

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Terdapat penelitian terdahulu pada tabel 2.1 hingga 2.10 yang dapat dijadikan pedoman dalam melakukan penelitian ini. Berikut rangkuman dari ke sepuluh penelitian terdahulu yang memiliki topik yang selaras dengan Faktor Eksternal Terhadap Prestasi Akademik Mahasiswa.

Tabel 2.1 Tabel Penelitian 1

Penelitian 1 [12]	
Nama Author	Totok Chamidy, Muhammad Ainul Yaqin, Suhartono Suhartono.
Judul Artikel Jurnal	The Influence of Internal and External Factors on Learning Achievement
Nama Jurnal, Vol, Edisi, Tahun	AICOLLIM 2022, ASSEHR 725, 2023
Metode	Kuantitatif; Survei, Multiple Linear Regression
Hasil	Faktor internal: intelegensi (10,04%) dan bakat (3,15%) berpengaruh signifikan. Faktor eksternal: dosen (12,8%) dan kurikulum (2,66%) paling dominan terhadap prestasi akademik mahasiswa. Koefisien determinasi internal 12,1%, eksternal 13,5%.

Tabel 2.2 Tabel Penelitian 2

Penelitian 2 [13]	
Nama Author	Fatia Amalia, Mylo Aleffa Rahman, Sabrina Nur Anisah
Judul Artikel Jurnal	Analisis Anova Two-Way: Pengaruh Kesehatan Mental Dan Tahun Akademik Terhadap Nilai Ipk Mahasiswa Dengan Bahasa Pemrograman Python
Nama Jurnal, Vol, Edisi, Tahun	Data Sciences Indonesia (DSI), Vol 4, 2025
Metode	Analisis ANOVA Dua Arah dengan bahasa pemrograman Python.
Hasil	Hasil penelitian menunjukkan bahwa tahun akademik dan kondisi kesehatan mental (depresi, kecemasan, serangan panik, dan tingkat kesehatan mental) berpengaruh signifikan terhadap IPK mahasiswa ($p < 0,05$). Namun, interaksi antara keduanya tidak signifikan ($p > 0,05$). Korelasi negatif antara IPK dan kesehatan mental ($r = -0,65$) menunjukkan bahwa semakin buruk kesehatan mental, semakin rendah nilai IPK mahasiswa.

Tabel 2.3 Tabel Penelitian 3

Penelitian 3 [14]	
Nama Author	Mohamed El Jihaoui, Oum El Kheir Abra, Khalifa Mansouri
Judul Artikel Jurnal	Factors Affecting Student Academic Performance: A Combined Factor Analysis of Mixed Data and Multiple Linear Regression Analysis
Nama Jurnal, Vol, Edisi, Tahun	IEEE Access, Vol. 13, 2025
Metode	Kuantitatif; Factor Analysis for Mixed Data (FAMD) dan Multiple Linear Regression (MLR) menggunakan data 1.073.450 observasi dari sistem pendidikan Maroko.

Hasil	Ditemukan empat faktor utama yang memengaruhi kinerja akademik mahasiswa, yaitu prior academic performance (76,6%), academic delay (14,34%), socioeconomic status (6,02%), dan class environment (3,03%), semuanya signifikan ($p < 0,001$). Model menjelaskan 88,53% variansi CGPA, menunjukkan bahwa kinerja akademik sebelumnya adalah faktor paling dominan.
-------	--

Tabel 2.4 Tabel Penelitian 4

Penelitian 4 [15]	
Nama Author	Jiansong Chen, Hongjing Pang
Judul Artikel Jurnal	Analyzing Factors Influencing Student Achievement: A Financial and Agricultural Perspective Using SPSS Statistical Analysis Software
Nama Jurnal, Vol, Edisi, Tahun	<i>Journal of Commercial Biotechnology</i> , Vol. 28, 2023.
Metode	Kuantitatif; survei (response rate 26.5%) dan analisis statistik menggunakan SPSS (deskriptif, regresi linier, ANOVA).
Hasil	Ditemukan empat faktor utama yang memengaruhi prestasi belajar siswa: perilaku guru (paling berpengaruh, berdampak positif pada 58% siswa), dukungan orang tua (18% dampak negatif pada siswa berprestasi rendah), lingkungan belajar, dan motivasi siswa (kurang berpengaruh). Studi ini menekankan pentingnya perilaku guru, dukungan orang tua, dan lingkungan belajar sebagai alat untuk meningkatkan prestasi akademik.

