengan akurasi 100% pada data latih maupun uji, dan mampu mengklasifikasikan seluruh kategori status bank dengan

	benar. Algoritma CART memperoleh akurasi 84,6%, sementara C4.5 mencapai akurasi rata-rata 83,34% dan akurasi uji 82,83%. Analisis <i>variable importance</i> melalui Random Forest menunjukkan bahwa 13 dari 16 rasio keuangan merupakan prediktor paling berpengaruh, menandakan bahwa sebagian besar rasio tersebut memiliki kontribusi signifikan dalam membedakan kondisi kesehatan bank. Secara keseluruhan, penelitian menyimpulkan bahwa C5.0 merupakan algoritma yang paling stabil dan akurat,
--	---

Jurnal 8

Nama Penulis	Maysas Yafi Urrochman, Hasyim Asy'ari, Abdur Ro'uf
Tahun Penelitian	2024
Judul Penelitian	<i>Using Machine Learning Techniques to Predict Financial Distress in Rural Banks in Indonesia</i> [27]
Penerbit	<i>Journal of Informatics Development</i>
Permasalahan	Penelitian ini mengidentifikasi masalah finansial yang dihadapi oleh bank-bank pedesaan di Indonesia, termasuk likuiditas dan manajemen yang buruk, yang mengarah pada kebangkrutan. Maka dibutuhkan model yang dapat memprediksi risiko finansial ini.
Metode	Penelitian ini menggunakan pendekatan dua tahap, yaitu klasifikasi dan regresi, dengan algoritma pohon keputusan (<i>Decision Tree</i>) dan <i>Naïve Bayes</i> untuk klasifikasi, serta <i>Random Forest Regression</i> untuk prediksi waktu terjadinya masalah keuangan. Data yang digunakan berasal dari laporan keuangan 150 bank pedesaan antara tahun 2015 hingga 2019.
Hasil dan Kesimpulan	Hasil penelitian menunjukkan bahwa algoritma <i>Decision Tree</i> memberikan akurasi 88%, sementara <i>Naïve Bayes</i> mencapai 82%. Model regresi <i>Random Forest</i> menunjukkan rata-rata kesalahan absolut 1,2 bulan dan kesalahan persentase absolut 3%. Penelitian ini menyimpulkan bahwa teknik <i>machine learning</i> dapat secara efektif memprediksi masalah finansial di bank-bank pedesaan, memberikan panduan bagi pengelola bank dan regulator untuk mengambil tindakan pencegahan yang tepat.

Jurnal 9

Nama Penulis	Resky Awalia, Farida Titik Kristanti
Tahun Penelitian	2023
Judul Penelitian	Prediksi Financial Distress Menggunakan <i>Artificial Neural Network</i> (ANN) pada Perusahaan Perbankan yang Terdaftar di Bursa Efek Indonesia (BEI) Periode 2017-2021 [28]

Penerbit	Jurnal Kajian Pendidikan Ekonomi dan Ilmu Ekonomi
Permasalahan	Penelitian ini mengidentifikasi masalah financial distress yang dihadapi oleh perusahaan perbankan di Indonesia, yang ditunjukkan dengan penurunan nilai <i>Earnings Per Share</i> (EPS) dan meningkatnya risiko kebangkrutan.
Metode	Penelitian ini menggunakan model <i>Artificial Neural Network</i> (ANN) untuk memprediksi <i>financial distress</i> . Data yang digunakan terdiri dari rasio keuangan seperti <i>current ratio</i> , <i>return on asset</i> , <i>debt to asset ratio</i> , dan <i>total asset turnover</i> , yang diambil dari laporan keuangan perusahaan perbankan yang terdaftar di BEI selama periode 2017 hingga 2021.
Hasil dan Kesimpulan	Hasil penelitian menunjukkan bahwa model ANN berhasil mencapai akurasi 88% dalam memprediksi kondisi kesehatan finansial bank. Penelitian ini menyimpulkan bahwa penggunaan teknik machine learning, khususnya ANN, efektif dalam memprediksi financial distress, memberikan wawasan yang berguna bagi pengelola bank dan regulator untuk mengambil tindakan pencegahan yang tepat.
Jurnal 10	
Nama Penulis	Peter Appiahene, Yaw Marfo Missah, Ussiph Najim
Tahun Penelitian	2020
Penerbit	Advances in Fuzzy Systems
Judul Penelitian	<i>Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks</i> [29]
Permasalahan	Penelitian ini berfokus pada efisiensi operasional bank di Ghana, terutama dalam konteks krisis keuangan yang terjadi antara 2015 hingga 2018. Krisis ini menimbulkan kekhawatiran mengenai efisiensi bank dan keamanan dana nasabah. Dengan meningkatnya kebutuhan untuk menganalisis kinerja bank, penting bagi para pemangku kepentingan untuk mengidentifikasi penyebab ketidakefisiensian dalam operasi perbankan.
Metode	Penelitian menggabungkan <i>Data Envelopment Analysis</i> (DEA) dengan tiga algoritma <i>machine learning</i> <i>Decision Tree</i> (C5.0), <i>Random Forest</i> , dan <i>Neural Network</i> untuk memprediksi efisiensi 444 cabang bank di Ghana. DEA digunakan untuk menentukan kelas efisiensi (efisien/inefisien), sedangkan ketiga algoritma diuji menggunakan data latih 70% dan data uji 30% dengan berbagai metrik performa accuracy, kappa, p-value.

