

BAB III

METODE PENELITIAN

3.1. Metode Penelitian

Penelitian kuantitatif merupakan suatu metode yang berfokus pada pengumpulan, analisis, dan interpretasi data numerik dengan tujuan mengukur dan menarik kesimpulan dari berbagai sampel yang digunakan. Metode ini dirancang untuk menjawab pertanyaan yang bersifat kuantitatif seperti “berapa lama,” “seberapa banyak,” dan “sejauh mana” dalam upaya untuk menguji hipotesis yang telah dirumuskan [36].

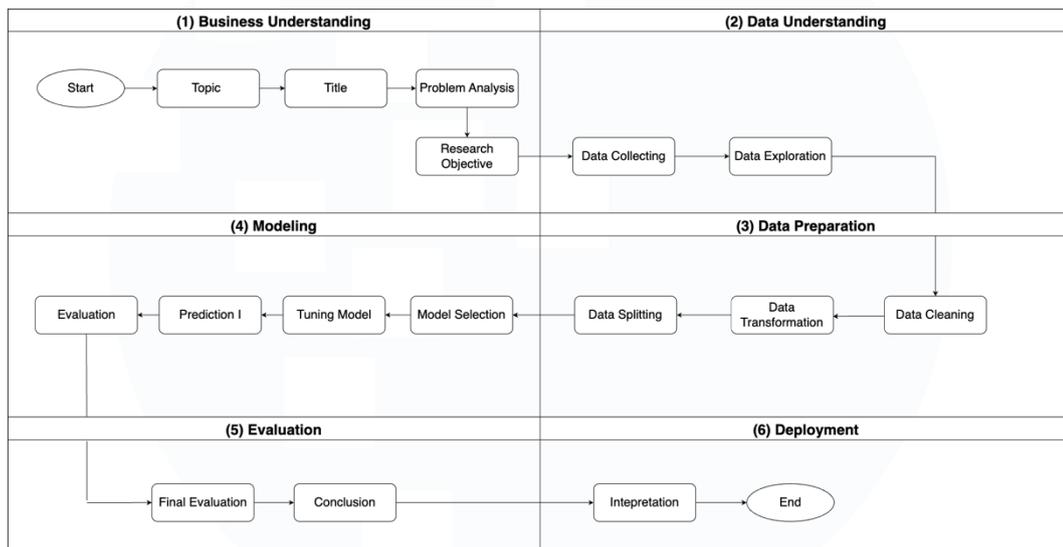
Pendekatan kuantitatif mengedepankan proses yang logis dan sistematis, dengan orientasi utama pada pengujian teori yang didasarkan pada prinsip-prinsip empirisme dan positivisme. Proses ini mengharuskan adanya pengumpulan data menggunakan instrumen yang terstandarisasi, seperti survei, kuesioner, dan ujian untuk memperoleh data numerik dari sampel yang besar. Reliabilitas data juga dijaga tinggi dengan melakukan perencanaan yang matang sebelum pengumpulan data dilakukan [37].

Dalam penelitian kuantitatif, analisis data dilakukan dengan menggunakan teknik statistik untuk mengidentifikasi pola, hubungan, atau perbedaan antar kelompok data yang dianalisis. Pengujian hipotesis menjadi bagian penting dalam menganalisis data yang dikumpulkan, dengan tujuan untuk mengevaluasi apakah hipotesis yang diajukan dapat diterima atau harus ditolak. Dengan demikian, hasil penelitian ini bersifat objektif dan dapat diuji kembali melalui prosedur yang sama untuk memastikan konsistensinya dalam penelitian lainnya [38].

Hasil analisis tersebut kemudian diinterpretasikan untuk memberikan pemahaman yang lebih mendalam mengenai fenomena yang diteliti, serta untuk mengevaluasi apakah hipotesis yang diajukan dapat diterima atau harus ditolak. Dengan demikian, penelitian kuantitatif memberikan hasil yang bersifat objektif dan dapat diuji kembali melalui prosedur yang sama [36] [37] [38].