Tabel 2.5 Tabel Penelitian 5

Penelitian 5 [16]	
Nama Author	Ibrahim Baba Suleiman, Oluwasogo Adekunle Okunade, Emmanuel Gbenga Dada, Uchenna Christiana Ezeanya
Judul Artikel Jurnal	Key Factors Influencing Students' Academic Performance
Nama Jurnal, Vol, Edisi, Tahun	<i>Journal of Electrical Systems and Information Technology</i> , Vol 11, 2024
Metode	Kuantitatif; analisis regresi linear untuk mengidentifikasi faktor-faktor yang memengaruhi prestasi akademik mahasiswa.
Hasil	Jam belajar ($r = 0,37$) dan nilai akademik sebelumnya ($r = 0,92$) berpengaruh positif signifikan terhadap prestasi. Aktivitas ekstrakurikuler, jam tidur, dan latihan soal berpengaruh sangat lemah. Disarankan penguatan kebiasaan belajar dan kebijakan pendidikan yang memperhatikan faktor sosial-ekonomi.

Tabel 2.6 Tabel Penelitian 6

Penelitian 6 [17]	
Nama Author	Yanuarini Nur Sukmaningtyas, Ronny Makhfuddin Akbar, Gita Rohma, Utami Asyafiiyah
Judul Artikel Jurnal	Penerapan Predictive Analytics untuk Analisis Faktor-faktor yang Mempengaruhi Performa Akademik Siswa
Nama Jurnal, Vol, Edisi, Tahun	<i>Arcitech: Journal of Computer Science and Artificial Intelligence</i> , Vol. 4, 2024
Metode	Pendekatan <i>predictive analytics</i> berbasis data dengan analisis korelasi dan eksplorasi faktor-faktor seperti jam belajar, jam tidur, nilai sebelumnya, latihan soal, dan kegiatan ekstrakurikuler.
Hasil	Ditemukan bahwa <i>previous scores</i> memiliki pengaruh paling kuat terhadap performa akademik siswa ($r = 0,92$), sementara jam belajar, jam tidur, dan latihan soal berpengaruh lebih lemah. Hasil menunjukkan pentingnya data historis dalam memprediksi keberhasilan akademik.

Tabel 2.7 Tabel Penelitian 7

Penelitian 7 [18]	
Nama Author	Abdullah Alshantqi, Abdallah Namoun
Judul Artikel Jurnal	Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification
Nama Jurnal, Vol, Edisi, Tahun	IEEE Access, Vol. 8, 2020
Metode	Menggunakan model <i>Hybrid Regression Model (HRM)</i> dan <i>Multi-Label Self-Organizing Map (MLSOM)</i> untuk memprediksi performa akademik mahasiswa dan mengidentifikasi faktor yang berpengaruh
Hasil	Model HRM menunjukkan akurasi prediksi lebih tinggi dibanding metode lain, semangat MLSOM berhasil mengungkap faktor-faktor dominan yang memengaruhi prestasi mahasiswa, seperti nilai akademik sebelumnya dan tingkat studi. Penelitian menegaskan pentingnya pendekatan prediktif yang juga menjelaskan faktor penyebab performa belajar.

Tabel 2.8 Tabel Penelitian 8

Penelitian 8 [19]	
Nama Author	Ömer Koçak, İdris Göksu, Yüksel Göktas
Judul Artikel Jurnal	The factors affecting academic achievement: A systematic review of meta analyses
Nama Jurnal, Vol, Edisi, Tahun	<i>International Online Journal of Education and Teaching (IOJET)</i> , Vol. 8, No. 1, 2021
Metode	Systematic review dan meta-analisis dari 254 variabel
Hasil	Variabel psikologis, strategi pembelajaran, karakteristik individu, keluarga, guru, dan sekolah dianalisis; sebagian besar variabel berpengaruh positif pada prestasi akademik. Effect size berkisar antara -0,799 hingga 2,987.