Hasil dan Kesimpulan	Hasil penelitian menunjukkan bahwa algoritma <i>Decision Tree</i> C5.0 memberikan model prediktif terbaik dengan akurasi 100% pada dataset uji, diikuti oleh <i>Random Forest</i> dengan akurasi 98,5%, dan <i>Neural Network</i> dengan akurasi 86,6%. Penelitian ini menyimpulkan bahwa bank di Ghana dapat memanfaatkan temuan ini untuk memprediksi efisiensi.
----------------------	--

Berdasarkan penelitian-penelitian terdahulu yang disajikan pada Tabel 2.1, terlihat bahwa berbagai pendekatan *machine learning* dan metode evaluasi telah banyak digunakan untuk menilai kesehatan serta risiko perbankan. Setelah pemaparan tiap studi, bagian ini merangkum kecenderungan umum, keunggulan algoritma, dan celah penelitian yang masih tersisa. Sejumlah penelitian menunjukkan bahwa algoritma *K-Means* dan *K-Medoids* mampu mengelompokkan bank berdasarkan rasio keuangan, dengan *K-Means* yang umumnya memberikan performa lebih baik pada validasi menggunakan *Davies–Bouldin Index* dan *Silhouette Coefficient* [21]. Selain itu, berbagai penelitian juga menunjukkan bahwa algoritma klasifikasi seperti *Random Forest*, *Decision Tree*, dan *Artificial Neural Network* (*ANN*) mampu memberikan tingkat akurasi tinggi dalam memprediksi risiko kebangkrutan, kredit bermasalah, maupun efisiensi operasional bank [19],[22]-[29]. Temuan tersebut menegaskan bahwa penggunaan teknik machine learning dapat menjadi alat yang efektif untuk mendukung pengambilan keputusan manajemen risiko dan evaluasi kesehatan lembaga keuangan.

Namun demikian, terdapat sejumlah keterbatasan yang sekaligus menjadi peluang pengembangan penelitian lebih lanjut. Pertama, sebagian besar studi sebelumnya hanya memanfaatkan rasio keuangan umum seperti *Return on Assets* (*ROA*), *Leverage*, dan rasio *profitabilitas*, tanpa mengacu secara spesifik pada kerangka *Risk Profile*, *Good Corporate Governance*, *Earnings*, dan *Capital* (*RGEC*) yang merupakan standar resmi penilaian tingkat kesehatan bank menurut Otoritas Jasa Keuangan (OJK). Oleh karena itu, penelitian ini memilih *RGEC* karena kerangka tersebut dinilai lebih komprehensif dalam merepresentasikan stabilitas, efisiensi, dan tata kelola bank secara menyeluruh, sehingga hasil analisis dapat lebih

relevan dengan konteks regulasi perbankan di Indonesia. Kedua, perbandingan langsung antara algoritma *K-Means* dan *K-Medoids* dalam konteks pengelompokan bank berbasis rasio RGEC masih sangat terbatas, sehingga efektivitas kedua algoritma tersebut dalam menangani data keuangan dengan potensi outlier dan variasi ekstrem perlu dieksplorasi lebih lanjut. Ketiga, sebagian besar penelitian terdahulu dilakukan pada sektor perbankan luar negeri, sedangkan kajian yang secara spesifik menganalisis Bank Perekonomian Rakyat (BPR) di Indonesia dengan pendekatan berbasis RGEC masih jarang ditemukan [19],[22]-[29].