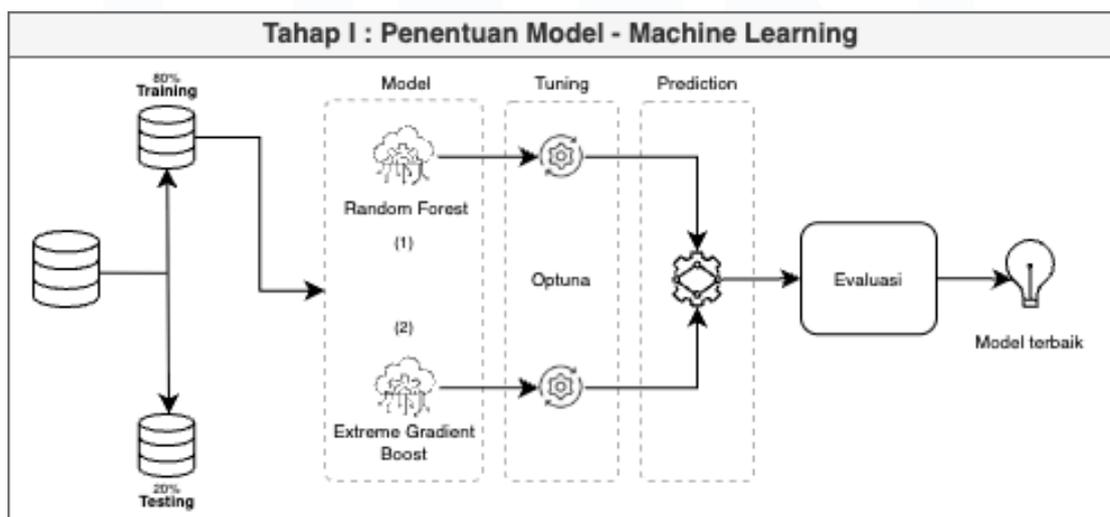
3.1.1 Alur Penelitian

Gambar 3.1 menyajikan alur penelitian dalam bentuk diagram flowchart. Dengan visualisasi alur penelitian ini, diharapkan pemahaman mengenai proses penelitian secara menyeluruh (end-to-end) dapat dipahami dengan jelas.

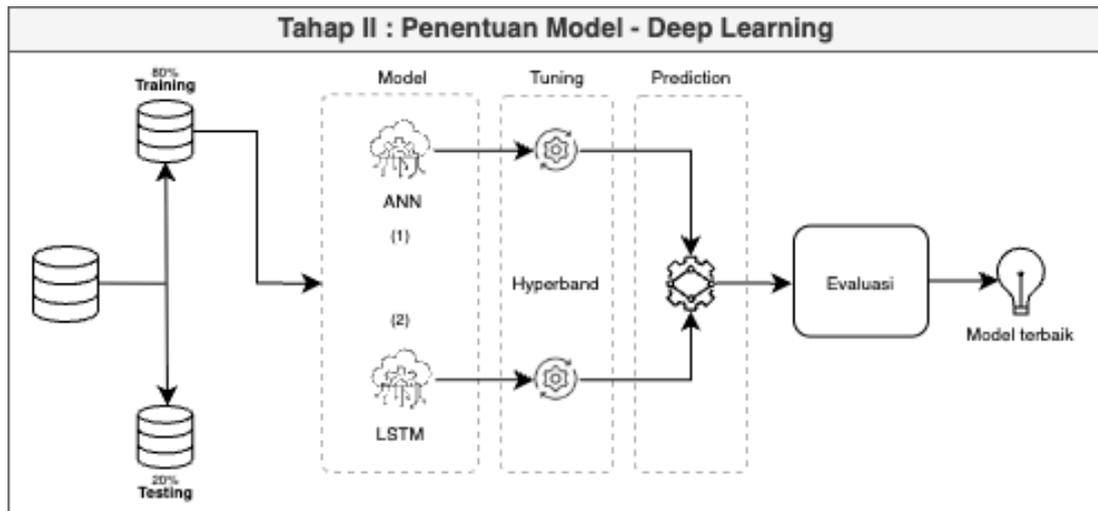


Gambar 3. 1 Alur Penelitian

Gambar 3.2 Kerangka Tahap I Proses Pemilihan Model Machine Learning dan Gambar 3.3 dibawah akan menampilkan Kerangka Tahap II Proses Pemilihan Model Deep Learning, dari situ model yang terbaik akan digunakan untuk proses deployment nanti