Tabel 2.9 Tabel Penelitian 9

Penelitian 9 [20]	
Nama Author	Rani Kurnia Putri, Darvi Mailisa Putri, Lilis Harianti Hasibuan
Judul Artikel Jurnal	Pengaruh Minat Belajar dan Motivasi Belajar terhadap Prestasi Belajar Mahasiswa Program Studi Matematika pada Mata Kuliah Statistika Deskriptif
Nama Jurnal, Vol, Edisi, Tahun	<i>JOSTECH Journal of Science and Technology</i> , Vol. 3, No. 2, 2023
Metode	Kuantitatif, regresi linear berganda
Hasil	Minat dan motivasi belajar berpengaruh positif dan signifikan terhadap prestasi belajar mahasiswa. Persamaan regresi: $Y = 34.983 + 0.745X_1 + 0.395X_2$. Uji F menunjukkan pengaruh simultan yang signifikan ($F = 33.422$, Sig. = 0.000). Koefisien determinasi ($R^2 = 0.485$) berarti kontribusi kedua variabel sebesar 48.5%, sedangkan sisanya dipengaruhi faktor lain.

Tabel 2.10 Tabel Penelitian 10

Penelitian 10 [21]	
Nama Author	Silvia Mona, Prasida Yunita
Judul Artikel Jurnal	Faktor-Faktor yang Berhubungan dengan Prestasi Belajar Mahasiswa
Nama Jurnal, Vol, Edisi, Tahun	Menara Ilmu: Jurnal Penelitian dan Kajian Ilmiah, Vol. XV, 2021
Metode	Survei analitik dengan pendekatan Cross Sectional, menggunakan kuesioner; analisis data dengan Chi-Square
Hasil	Ditemukan hubungan signifikan antara motivasi, minat belajar, dan lingkungan sosial dengan prestasi akademik mahasiswa ($p < 0,05$). Sebagian besar mahasiswa dengan motivasi dan minat tinggi memperoleh prestasi belajar sangat memuaskan.

Berbagai penelitian yang disajikan pada tabel 2.1 hingga 2.10 telah menunjukkan bahwa faktor eksternal memiliki pengaruh yang signifikan terhadap prestasi akademik mahasiswa. Faktor-faktor seperti dukungan orang tua, perilaku dosen, lingkungan belajar, serta kondisi sosial dan ekonomi secara konsisten ditemukan sebagai penentu utama dalam peningkatan capaian prestasi akademik mahasiswa [15], [16]. Perilaku dosen yang profesional dan lingkungan akademik yang kondusif juga dapat meningkatkan pemahaman mahasiswa terkait materi yang diberikan serta kedisiplinan dalam belajar. Serta dukungan orang tua yang baik pun dapat meningkatkan motivasi dan fokus belajar mahasiswa [21], [20]. Penelitian lain juga menunjukkan bahwa fasilitas kampus, kualitas interaksi sosial, dan ketersediaan sumber belajar juga dapat memengaruhi prestasi akademik mahasiswa [19], [14]. Selain itu, bukti bahwa terdapat korelasi positif yang kuat antara nilai akademik mahasiswa dan variabel eksternal yang diperkuat oleh metode kuantitatif dan statistik seperti regresi linier, dan ANOVA [18]. Secara keseluruhan, hasil penelitian tersebut telah menunjukkan bahwa faktor eksternal itu sangat penting untuk mendukung dan membentuk keberhasilan akademik mahasiswa.

2.2 Tinjauan Teori

2.2.1 Prestasi Akademik

Salah satu indikator penting untuk menilai keberhasilan seorang mahasiswa selama di pendidikan perguruan tinggi adalah prestasi akademik mereka. Terdapat banyak faktor yang biasanya digunakan untuk mengukur prestasi ini, seperti lama nya studi serta Indeks Prestasi Kumulatif (IPK). [22]. Prestasi akademik mencerminkan perkembangan kemampuan pengetahuan, dan emosional mahasiswa setelah menjalani proses pendidikan perguruan tinggi ini. Dorongan untuk berprestasi menjadi faktor penting yang telah memotivasi mahasiswa dalam meningkatkan kualitas akademiknya, yang tercermin dari hasil belajar yang optimal. Selain faktor internal seperti *self-esteem* dan motivasi belajar, faktor eksternal seperti *social support* atau dukungan sosial dari orang tua juga berperan signifikan dalam memengaruhi pencapaian prestasi akademik mahasiswa [23].