Selain itu, penelitian-penelitian sebelumnya umumnya berhenti pada tahap segmentasi tanpa melanjutkan ke proses klasifikasi prediktif. Oleh karena itu, penelitian ini tidak hanya membandingkan performa algoritma *K-Means* dan *K-Medoids*, tetapi juga mengembangkan model *Random Forest* sebagai tahap klasifikasi untuk memprediksi tingkat kesehatan BPR berdasarkan hasil klasterisasi yang diperoleh. Pemilihan *Random Forest* didasarkan pada kemampuannya dalam menghasilkan prediksi yang akurat, stabil, serta kemampuannya mengidentifikasi variabel-variabel keuangan yang paling berpengaruh terhadap tingkat kesehatan bank. Terakhir, penelitian sebelumnya menggunakan data hingga periode 2019–2021, sehingga penelitian ini menyajikan analisis yang lebih terkini terkait kondisi kesehatan bank di era perekonomian saat ini [19],[22]-[29]. Dengan mempertimbangkan gap-gap tersebut, penelitian ini bertujuan untuk membandingkan algoritma *K-Means* dan *K-Medoids* dalam segmentasi tingkat kesehatan bank berdasarkan rasio RGEC, sekaligus memberikan kontribusi pada pengembangan literatur terkait evaluasi kesehatan bank di Indonesia.

2.2 Tinjauan Teori

2.2.1 Bank Perekonomian Rakyat

Bank Perekonomian Rakyat (BPR) merupakan salah satu jenis bank di Indonesia dengan lingkup usaha terbatas, yang berfokus pada penghimpunan dana masyarakat dan penyaluran kredit skala kecil, terutama kepada pelaku usaha mikro, kecil, dan menengah (UMKM). Perubahan nama dari Bank

Perkreditan Rakyat menjadi Bank Perekonomian Rakyat berdasarkan Undang-Undang Nomor 4 Tahun 2023 tentang Pengembangan dan Penguatan Sektor Keuangan (UU P2SK) dimaksudkan untuk memperkuat peran BPR dalam sistem keuangan nasional serta meningkatkan kepercayaan masyarakat[1][2].

Secara teoritis, intermediasi keuangan, BPR berfungsi sebagai perantara antara pihak yang memiliki surplus dana dan pihak yang membutuhkan pembiayaan, dengan fokus pada sektor usaha mikro, kecil, dan menengah (UMKM). Keterkaitan BPR dengan konsep keuangan inklusif atau *financial inclusion* tercermin dari kemampuannya menjangkau masyarakat dan pelaku usaha yang tidak terlayani secara optimal oleh bank umum, khususnya di wilayah pedesaan dan daerah [30][31]. Oleh karena itu, BPR berperan penting dalam memperkuat ekonomi kerakyatan, meningkatkan inklusi keuangan, serta menjaga stabilitas ekonomi lokal.

2.2.2 Penilaian Tingkat Kesehatan Bank

Penilaian tingkat kesehatan bank merupakan mekanisme untuk menilai kemampuan bank dalam menjalankan kegiatan usaha secara normal, memenuhi kewajiban kepada nasabah, serta menjaga likuiditas dan solvabilitas dalam menghadapi risiko. Otoritas Jasa Keuangan (OJK) melakukan penilaian ini sebagai bagian dari fungsi pengaturan dan pengawasan lembaga jasa keuangan sebagaimana diatur dalam Undang-Undang Nomor 21 Tahun 2011 tentang OJK [7][32][33].

Sebelumnya, penilaian kesehatan bank di Indonesia menggunakan pendekatan *Capital, Asset Quality, Management, Earnings, Liquidity* (CAMEL) yang berfokus pada aspek keuangan dan manajerial. Seiring meningkatnya kompleksitas risiko perbankan, pendekatan tersebut dinilai belum cukup komprehensif. Sebagai respon, diterapkan pendekatan *Risk Profile, Good Corporate Governance, Earnings, Capital* (RGEC), yang awalnya digunakan untuk bank umum melalui PBI No. 13/1/PBI/2011 dan kemudian diadaptasi untuk BPR melalui POJK No. 20/POJK.03/2014 serta SEOJK No. 03/SEOJK.03/2022 tentang Penilaian Tingkat Kesehatan BPR dan BPRS [34].

Seluruh rasio dan kriteria penilaian yang digunakan dalam mengacu pada ketentuan yang tercantum dalam Surat Edaran Otoritas Jasa Keuangan (SEOJK) No. 03/SEOJK.03/2022 tentang Penilaian Tingkat Kesehatan Bank Perkreditan Rakyat (BPR) dan Bank Pembiayaan Rakyat Syariah (BPRS), serta diperkuat dengan referensi dari berbagai jurnal ilmiah terkait. Ketentuan tersebut menjabarkan metode evaluasi tingkat kesehatan bank berdasarkan empat faktor utama, yaitu *Risk Profile*, *Good Corporate Governance*, *Earnings*, dan *Capital* (RGEC) :