Gambar 3. 2 Kerangka Penentuan Model - Machine Learning



Gambar 3. 3 Kerangka Penentuan Model - Deep Learning

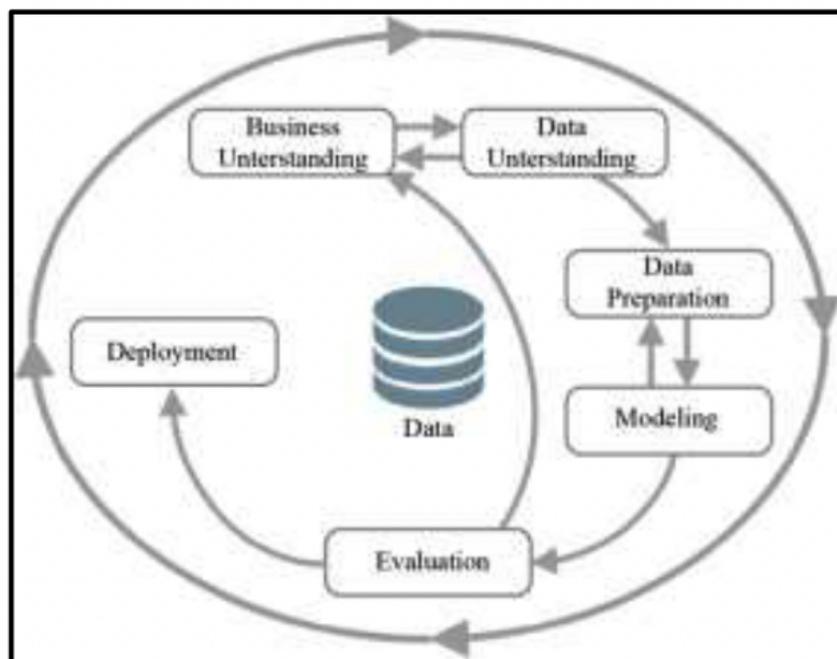
3.1.2 Kerangka Kerja Data Mining dan Data Analisis

Dalam penelitian ini, jenis kerangka yang akan diterapkan, yaitu kerangka kerja Data Mining dan. Berikut adalah tabel yang membandingkan kerangka kerja: Data Mining, SEMMA dan CRISP-DM.

Tabel 3. 1 Tabel Penentuan Metode Penelitian

Indikator	SEMMA [39]	CRISP-DM [40], [41], [42]
Siklus	<ol style="list-style-type: none"> 1. Sample 2. Explore 3. Modify 4. Model 5. Assessment 	<ol style="list-style-type: none"> 1. Business Understanding 2. Data Understanding 3. Data Preparation 4. Modeling 5. Evaluation 6. Deployment
Kelebihan	<ul style="list-style-type: none"> • Berfokus pada manajemen data dan pemodelan 	<ul style="list-style-type: none"> • Meliputi seluruh siklus hidup proyek data science, mulai dari memahami masalah bisnis hingga implementasi, sehingga cocok untuk berbagai jenis proyek.
Kekurangan	<ul style="list-style-type: none"> • Kurang memperhatikan masalah dalam tahap implementasi 	<ul style="list-style-type: none"> • Struktur yang sangat fleksibel membuatnya sangat cocok untuk digunakan berbagai macam proyek

Berdasarkan tabel perbandingan yang ada, dapat disimpulkan bahwa metodologi CRISP-DM adalah pilihan yang lebih efektif dan sesuai untuk digunakan dalam penelitian ini, dibandingkan dengan SEMMA. Hal ini disebabkan oleh fakta bahwa CRISP-DM mencakup seluruh siklus hidup proyek data science, mulai dari memahami masalah bisnis hingga implementasi, yang membuatnya lebih komprehensif dan lebih cocok untuk berbagai jenis proyek. Sebaliknya, SEMMA lebih berfokus pada manajemen data dan pemodelan, tetapi memiliki kekurangan dalam hal implementasi yang tidak terlalu diperhatikan pada tahap akhir proyek.