2.2.2 Faktor Eksternal

Faktor eksternal dalam konteks penelitian ini merupakan suatu kondisi atau pengaruh dari luar diri mahasiswa yang dapat memengaruhi proses dan hasil belajar mereka [5]. Faktor-faktor yang dianggap eksternal tersebut dapat mencakup lingkungan keluarga, kampus, dan masyarakat yang berperan penting dalam membentuk perilaku serta motivasi belajar mahasiswa [24]. Lingkungan keluarga yang suportif mampu meningkatkan semangat dan konsentrasi belajar, sedangkan kondisi sosial ekonomi yang stabil berkontribusi terhadap tersedianya fasilitas pendidikan mahasiswa yang memadai [25]. Dukungan dan keterlibatan orang tua juga berperan dalam menumbuhkan tanggung jawab akademik serta memperkuat kedisiplinan dalam menjalani perkuliahan [26]. Selain itu, interaksi sosial di lingkungan kampus dan dukungan teman sebaya juga turut mendorong menciptakan suasana belajar yang positif dan kolaboratif, yang pada akhirnya dapat meningkatkan prestasi akademiknya [27]. Secara keseluruhan, faktor eksternal tersebut seperti dukungan keluarga, lingkungan sosial, dan kondisi ekonomi dapat memberikan pengaruh yang cukup signifikan terhadap keberhasilan akademik mahasiswa, baik secara langsung melalui bantuan material dan emosional, maupun secara tidak langsung melalui peningkatan motivasi dan kesejahteraan psikologis dalam proses belajar [28].

2.2.3 Big Data Analytics

Big Data Analytics merupakan pendekatan analisis data yang digunakan untuk mengekstraksi wawasan dan pola tersembunyi dari kumpulan data berskala besar, kompleks, dan beragam. Karakteristik utama *Big Data* dikenal sebagai 5V, yaitu *volume*, *velocity*, *variety*, *veracity*, dan *value* [29]. *Volume* menunjukkan jumlah data yang sangat besar; *velocity* menggambarkan kecepatan data dihasilkan dan diproses, *variety* mengacu pada beragamnya tipe data, baik terstruktur maupun tidak terstruktur, *veracity* berhubungan dengan keakuratan dan konsistensi data, sedangkan *value* menekankan pentingnya nilai informasi yang dapat dihasilkan dari proses analisis tersebut [30]. Penerapan Big Data Analytics ini dapat berperan penting dalam memahami

dinamika pembelajaran dan meningkatkan efektivitas sistem akademik. Melalui analisis terhadap data mahasiswa seperti kehadiran, hasil akademik, partisipasi kelas, dan umpan balik dosen, institusi dapat mengenali pola-pola perilaku belajar dan mengidentifikasi faktor-faktor yang berkontribusi terhadap keberhasilan akademik [31]. Selain itu, *Big Data Analytics* mampu mendukung proses manajemen pendidikan yang lebih efisien melalui pemantauan performa akademik secara real-time, prediksi kebutuhan sumber daya, serta peningkatan layanan pembelajaran digital.

2.3 Metode Analisis

2.3.1 Statistik Deskriptif

Statistik deskriptif berfungsi sebagai langkah awal yang sistematis untuk merangkum, menyajikan, dan menggambarkan karakteristik penting dari dataset penelitian sehingga memudahkan dalam memahami pola, penyebaran, dan kelainan data sebelum melakukan analisis inferensial atau pemodelan lebih lanjut [32]. Statistik deskriptif ini mencakup perhitungan ukuran kecenderungan pusat, seperti rata-rata dan median, serta ukuran penyebaran, seperti varians dan simpangan baku. Ukuran-ukuran ini memberikan informasi terkait nilai pusat dan penyebaran data sampel [33].