1) *Risk Profile*

Pada aspek profil risiko, penilaian dilakukan untuk mengukur tingkat risiko inheren yang dihadapi bank serta efektivitas penerapan manajemen risiko dalam operasionalnya. Delapan jenis risiko menjadi perhatian, meliputi risiko kredit, pasar, likuiditas, operasional, hukum, strategis, kepatuhan, dan reputasi. Setiap risiko dinilai berdasarkan prinsip penilaian kesehatan bank dan dikelompokkan ke dalam lima peringkat, dari peringkat 1 (*low*) hingga peringkat 5 (*high*) [7]. Dalam penelitian ini, fokus profil risiko diarahkan pada dua indikator utama likuiditas, yaitu *Loan to Deposit Ratio* (LDR) dan *Non-Performing Loan* (NPL).

a) *Loan to Deposit Ratio (LDR)*

Loan to Deposit Ratio (LDR) digunakan untuk mengukur kemampuan bank dalam menyalurkan dana pihak ketiga (DPK) dalam bentuk kredit. Rasio yang terlalu tinggi mengindikasikan risiko likuiditas, sedangkan rasio yang terlalu rendah menunjukkan kurang optimalnya penyaluran kredit. OJK menetapkan kisaran LDR sehat umumnya berada pada interval 75%–85% dengan toleransi hingga 100% [35][36]. Matriks Kriteria Penetapan Peringkat LDR ditunjukkan pada tabel 2.2.

$$LDR = \frac{\text{Total Kredit}}{\text{Dana Pihak Ketiga}} \times 100\% \quad (2.1)$$

Tabel 2. 2 Matriks Kriteria Penetapan Peringkat LDR [36]

Peringkat	Bobot	Matriks
1	LDR < 75%	Sangat Sehat
2	LDR 75% - 85%	Sehat
3	LDR 85% - 100%	Cukup Sehat
4	LDR 100% - 120%	Kurang Sehat
5	LDR > 120%	Tidak Sehat

b) Non-Performing Loan (NPL)

Non-Performing Loan (NPL) menggambarkan proporsi kredit bermasalah terhadap total kredit yang disalurkan. Semakin tinggi rasio NPL, semakin besar potensi kerugian bank dan semakin rendah kualitas asset [35][36]. Matriks Kriteria Penetapan Peringkat NPL ditunjukkan pada Tabel 2.3.

$$NPL = \frac{\text{Kredit Bermasalah}}{\text{Total Kredit}} \times 100\% \quad (2.2)$$

Tabel 2. 3 Matriks Kriteria Penetapan Peringkat NPL [36]

Peringkat	Bobot	Matriks
1	NPL < 2%	Sangat Sehat
2	2% ≤ NPL < 3,5%	Sehat
3	3,5% ≤ NPL < 5%	Cukup Sehat
4	5% ≤ NPL < 8%	Kurang Sehat
5	NPL ≥ 8%	Tidak Sehat

2) Good Corporate Governance (GCG)

Good Corporate Governance (GCG) merupakan penilaian atas kualitas penerapan tata kelola bank yang baik dalam kegiatan operasional sehari-hari. Penilaian GCG dilakukan melalui *self-assessment* yang mencakup prinsip-prinsip transparansi, akuntabilitas, responsibilitas, independensi, dan kewajaran. Selanjutnya, Skor GCG tidak dihitung dengan rumus matematis, melainkan ditetapkan dalam bentuk peringkat, Namun digunakan untuk hasil penilaian dari manajemen bank [35]. Matriks Kriteria Peringkat GCG ditunjukkan pada tabel 2.4

Tabel 2. 4 Matriks Kriteria Peringkat GCG [35]

Peringkat	Bobot	Matriks
1	GCG <1,5	Sangat Sehat
2	GCG 1,5 ≤ GCG <2,5	Sehat
3	GCG 2,5 ≤ GCG <3,5	Cukup Sehat
4	GCG 3,5 ≤ GCG <4,5	Kurang Sehat
5	GCG 4,5 ≤ GCG <5	Tidak Sehat

3) Earnings

a) Biaya Operasional terhadap Pendapatan Operasional (BOPO)

Digunakan untuk mengukur tingkat efisiensi dan kemampuan bank dalam mengelola aktivitas operasionalnya. Semakin kecil rasio BOPO menunjukkan semakin efisien bank dalam menjalankan operasional. Sebaliknya, rasio yang besar menunjukkan tingginya biaya yang dikeluarkan dibandingkan pendapatan yang diperoleh [37]. Matriks Kriteria Penetapan Peringkat BOPO ditunjukkan pada tabel 2.5.

$$\text{BOPO} = \frac{\text{Beban Operasional}}{\text{Pendapatan Operasional}} \times 100\% \quad (2. 3)$$