Gambar 3. 4 Metodologi CRISP-DM

Sumber: [42]

Penerapan metodologi CRISP-DM dalam penelitian ini juga didukung oleh literatur yang telah dibahas sebelumnya, seperti yang dijelaskan oleh [39] [40] dan [41]. Sementara itu, SEMMA yang dijelaskan oleh [40] dan [42] meskipun efektif dalam tahap eksplorasi dan pemodelan data, tidak sepenuhnya mendukung tahap implementasi yang sangat penting dalam konteks penelitian ini.

3.2 Tahapan Penelitian

Penelitian ini melibatkan beberapa tahapan yang dimulai dari *perencanaan* hingga *evaluasi*. Proses ini bertujuan untuk membandingkan beberapa model yang menggunakan *Machine Learning* dan *Deep Learning*, dan memilih model terbaik untuk digunakan dalam tahap *deployment*. Berikut adalah tahapan yang dijalani dalam penelitian ini:

3.2.1 Business Understanding

Tahapan awal ini melibatkan pemahaman mendalam terhadap situasi penelitian yang relevan. Pada fase ini, penting untuk menetapkan tujuan penelitian yang jelas, seperti jenis data mining yang digunakan (misalnya regresi) serta kriteria keberhasilan data mining, seperti metrik evaluasi yang ingin dicapai. Selain itu, penelitian ini juga bertujuan untuk memahami akar permasalahan dalam konteks logam mulia yang menjadi topik penelitian.

3.2.2 Data Preparation

Pada fase ini, peneliti mengumpulkan data dari sumber yang terpercaya dan melakukan eksplorasi data. Visualisasi data digunakan untuk memberi pemahaman yang lebih baik mengenai atribut dan pola yang ada. Data yang digunakan dalam penelitian ini dikumpulkan melalui hasil yang telah dikumpulkan pada Jupyter Notebook untuk memberikan gambaran yang jelas tentang data yang digunakan dalam analisis.

3.2.3 Modeling

Pada tahap modeling, beberapa model Machine Learning dan Deep Learning akan diterapkan, dengan tujuan untuk menguji dan membandingkan kinerja masing-masing model. Model yang digunakan dalam penelitian ini adalah ANN dan LSTM untuk deep learning sementara XGBoost dan Random Forest akan digunakan untuk Model-model ini akan disesuaikan menggunakan teknik **Hyperband** dan **Optuna** untuk optimasi parameter. Kinerja model akan dievaluasi menggunakan metrik seperti **RMSE**, **MAE**, dan **MAPE**. Penelitian ini hanya bertujuan untuk membandingkan model yang diuji, dan bukan untuk

menciptakan model hibrid. Model dengan performa terbaik akan dipilih untuk langkah deployment.

3.2.4 Evaluation

Pada tahap modeling, beberapa model Machine Learning dan Deep Learning akan diterapkan, dengan tujuan untuk menguji dan membandingkan kinerja masing-masing model. Model yang digunakan dalam penelitian ini adalah ANN dan LSTM Model-model ini akan disesuaikan menggunakan teknik **Hyperband** dan **Optuna** untuk optimasi parameter. Kinerja model akan dievaluasi menggunakan metrik seperti **RMSE**, **MAE**, dan **MAPE**. Penelitian ini hanya bertujuan untuk membandingkan model yang diuji, dan bukan untuk menciptakan model hibrid. Model dengan performa terbaik akan dipilih untuk langkah deployment.