Rumus untuk menghitung rata-rata adalah:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Sedangkan variansi dapat dihitung dengan:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Sedangkan standar deviasi dihitung dengan:

$$s = \sqrt{s^2}$$

Selain ukuran numerik, visualisasi grafik seperti *histogram*, *boxplot*, dan *density plot* penting untuk mengidentifikasi bentuk distribusi, keberadaan outlier, dan ketidaksimetrian (skewness) data sehingga asumsi-asumsi lanjutan untuk uji parametrik ini dapat dievaluasi. Proses eksplorasi deskriptif meliputi pengecekan *missing values* dan frekuensi kategori pada variabel kategorikal. Proses ini juga melibatkan pembuatan tabel silang sederhana untuk memahami struktur data sebelum memilih uji statistik yang sesuai [34]. Statistik deskriptif memberikan landasan untuk mengambil keputusan metodologis, seperti mentransformasi variabel, mengambil logaritma dari suatu variabel, dan mengkategorikan variabel untuk analisis korelasi dan regresi lebih lanjut [32].

2.3.2 Analisis Korelasi Spearman

Spearman rank correlation sering digunakan untuk mengukur kekuatan dan arah hubungan antara dua variabel ordinal atau antara variabel yang tidak memenuhi asumsi normalitas, sehingga menjadi pilihan yang bagus dalam banyak studi akademik yang memanfaatkan data berskala ordinal atau data metrik dengan pencilan. Nilai koefisiennya berkisar dari -1 sampai $+1$ dan interpretasinya mengikuti tanda dan magnitudo tersebut. Metode ini sebenarnya didefinisikan sebagai koefisien Pearson yang dihitung atas peringkat (*rank*) dari setiap variabel, sehingga nilai Spearman ini juga dapat ditulis sebagai Pearson pada variabel berperingkat dan sifat-sifat inferensialnya mengikuti transformasi tersebut ketika ukuran sampel cukup besar [35]. Simbol ukuran populasinya adalah ρ dan ukuran sampelnya r_s . Formula r_s untuk korelasi Spearman adalah sebagai berikut:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

Di mana d_i adalah selisih peringkat setiap pasangan observasi ke- i , dan n adalah jumlah observasi. Dalam situasi di mana terdapat nilai seri (*ties*), definisi yang lebih umum menempatkan r_s sebagai kovarians antar peringkat dibagi produk simpangan baku peringkat, dan praktik yang disarankan adalah menggunakan korelasi Pearson pada peringkat atau menerapkan koreksi untuk

ties agar estimasi dan uji signifikansi menjadi akurat [35]. Dari sisi asumsi dan kelebihan/kelemahan, Spearman tidak memerlukan distribusi normal maupun linearitas, dan relatif lebih tahan terhadap outlier dibanding korelasi parametris seperti *Pearson correlation coefficient*. Oleh karena itu, pada data distribusi menyimpang, data ordinal, atau data dengan outlier, Spearman sering kali memberikan estimasi yang lebih stabil dan valid [36]. Namun demikian, terdapat *trade-off* metodologis, yaitu bahwa koefisien Spearman dapat memiliki variabilitas yang lebih besar dibandingkan koefisien Pearson ketika data berdistribusi normal dan bebas outlier, sehingga pada kondisi tertentu Pearson dapat memberikan estimasi yang lebih efisien [37].

2.3.3 Regresi Linear Berganda

Regresi linear berganda merupakan teknik statistik yang memodelkan hubungan linier antara sebuah variabel dependen kontinu dan dua atau lebih variabel independen sehingga dapat digunakan untuk prediksi dan pengujian hipotesis sebab-akibat dalam kerangka linear [38]. Model umum dinyatakan sebagai:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon,$$

Model tersebut merepresentasikan hubungan linier antara satu variabel dependen Y dan sejumlah variabel independen X_1, X_2, \dots, X_k di mana setiap koefisien β_j menggambarkan perubahan rata-rata pada Y akibat perubahan satu satuan pada X_j dengan variabel lain dianggap konstan [39]. Struktur model ini memungkinkan evaluasi kontribusi relatif masing-masing prediktor terhadap variasi respons, sehingga mampu menangkap efek simultan dalam satu kerangka analitis. Komponen β_0 berfungsi sebagai intercept yang menunjukkan nilai ekspektasi Y ketika seluruh prediktor bernilai nol, sedangkan ε mencerminkan deviasi antara nilai observasi dan nilai yang diprediksi model, yang diasumsikan memiliki nilai harapan nol serta distribusi yang identik dan independen [39].