Tabel 2. 5 Matriks Kriteria Penetapan Peringkat BOPO [37]

Peringkat	Bobot	Matriks
1	BOPO ≤ 83%	Sangat Sehat
2	83% < BOPO ≤ 85%	Sehat
3	85% < BOPO ≤ 87%	Cukup Sehat
4	87% < BOPO ≤ 89%	Kurang Sehat
5	BOPO > 89%	Tidak Sehat

b) Return on Assets (ROA)

ROA merupakan rasio yang digunakan untuk mengukur kemampuan bank dalam menghasilkan laba bersih dari total aset yang dimiliki. Rasio ini menjadi salah satu indikator utama profitabilitas bank, karena semakin besar nilai ROA menunjukkan semakin efektif bank dalam memanfaatkan asetnya untuk memperoleh keuntungan [36] [37]. Matriks Kriteria Penetapan Peringkat ROA ditunjukkan pada tabel 2.6.

$$ROA = \frac{Laba Sebelum Pajak}{Total Aset} \times 100\% \quad (2.4)$$

Tabel 2. 6 Matriks Kriteria Penetapan Peringkat ROA [36] [37].

Peringkat	Bobot	Matriks
1	ROA $\geq 1.5\%$	Sangat Sehat
2	$1.25\% < ROA \leq 1.5\%$	Sehat
3	$0.5\% < ROA \leq 1.25\%$	Cukup Sehat
4	$0\% < ROA \leq 0.5\%$	Kurang Sehat
5	ROA $\leq 0\%$	Tidak Sehat

4) Capital

a) Capital Adequacy Ratio (CAR)

CAR adalah rasio kecukupan modal yang digunakan untuk mengukur kemampuan bank dalam menutup risiko kerugian yang mungkin timbul dari penyaluran kredit atau kegiatan operasional lainnya. Rasio ini mencerminkan seberapa kuat permodalan bank dalam menjaga stabilitas dan keberlangsungan usaha, sekaligus sebagai indikator kepatuhan terhadap ketentuan minimum modal yang ditetapkan regulator [35] [36]. Matriks Kriteria Penetapan Peringkat CAR ditunjukkan pada tabel 2.7

$$CAR = \frac{Modal}{Aset Tertimbang Menurut Risiko} \times 100\% \quad (2.5)$$

Tabel 2. 7 Matriks Kriteria Penetapan Peringkat CAR [35][36]

Peringkat	Bobot	Matriks
1	CAR $\geq 11\%$	Sangat Sehat
2	$9.5\% \leq CAR < 11\%$	Sehat
3	$8\% \leq CAR \leq 9.5\%$	Cukup Sehat
4	$6.5\% \leq CAR \leq 8\%$	Kurang Sehat
5	CAR $< 6.5\%$	Tidak Sehat

2.2.3 Machine Learning

Machine Learning (ML) merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma yang mampu belajar dari data untuk melakukan prediksi atau pengambilan keputusan tanpa diprogram secara eksplisit [9][38]. ML bekerja dengan cara mengenali pola dari data historis untuk kemudian diaplikasikan pada data baru sehingga dapat menghasilkan output yang relevan. Secara umum, ML terbagi menjadi tiga kategori utama, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*, yang masing-masing memiliki pendekatan berbeda dalam proses pembelajaran data. Supervised learning digunakan untuk klasifikasi atau regresi dengan label yang sudah diketahui, sedangkan unsupervised learning digunakan untuk menemukan pola atau kelompok pada data yang belum terlabeli. Sementara itu, *reinforcement learning* berfokus pada proses pembelajaran berbasis pengalaman, di mana sebuah agen belajar melalui interaksi dengan lingkungannya untuk mencapai tujuan tertentu dengan memaksimalkan *reward* dan meminimalkan *punishment*. Pendekatan ini banyak digunakan pada sistem yang membutuhkan pengambilan keputusan berkelanjutan, seperti robotika, permainan (*game*) [39][40].

2.3 Framework/Algoritma yang digunakan

2.3.1 CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah framework yang sangat banyak digunakan dalam industri untuk mengelola siklus hidup proyek data mining dan analisis data karena pendekatannya yang terstruktur, dibangun oleh 5 perusahaan yaitu Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA, lalu dikembangkan melalui serangkaian *workshops* yang berlangsung antara tahun 1997 hingga 1999 [41]. CRISP-DM memiliki 6 fase yaitu:

1. *Business Understanding*

Fase awal yang berfokus pada pemahaman tujuan dan kebutuhan bisnis. Informasi yang tersedia digunakan untuk merumuskan masalah yang harus diselesaikan serta menyusun rencana aksi untuk mencapai tujuan.