3.2.5 Deployment

Setelah memilih model terbaik berdasarkan hasil evaluasi, model tersebut akan dipersiapkan untuk **deployment**. Pada tahap ini, model yang terpilih akan diterapkan untuk melakukan prediksi sesuai dengan tujuan yang telah ditetapkan di awal penelitian. Proses deployment mencakup langkah-langkah seperti pengujian akhir, penerapan di sistem produksi, serta monitoring kinerja model. Model yang diterapkan akan terus dipantau untuk memastikan kinerjanya tetap optimal.

3.3 Teknik Pengumpulan Data

Periode pengumpulan data dalam penelitian ini terdiri dari dua sumber utama. Pertama, data historis biodata mahasiswa Universitas Multimedia Nusantara yang telah dikumpulkan sejak tahun akademik 2019 hingga akhir 2024. Kedua, data diperoleh melalui survei daring menggunakan *Google Form* yang dipublikasikan pada 15 April 2025 dan kembali disebar pada 10 Mei 2025 untuk menjangkau responden tambahan. Tautan form yang digunakan adalah <https://bit.ly/formulir-biodata-akademik-umn>. Meskipun seluruh data yang terkumpul dicatat, hanya data

yang paling relevan dan memenuhi kriteria kualitas yang akan dipilih untuk dianalisis lebih lanjut dalam proses pemodelan prediktif.

3.3.1 Populasi dan Sampel

Dalam penelitian ini, purposive sampling digunakan sebagai prosedur pemilihan sampel, yang sering diterapkan dalam berbagai paradigma penelitian untuk memastikan pemilihan sampel yang berkualitas tanpa bias. Pendekatan ini meningkatkan keandalan dan kredibilitas temuan penelitian [38]. Purposive sampling memilih kasus-kasus yang kaya informasi untuk analisis mendalam. Ukuran sampel dan pemilihan kasus ditentukan berdasarkan tujuan penelitian.

Populasi dalam penelitian ini mencakup seluruh mahasiswa aktif Universitas Multimedia Nusantara dari tahun akademik 2019 hingga 2025. Sampel yang digunakan adalah mahasiswa yang telah mengisi *Google Form* biodata akademik, yang memuat informasi terkait profil demografis, data akademik, dan capaian pembelajaran. Pemilihan sampel mempertimbangkan kelengkapan data serta kesesuaian terhadap fokus penelitian, yakni prediksi kinerja akademik berdasarkan variabel numerik dan demografis.

Tabel 3. 2 Tabel Periode Jangka Waktu Data

Jangka Waktu Data	Rank	Sumber	Nama Data
Periode 2019-2024	1	Data historis dari sistem internal UMN	Mahasiswa UMN
Periode 2025	2	Google Form Survey	Survei tambahan untuk melengkapi/memverifikasi data

Tabel 3.2 merangkum dua sumber data utama dalam penelitian ini. Data pertama merupakan biodata mahasiswa UMN dari tahun akademik 2019 hingga

2024 yang telah dikumpulkan sebelumnya. Data kedua diperoleh dari survei tambahan yang disebarakan melalui Google Form pada 10 Mei 2025 untuk memperbarui atau melengkapi informasi yang diperlukan dalam analisis.

3.3.2 Periode Pengambilan Data

Periode pengambilan data dalam penelitian ini dilakukan dengan menggunakan survey yang disebarakan kepada mahasiswa UMN, serta dengan memanfaatkan data biodata mahasiswa UMN yang sudah dikumpulkan dari periode sebelumnya melalui google form. Data yang digunakan mencakup informasi akademik, demografi, dan prestasi belajar mahasiswa. Data tersebut akan digunakan untuk analisis dan prediksi berdasarkan variabel-variabel yang relevan.

Data yang dikumpulkan akan digunakan dalam analisis prediktif untuk memproyeksikan kinerja akademik mahasiswa berdasarkan data yang sudah ada. Jangka waktu pengumpulan data adalah periode akademik 2019 - 2023, dengan biodata mahasiswa UMN yang mencakup informasi kinerja akademik dan demografi mahasiswa yang telah dikumpulkan.

3.4 Variabel Penelitian

3.4.1 Variabel Independent

Kategori pertama terdiri atas data historis aktivitas dan kehadiran mahasiswa, yang menjadi dasar dalam menganalisis dan memprediksi tingkat keberhasilan akademik. Data ini mencakup metrik seperti kehadiran, nilai matakuliah.