2.3.4 Metrik Evaluasi

2.3.4.1 MAE

Mean Absolute Error (MAE) adalah metrik evaluasi regresi yang mengukur rata-rata nilai mutlak selisih antara nilai observasi aktual dan nilai prediksi sehingga memberikan gambaran langsung tentang ukuran kesalahan prediksi dalam satuan variabel [40]. Secara matematis, MAE didefinisikan sebagai :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

dengan y_i nilai sebenarnya, \hat{y}_i nilai prediksi, dan n jumlah observasi, rumus MAE mudah dipahami karena setiap deviasi dihitung tanpa proses penguadratan [41]. MAE menggunakan selisih absolut sehingga metrik ini lebih tahan terhadap nilai kesalahan yang sangat ekstrem dibandingkan MSE atau RMSE. Ketika digunakan sebagai fungsi kehilangan, MAE menghasilkan estimasi yang mendekati median kesalahan, sehingga memberikan hasil yang lebih stabil pada data yang mengandung outlier. Dengan karakteristik tersebut, MAE menjadi pilihan yang tepat ketika dibutuhkan ukuran kesalahan yang sederhana, mudah ditafsirkan, dan memberikan penalti yang seimbang untuk setiap penyimpangan [41]. Kelemahan MAE adalah tidak dapat membedakan besar-kecilnya kesalahan karena tidak memberi bobot lebih pada kesalahan besar, sehingga model dengan beberapa outlier besar dapat tampak serupa dengan model yang memiliki banyak kesalahan kecil [40].

2.3.4.2 RMSE

RMSE adalah metrik evaluasi yang mengukur rata-rata *error* antara prediksi dan nilai aktual, dengan memberikan penalti lebih besar pada *error* yang besar dibandingkan MAE. [41]. Secara matematis, RMSE didefinisikan sebagai

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

dengan y_i sebagai nilai sebenarnya, \hat{y}_i sebagai nilai prediksi, dan n jumlah observasi. Rumus ini menunjukkan bahwa setiap selisih prediksi terlebih dahulu dikuadratkan sebelum dirata-ratakan, lalu diakarkan kembali agar hasilnya berada pada satuan yang sama dengan variabel dependen. RMSE sering dipilih ketika *error* perlu mendapat perhatian khusus karena kuadrat selisih membuat nilai yang jauh menyimpang menjadi lebih dominan dalam perhitungan [40], [41]. RMSE memiliki keterkaitan erat dengan deviasi standar residual ketika error diasumsikan berdistribusi normal, sehingga metrik ini dianggap optimal untuk model yang dibangun dengan asumsi error berdistribusi normal. [40]. Oleh karena itu, RMSE digunakan secara luas dalam evaluasi model prediksi yang dianggap menghasilkan pola *error* yang hampir identik dengan distribusi normal. Karena nilai RMSE bergantung pada skala variabel dependen, interpretasi RMSE harus mempertimbangkan konteks. Tanpa normalisasi atau penggunaan ukuran error relatif, perbandingan antar model dengan skala yang berbeda dapat menjadi tidak akurat. [41].