2. *Data Understanding*

Tahap ini dimulai dengan pengumpulan data yang relevan, kemudian dilakukan pemeriksaan kualitas, pencarian wawasan penting, serta identifikasi potensi masalah. Informasi tersembunyi yang bermanfaat juga dapat terungkap pada fase ini.

3. *Data Preparation*

Proses pembersihan dan pengolahan dilakukan agar data mentah berubah menjadi data siap analisis. Hasil akhir pada tahap ini berupa data yang bebas dari kesalahan yang dapat memengaruhi keakuratan analisis.

4. *Modeling*

Berbagai teknik pemodelan diterapkan untuk membangun model yang sesuai dengan tujuan analisis. Parameter disesuaikan agar diperoleh performa yang optimal serta efisiensi komputasi.

5. *Evaluation*

Model yang dihasilkan melalui proses sebelumnya diuji lebih lanjut untuk memastikan kesesuaian dengan tujuan penelitian dan kemampuan menyelesaikan masalah. Evaluasi memberikan gambaran mengenai kelayakan penerapan model dalam konteks nyata.

6. Deployment

Tahap akhir berupa penerapan hasil analisis ke dalam sistem nyata, misalnya dalam bentuk aplikasi atau situs web, sehingga dapat dimanfaatkan secara praktis dalam kegiatan sehari-hari.

2.3.2 Standardisasi Data

Proses standardisasi dilakukan menggunakan StandardScaler untuk menyetarakan skala setiap fitur, karena algoritma berbasis jarak sangat sensitif terhadap perbedaan skala. Tanpa penyetaraan, variabel dengan rentang nilai besar dapat mendominasi perhitungan jarak dan menurunkan kualitas *clustering* [42]. Berikut merupakan rumus StandardScaler yang ditunjukkan pada rumus (2.6)–(2.8):

- 1) StandardScaler

$$z = \frac{x_i - \mu}{\sigma} \quad (2.6)$$

- 2) Rata-rata (Mean)

$$\text{Mean}(\mu) = \frac{1}{N} \sum_{i=1}^N (x_i) \quad (2.7)$$

- 3) Standar Deviasi (*Standard Deviation*)

$$(\sigma) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2.8)$$

2.3.3 K-means

Algoritma *K-Means* merupakan salah satu metode *clustering* yang paling umum digunakan dalam analisis data. Prinsip dasarnya adalah membagi sekumpulan data menjadi sejumlah k *cluster* dengan tujuan meminimalkan jumlah kuadrat jarak antara setiap titik data terhadap pusat *cluster* atau *centroid*. Proses optimisasi dilakukan secara iteratif melalui dua tahap utama, yaitu tahap

assignment di mana setiap data ditempatkan pada *cluster* dengan jarak terdekat terhadap *centroid*, dan tahap update *centroid* di mana posisi *centroid* diperbarui berdasarkan rata-rata posisi seluruh anggota *cluster* hingga proses konvergen atau perubahan antar iterasi menjadi sangat kecil [43][44]. Parameter penting yang mempengaruhi performa *K-Means* meliputi pemilihan jumlah *cluster* (k) dan inisialisasi *centroid*, karena keduanya menentukan stabilitas hasil. Algoritma ini unggul dalam kecepatan komputasi dan efisiensi pada dataset besar, namun memiliki keterbatasan seperti sensitivitas terhadap *outlier*, variasi hasil akibat inisialisasi awal, dan asumsi bahwa *cluster* berbentuk konveks dan memerlukan penentuan jumlah *cluster* secara manual sebelum proses dijalankan [45].

2.3.4 *K-medoids*

Algoritma *K-Medoids* merupakan varian dari *K-Means* yang bekerja dengan prinsip serupa, tetapi menggunakan *medoid* sebagai pusat *cluster*, yaitu titik data aktual dengan total jarak minimum terhadap seluruh anggota *cluster*. Pendekatan berbasis *medoid* ini membuat *K-Medoids* lebih *robust* terhadap *outlier*, karena pusat *cluster* tidak dipengaruhi oleh nilai ekstrem. Tujuan utama algoritma adalah meminimalkan total jarak absolut antara anggota *cluster* dan medoidnya. [46]. Parameter penting dalam *K-Medoids* meliputi jumlah *cluster* (k) dan metode inisialisasi, yaitu *k-medoids++*. Kelebihan utamanya terletak pada stabilitas hasil *clustering* pada data yang memiliki *outlier* atau distribusi tidak normal, karena pusat *cluster* merupakan titik nyata. Namun, algoritma ini memiliki kompleksitas komputasi lebih tinggi dibandingkan *K-Means* akibat proses pemilihan dan pembaruan *medoid* yang memerlukan iterasi intensif, terutama pada dataset berukuran besar. Beberapa parameter penting yang perlu ditentukan dalam algoritma ini antara lain jumlah *cluster* (k), metode inisialisasi (seperti *k-medoids++*). Kelebihan utama *K-Medoids* adalah kemampuannya dalam menghasilkan pembagian *cluster* yang lebih stabil pada data yang mengandung *outlier* atau distribusi yang tidak normal, karena pusat *cluster* selalu berupa titik data nyata. Namun demikian, kelemahan utama algoritma ini adalah kompleksitas