Kategori kedua adalah data yang digunakan sebagai input untuk proses pemodelan menggunakan empat algoritma, yaitu dua algoritma deep learning Long Short-Term Memory dan Artificial Neural Network serta dua algoritma machine learning XGBoost dan Random Forest. Keempat model ini akan dianalisis dan dibandingkan untuk menentukan metode terbaik dalam memprediksi kinerja akademik mahasiswa berdasarkan data kehadiran dan aktivitas tersebut.

3.4.2 Variabel Independent

Variabel dependen dalam penelitian ini adalah tingkat keberhasilan akademik mahasiswa, yang direpresentasikan dalam bentuk nilai akhir suatu mata kuliah tertentu atau Indeks Prestasi Semester. Nilai ini merupakan output yang diprediksi oleh masing-masing algoritma berdasarkan input dari variabel-variabel independen. Hasil dari keempat model akan dibandingkan untuk menentukan model prediksi yang paling akurat dan efisien dalam konteks data mahasiswa UMN.

3.5 Teknik Analisis Data

Penelitian ini menerapkan teknik analisis data berbasis data mining dan machine learning, dengan tahapan utama meliputi pra-pemrosesan data eksplorasi data awal pembangunan model prediksi, serta evaluasi dan optimasi model. Proses ini mencakup transformasi data mentah menjadi bentuk yang siap dianalisis, identifikasi pola atau korelasi dalam data, serta penerapan algoritma pembelajaran mesin dan pembelajaran mendalam untuk membangun model yang mampu memprediksi performa akademik mahasiswa secara akurat.

Bahasa pemrograman yang digunakan dalam penelitian ini dipilih untuk mendukung proses data mining dan perancangan model. Dari berbagai pilihan yang tersedia, penelitian ini akan memfokuskan penggunaan pada dua bahasa pemrograman, yaitu Python dan R. Berikut ini adalah perbandingan antara kedua bahasa pemrograman yang akan diterapkan dalam penelitian ini.

Tabel 3. 3 Tabel Pemilihan Bahasa Pemrograman

Indikator	Python [33]	R [33]
Kelebihan	<ul style="list-style-type: none"> • Serbaguna untuk berbagai proyek • komunitas besar • Memiliki berbagai Pustaka luas seperti Pandas, Scikit-learn 	<ul style="list-style-type: none"> • Unggul dalam analisis statistik • Banyak pustaka untuk analisis data • Cocok untuk eksplorasi pada tahap awal
Kekurangan	<ul style="list-style-type: none"> • Tidak memiliki repositori pustaka umum • Sulit mencari fungsi spesifik • Kompleks mengintegrasikan pustaka tertentu 	<ul style="list-style-type: none"> • Sulit dipelajari bagi pemula • Indeks vektor mulai dari satu • Sintaks beberapa fungsi tidak selalu jelas

Dari tabel di atas, dapat dilihat bahwa Python memiliki kelebihan dalam hal fleksibilitas dan kompatibilitas dengan berbagai proyek, terutama yang melibatkan data science dan pengembangan berbasis machine learning, dengan dukungan pustaka seperti Pandas dan Scikit-learn. Selain itu, Python memiliki komunitas besar yang memberikan dukungan banyak bagi pengguna, serta kemudahan penggunaan melalui Jupyter Notebook, yang digunakan dalam penelitian ini untuk memfasilitasi proses analisis data.

Namun, R unggul dalam analisis statistik yang mendalam dan eksplorasi data, menjadikannya pilihan utama untuk proyek yang membutuhkan pemodelan statistik dan visualisasi data yang mendalam. Walaupun demikian, R sering dianggap sulit dipelajari bagi pemula, dengan sintaks yang unik dan beberapa fungsi yang tidak selalu intuitif.