2.3.4.3 MSE

Mean Squared Error (MSE) merupakan ukuran rata-rata dari kuadrat perbedaan antara nilai prediksi dan nilai aktual, dan metrik ini memiliki peran teoretis penting dalam evaluasi model karena dapat menunjukkan seberapa jauh rata-rata kesalahan prediksi [40]. Secara matematis, MSE dirumuskan sebagai:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

dengan y_i sebagai nilai sebenarnya, \hat{y}_i sebagai nilai prediksi, dan n sebagai jumlah observasi. Rumus tersebut menunjukkan bahwa setiap

selisih prediksi dikuadratkan terlebih dahulu agar kesalahan yang lebih besar memiliki pengaruh lebih kuat dalam hasil akhir. MSE juga digunakan sebagai fungsi kehilangan dalam metode *Ordinary Least Squares* (OLS) karena bentuknya yang diferensiabel sehingga memudahkan proses estimasi parameter secara analitis [40]. Secara teoretis, MSE dapat diuraikan menjadi varians dan bias kuadrat, sehingga membantu menjelaskan bagaimana perubahan kompleksitas model dapat menurunkan kesalahan sistematis tetapi meningkatkan ketidakstabilan prediksi, atau sebaliknya [40]. Metrik ini sangat sensitif terhadap nilai kesalahan yang besar karena adanya proses penguadratan, yang bermanfaat ketika kesalahan besar perlu diperhatikan, tetapi dapat kurang sesuai bila data mengandung nilai ekstrem yang tidak representatif [40], [41]. Dalam praktiknya, MSE sering disajikan bersama RMSE agar interpretasinya kembali ke satuan asli variabel, dan biasanya dilengkapi dengan MAE sebagai ukuran yang lebih robust untuk memberikan gambaran performa model secara lebih menyeluruh [41].

2.3.4.4 R^2

adalah ukuran yang menunjukkan seberapa besar proporsi variasi pada variabel dependen yang dapat dijelaskan oleh variabel independen dalam model regresi. R^2 ini memiliki nilai antara 0 dan 1 dan membuatnya mudah dipahami sebagai persentase variabilitas yang berhasil dijelaskan. [42]. Secara matematis, R^2 untuk model linear dengan intercept dapat dituliskan sebagai:

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

Di mana \bar{y} adalah rata-rata observasi, SS_{reg} adalah jumlah kuadrat yang dijelaskan model, dan SS_{tot} adalah total jumlah kuadrat [42]. Metrik ini sering digunakan untuk menggambarkan kualitas kecocokan model karena memberikan ringkasan yang jelas mengenai seberapa besar varians data yang dapat diterangkan oleh prediktor. Meskipun demikian, interpretasi R^2 perlu dilakukan dengan hati-hati karena nilai R^2 akan meningkat ketika

jumlah prediktor ditambah, bahkan jika prediktor tersebut tidak benar-benar memberikan kontribusi berarti; oleh sebab itu, nilai *adjusted R²* lebih tepat digunakan ketika membandingkan model dengan jumlah variabel berbeda [42]. Selain itu, *R²* tidak dapat menggantikan pemeriksaan terhadap asumsi regresi, tidak memberikan informasi mengenai bias prediksi, dan tidak selalu mencerminkan kemampuan model dalam memprediksi data baru. Oleh karena itu, laporan penelitian yang komprehensif biasanya mengombinasikan *R²* dengan metrik kesalahan seperti MAE atau RMSE serta prosedur validasi silang untuk menilai performa model secara lebih menyeluruh [40], [42].

2.4 Tools/Software yang Digunakan

2.4.1 Jupyter Notebook

Jupyter Notebook merupakan sebuah platform *open-source* yang memungkinkan untuk menulis, menjalankan kode yang telah disusun, menampilkan visualisasi, serta mendokumentasikan proses analisis dalam satu dokumen interaktif. Platform ini mendukung berbagai bahasa pemrograman, termasuk *Python*, dan banyak digunakan dalam penelitian ilmiah maupun pendidikan. Penggunaannya ini dapat memberikan kemudahan dalam melakukan eksperimen data, dokumentasi hasil, serta kolaborasi antarpeneliti secara transparan [43]. Selain itu, Jupyter Notebook juga telah dimanfaatkan sebagai media pembelajaran interaktif di bidang akademik yang memungkinkan mahasiswa untuk dapat menggabungkan kode, visualisasi, dan narasi dalam satu dokumen berbasis web [43]. Pendekatan ini membantu meningkatkan pemahaman konsep-konsep ilmiah yang kompleks melalui eksplorasi berbasis komputasi.