komputasi yang lebih tinggi dibandingkan *K-Means*, karena proses pemilihan dan pembaruan *medoid* memerlukan iterasi yang lebih intensif terutama pada dataset berukuran besar [46].

2.3.5 Davies Bouldin Index

Davies Bouldin Index (DBI) merupakan salah satu metode yang digunakan untuk melakukan validasi hasil dari algoritma *clustering*. Validasi dilakukan dengan mempertimbangkan dua aspek utama, yaitu tingkat kedekatan data dengan pusat *cluster* (kohesi) serta jarak antar pusat *cluster* (separasi). *Cluster* yang baik dicirikan dengan nilai kohesi yang rendah serta nilai separasi yang tinggi. Penentuan jumlah *cluster* optimal pada DBI didasarkan pada nilai indeks yang mendekati angka 0. Semakin kecil nilai DBI, semakin baik kualitas *cluster* yang dihasilkan karena menunjukkan pengelompokan yang lebih tepat dan optimal [15][47].

2.3.6 Silhouette Coefficient

Silhouette Score merupakan salah satu metode yang digunakan untuk mengevaluasi kualitas hasil *clustering*. Evaluasi dilakukan dengan mengukur seberapa dekat suatu data terhadap *cluster* tempatnya berada (kohesi) serta seberapa jauh jarak data tersebut dengan *cluster* lain yang paling dekat (separasi). Nilai *Silhouette Score* berada pada rentang -1 hingga 1. Semakin mendekati angka 1, maka semakin baik kualitas *clustering* yang dihasilkan karena menunjukkan bahwa data terkelompok dengan tepat. Sebaliknya, nilai mendekati 0 menunjukkan data berada di batas dua *cluster*, sementara nilai negatif mengindikasikan data ditempatkan pada *cluster* yang salah. Dengan demikian, *Silhouette Score* membantu dalam menentukan jumlah *cluster* optimal yang memberikan pengelompokan paling sesuai [15][48].

2.3.7 Random Forest

Random Forest merupakan algoritma *ensemble learning* yang membangun banyak pohon keputusan dari sampel data dan subset fitur yang dipilih secara acak. Setiap pohon memberikan prediksi, kemudian hasilnya digabungkan

melalui voting untuk klasifikasi atau rata-rata untuk regresi. Mekanisme ini efektif dalam mengurangi varians model, meminimalkan risiko *overfitting*, dan meningkatkan akurasi secara konsisten. Algoritma ini juga mampu menangani data berdimensi tinggi serta menyediakan pengukuran *feature importance* untuk menilai kontribusi masing-masing variabel. Dengan karakteristik yang stabil dan fleksibel, *Random Forest* banyak digunakan dalam berbagai bidang seperti keuangan, kesehatan, dan analitik sosial. [16][49].

2.3.8 Evaluasi Random Forest

Evaluasi model klasifikasi dilakukan untuk menilai akurasi dan konsistensi prediksi model terhadap data uji. Setiap metrik memberikan perspektif berbeda mengenai kualitas klasifikasi, baik dari sisi ketepatan prediksi maupun kemampuan model membedakan tiap kelas [50][51]. Berikut penjelasannya:

1) Confusion Matrix

Confusion Matrix merupakan alat evaluasi yang menampilkan hasil perbandingan antara label aktual dengan label yang diprediksi oleh model dalam bentuk tabel dua dimensi. Matriks ini memberikan informasi detail tentang jumlah prediksi yang benar dan salah untuk setiap kelas, sehingga memudahkan dalam menganalisis kesalahan klasifikasi. *Confusion Matrix* menjadi dasar untuk menghitung berbagai metrik turunan seperti *accuracy*, *precision*, *recall*, dan *F1-Score* [50][51].