Dari perbandingan tersebut, dapat disimpulkan bahwa Python lebih cocok digunakan dalam penelitian ini karena kemudahan dalam pengolahan data, integrasi pustaka machine learning, serta kemampuan pengolahan data yang lebih efisien. Penggunaan Python dalam penelitian ini didukung oleh kemampuan bahasa tersebut untuk menangani proses data besar dengan pustaka kuat seperti Pandas dan Scikit-learn. Jurnal terbaru yang membahas perbandingan kedua bahasa pemrograman ini menunjukkan bahwa Python lebih unggul dalam analisis data besar dan machine learning, sementara R tetap menjadi pilihan utama untuk statistik mendalam dan visualisasi data. Oleh karena itu, dalam penelitian ini, Python dipilih sebagai bahasa pemrograman utama untuk implementasi model [33].

3.6 Teknik Tuning Model

3.6.1 Teknik I Optuna Tuning – Machine Learning

Pada Tahap I, Optuna dipilih sebagai teknik untuk melakukan tuning pada model Machine Learning seperti XGBoost dan Random Forest, yang akan digunakan untuk memilih model terbaik dalam prediksi kinerja akademik mahasiswa UMN. Alasan pemilihan Optuna terletak pada fleksibilitas dan kemampuan modularnya dalam mengoptimalkan hyperparameter, khususnya pada XGBoost dan Random Forest. Dengan adanya API define-by-run, teknik ini memfasilitasi integrasi yang lebih mudah selama pengembangan [15].

Pendekatannya yang selektif dan adaptif, bersama dengan efisiensinya dalam mengeliminasi trial yang tidak memberikan hasil yang diinginkan, memungkinkan pengelolaan sumber daya dan waktu yang lebih optimal. Gabungan antara optimasi Gradien dan Bayesian Optimization memungkinkan Optuna untuk secara efisien menyediakan pengaturan model yang optimal dan efektif [15]. Fleksibilitas Optuna dalam menyesuaikan pengaturan hyperparameter secara dinamis menjadikannya pilihan ideal dalam penelitian ini, yang berfokus pada model untuk prediksi kinerja mahasiswa UMN.

3.6.2 Teknik II Hyperband Tuning – Deep Learning

Pada tahap II, Hyperband dipilih sebagai teknik untuk melakukan tuning pada model deep learning, seperti LSTM dan ANN, yang digunakan untuk memprediksi nilai akademik mahasiswa berdasarkan kehadiran mereka. Hyperband dipilih karena kemampuannya dalam mengoptimalkan proses tuning model deep learning yang memerlukan sumber daya besar dan waktu pelatihan yang cukup lama [31]. Dengan pendekatan berbasis bandit, Hyperband dapat secara adaptif menyaring konfigurasi yang tidak efektif lebih awal, menghemat sumber daya dan waktu pelatihan. Teknik ini memungkinkan penyaringan konfigurasi yang tidak layak atau tidak sesuai, sekaligus mengalokasikan lebih banyak sumber daya ke konfigurasi yang menjanjikan. Pemilihan Hyperband pada tahap II didasarkan pada kemampuannya untuk menangani proses pelatihan model deep learning yang lebih kompleks, seperti LSTM dan ANN. Dengan kecepatan tinggi dan kemampuan adaptif dalam menyempurnakan model, Hyperband memungkinkan pengaturan model yang optimal untuk memaksimalkan akurasi prediksi nilai mahasiswa [31]. Pendekatannya yang efisien menjadikan Hyperband pilihan ideal untuk tahap II ini, di mana model deep learning akan lebih efektif dalam mempelajari pola dari data kehadiran mahasiswa. Pemilihan kedua teknik ini—Optuna pada Tahap I untuk model machine learning, dan Hyperband pada Tahap II untuk model deep learning didasarkan pada kemampuan adaptif keduanya dalam mengoptimalkan dan menyesuaikan model. Integrasi kedua teknik tuning ini diharapkan akan menghasilkan model yang lebih efisien dan performatif, dengan masing-masing teknik berfokus pada keunggulan yang berbeda untuk meningkatkan akurasi dan kecepatan dalam memprediksi kinerja akademik mahasiswa.

UNIVERSITAS
MULTIMEDIA
NUSANTARA