Jupyter Notebook dikenal sebagai akronim dari tiga bahasa pemrograman utama yang didukungnya, yaitu Python (Py), dan R, memberikan fleksibilitas bagi penggunaannya untuk dapat menyesuaikan bahasa sesuai kebutuhan analisis. Panduan penggunaan dan analisis data berbasis Jupyter juga banyak tersedia secara *online* untuk mendukung pembelajaran mandiri dan penerapan

dalam penelitian [44]. Melalui kemampuannya dalam integrasi penulisan kode, visualisasi, dokumentasi, serta kolaborasi, Jupyter Notebook menjadi platform utama dalam penelitian ini untuk melakukan analisis data secara interaktif, transparan, dan efisien.

2.4.2 Python

Python adalah bahasa pemrograman tingkat tinggi yang populer di kalangan peneliti dan profesional pada bidang data science dikarenakan sintaksnya yang sederhana dan kemampuannya dalam pengolahan data berskala besar. Keunggulannya terletak pada cakupan *library* yang cukup luas, seperti Pandas, NumPy untuk komputasi numerik, serta Matplotlib dan Seaborn untuk visualisasi data. Penggunaan Python memungkinkan integrasi yang mulus antara analisis data, pembuatan model prediksi, dan visualisasi hasil dalam satu platform yang terkoordinasi [45].

Python juga mendukung berbagai teknik *machine learning* dan metode statistik yang penting dalam analisis data pendidikan. Berbagai studi telah menunjukkan pemanfaatannya dalam penerapan model regresi untuk memprediksi prestasi akademik mahasiswa [46], pengembangan sistem otomatis analisis hasil akademik berbasis Python dan Streamlit [47], serta tinjauan komprehensif mengenai *library* Python untuk data science yang mencakup proses pengumpulan, analisis, dan visualisasi data [48]. Dengan dukungan tersebut, Python berperan sebagai alat utama dalam penelitian ini, menyediakan fondasi yang kuat untuk analisis data dan pengembangan model prediksi yang efektif.

2.4.3 Library Statistik & Visualisasi

Library Python seperti Pandas, NumPy, Matplotlib, Seaborn, dan Scikit-learn merupakan komponen penting dalam pengolahan serta analisis data dalam penelitian ini. Pandas menyediakan struktur *DataFrame* yang memudahkan manipulasi data tabular, sementara NumPy mendukung komputasi numerik secara efisien. Matplotlib dan Seaborn digunakan untuk membuat visualisasi data yang membantu dalam memahami pola serta

hubungan antar variabel. Sementara itu, Scikit-learn menyediakan berbagai algoritma pembelajaran mesin dan fungsi evaluasi model, seperti *RMSE*, *MSE*, *MAE*, dan R^2 , yang digunakan untuk menilai performa model prediksi [45]. Penggunaan *library* ini memungkinkan proses eksplorasi data, pembuatan model, dan evaluasi hasil dilakukan secara sistematis dan efisien. Selain itu, *library* tersebut mendukung otomatisasi proses analisis, mengurangi potensi kesalahan manual, dan meningkatkan reproduktibilitas hasil. Pemanfaatannya juga meluas pada sistem analisis akademik berbasis Python dan Streamlit untuk ekstraksi data serta visualisasi hasil secara interaktif [47]. Selain itu, telah dilakukan tinjauan mendalam terhadap perpustakaan-perpustakaan Python, mencakup topik-topik seperti pengumpulan data, analisis, pemrosesan, dan visualisasi data [48]

2.4.4 Power BI

Power BI merupakan sebuah platform business intelligence (BI) dan visualisasi data yang dikembangkan oleh Microsoft. Power BI dirancang untuk mengubah data mentah menjadi wawasan bisnis yang bermakna melalui dashboard yang interaktif, custom report, dan pembaruan data secara real-time [49]. Power BI memungkinkan pengguna nya dari berbagai latar belakang, seperti *business analyst*, *project manager*, *data scientist* untuk menghubungkan, mengimpor, dan mengintegrasikan data dari berbagai sumber termasuk database, layanan cloud (Azure, AWS, Google Cloud), serta file seperti Excel, CSV, dan JSON ke dalam satu interface yang intuitif dan mudah digunakan [50].