2) Accuracy (Akurasi)

Akurasi mengukur proporsi prediksi yang benar terhadap seluruh data uji. Metrik ini paling umum digunakan untuk menilai kinerja keseluruhan model.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.9)$$

di mana TP (*True Positive*) dan TN (*True Negative*) adalah jumlah prediksi benar untuk kelas positif dan negatif, sedangkan FP (*False Positive*) dan FN (*False Negative*) adalah jumlah kesalahan prediksi. Akurasi yang tinggi menunjukkan kemampuan model untuk memprediksi dengan benar sebagian besar data[50][51]

3) *Precision* (Presisi)

Presisi mengukur sejauh mana prediksi positif model benar-benar relevan. Metrik ini penting untuk menghindari *false positive*, terutama pada kasus prediksi risiko keuangan atau kesehatan bank.

$$Precision = \frac{TP}{TP+FP} \quad (2. 10)$$

Nilai presisi yang tinggi menunjukkan bahwa model jarang salah dalam mengklasifikasikan data negatif sebagai positif [50][51]

4) *Recall* (*Sensitivitas* atau *True Positive Rate*)

Recall menunjukkan kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk dalam kelas positif.

$$Recall = \frac{TP}{TP+FN} \quad (2. 11)$$

Nilai *recall* tinggi berarti model jarang gagal mendeteksi data positif yang sebenarnya, sehingga cocok untuk sistem yang menuntut deteksi risiko secara komprehensif [50][51].

5) *F1-Score*

F1-Score merupakan rata-rata harmonik antara *Precision* dan *Recall*, digunakan ketika terdapat ketidakseimbangan data antar kelas.

$$F1-Score = 2x \frac{Precision \times Recall}{Precision + Recall} \quad (2. 12)$$

Nilai *F1-Score* tinggi menunjukkan keseimbangan yang baik antara kemampuan mendeteksi kelas positif (*Recall*) dan keakuratan hasil prediksi (*Precision*) [50][51].

6) *Area Under Curve (AUC) - Receiver Operating Characteristic (ROC)*

ROC-AUC mengukur kemampuan model untuk membedakan antara kelas positif dan negatif dengan menghitung luas area di bawah kurva ROC. Nilai AUC berkisar antara 0 hingga 1, di mana nilai mendekati 1 menandakan bahwa model memiliki kemampuan diskriminatif yang sangat baik terhadap kedua kelas. *ROC-AUC* sering digunakan untuk mengevaluasi performa model klasifikasi multikategori maupun biner [50][51].

2.4 Tools/software yang digunakan

2.4.1 *Jupyter Notebook*

Jupyter Notebook merupakan perangkat lunak interaktif yang banyak digunakan dalam *data science* untuk menggabungkan kode, teks, dan visualisasi dalam satu dokumen. Platform ini memudahkan proses eksplorasi data karena hasil komputasi dapat ditampilkan secara langsung tanpa menjalankan ulang seluruh skrip. Fitur interaktif tersebut mendukung proses analisis, eksperimen, dan pengembangan model secara lebih efisien. *Jupyter Notebook* juga menyediakan dukungan untuk berbagai bahasa pemrograman melalui sistem kernel. Dengan fleksibilitas tersebut, platform ini menjadi alat yang umum digunakan dalam penelitian dan pengembangan berbasis machine learning [52].

2.4.2 *Python*

Python merupakan bahasa pemrograman tingkat tinggi yang diperkenalkan oleh Guido van Rossum dan dikenal karena sintaksnya yang sederhana serta mudah dipahami. Bahasa ini mendukung berbagai paradigma pemrograman seperti objektif, fungsional, dan imperatif sehingga memberikan fleksibilitas dalam pengembangan kode. Popularitas *python* didorong oleh *library* yang lengkap dan *framework* yang luas serta komunitas pengguna yang sangat aktif. Python banyak digunakan di berbagai bidang, termasuk analisis data, pengembangan web, kecerdasan buatan, dan *machine learning*. Dengan

fleksibilitas tersebut, Python menjadi salah satu bahasa utama dalam pembangunan aplikasi modern dan penelitian berbasis data. [53].

2.4.3 Streamlit

Streamlit merupakan *framework open-source* berbasis *Python* yang digunakan untuk pembuatan aplikasi *web* interaktif secara cepat tanpa perlu pengetahuan *HTML*, *CSS*, atau *JavaScript*. *Framework* ini bekerja dengan konsep *reactive programming*, sehingga setiap perubahan input pengguna langsung memicu pembaruan tampilan. Mekanisme ini membuat proses penyajian output analisis dan prediksi menjadi lebih efisien dan responsif. *Streamlit* sangat tepat digunakan untuk *deployment* model *machine learning* karena mampu mengintegrasikan pemrosesan data, model, dan visualisasi dalam satu *workflow*. Dengan kesederhanaan implementasinya, *Streamlit* mendukung pengembangan aplikasi analitik yang praktis dan mudah diakses oleh pengguna akhir. [54].

