

BAB III

METODE PENELITIAN

3.1. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif, yang menekankan pengumpulan, analisis, dan interpretasi data numerik untuk mengukur dan memahami bagaimana variabel berinteraksi satu sama lain. Metode ini berguna untuk menilai hubungan dan pengaruh faktor eksternal terhadap prestasi akademik mahasiswa karena memungkinkan pengumpulan data yang dapat diukur dari sampel representatif. [51]. Faktor-faktor yang dianalisis mencakup faktor keluarga, faktor pilihan studi, faktor ekonomi dan pekerjaan, faktor kesehatan dan gaya hidup, faktor akademik dan dosen, serta faktor fasilitas kampus. Data diperoleh melalui kuesioner terstandar guna memastikan konsistensi data dan mengurangi bias dalam analisis.

Metode kuantitatif ini dilakukan menggunakan teknik analisis berbasis bahasa pemrograman Python untuk memastikan proses integrasi data dari berbagai sumber, baik data dari akademik maupun data dari non-akademik. Analisis dimulai dengan statistik deskriptif untuk mempelajari distribusi data dan karakteristik data. Selanjutnya, analisis korelasi digunakan untuk mengevaluasi hubungan antarvariabel. Terakhir, untuk memperhitungkan pengaruh faktor-faktor eksternal terhadap prestasi akademik siswa, regresi linear berganda digunakan. Penelitian ini juga menggunakan MAE, RMSE, MSE, dan R^2 sebagai evaluasinya. Untuk mempermudah interpretasi pola hubungan antarvariabel, hasil analisis kemudian dipresentasikan menggunakan dashboard deskriptif melalui tool Microsoft Power BI. Selanjutnya, temuan tersebut ditafsirkan untuk memahami fenomena yang diamati dan mengevaluasi hipotesis yang diajukan. Tujuan interpretasi adalah untuk menunjukkan sejauh mana faktor luar berkontribusi terhadap prestasi akademik siswa. Faktor-faktor seperti dukungan sosial, lingkungan kampus, dan keluarga memengaruhi keinginan siswa untuk belajar dan kinerja akademik mereka [52]. Metode ini memastikan bahwa hasil penelitian benar secara numerik dan relevan

untuk pengambilan keputusan pendidikan dan pengembangan kebijakan akademik di perguruan tinggi Indonesia.

3.1.1 Metode Framework

Dalam penelitian ini, jenis *framework* yang akan digunakan yaitu *framework* KDD. Berikut adalah tabel yang membandingkan *framework* KDD dan CRISP-DM:

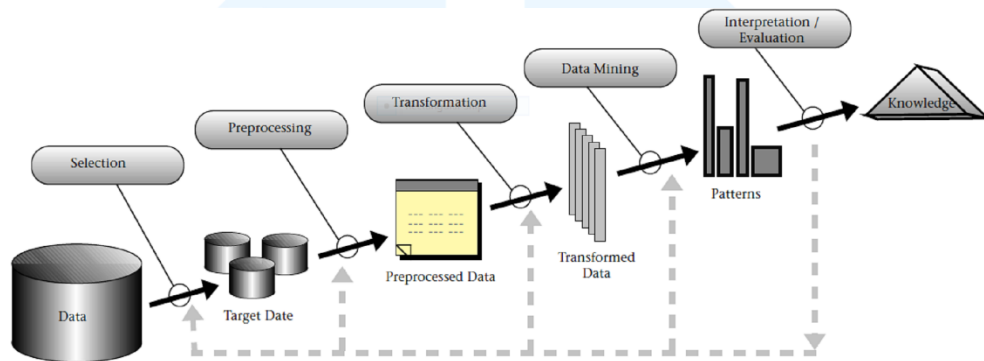
Tabel 3.1 Perbandingan Kerangka Kerja Data Mining

Indikator	KDD	CRISP-DM
Pendekatan	Sistematis, berurutan, dan filosofis, menekankan tahapan eksplisit dari pemilihan data hingga penemuan <i>knowledge</i> nya.	Iteratif dengan enam fase umum yang menekankan fleksibilitas, evaluasi model, serta kebutuhan konteks bisnis yang berulang.
Fokus Penelitian	Sangat sesuai untuk riset akademik yang tujuannya adalah mengungkapkan pola tersembunyi dan menemukan hubungan baru antar variabel (<i>knowledge extraction</i>)	Lebih cocok untuk <i>project</i> berbasis industri, yang menekankan manajemen <i>project</i> , fleksibilitas tinggi, dan <i>deployment</i> model ke lingkungan operasional.
Kelebihan	Memiliki struktur analisis yang jelas, logis, dan kuat untuk validasi hasil. Setiap tahap memiliki fokus filosofis yang ketat menuju <i>knowledge</i> .	Memiliki siklus yang fleksibel dan mudah diadaptasi ke berbagai jenis masalah bisnis, menyediakan check-point manajemen <i>project</i> yang rinci.
Kelemahan	Kurang fleksibel jika ada kebutuhan mendadak untuk kembali ke tahap awal, urutan tahapan cenderung lebih ketat.	Strukturnya kurang menekankan aspek filosofis dari penemuan <i>knowledge</i> itu sendiri, dan lebih berfokus pada hasil analisis yang cepat.
Kemampuan Mengungkap Pola	Lebih efektif dalam menemukan pola baru dari dataset kompleks (terutama setelah proses transformasi data yang ketat) karena fokusnya adalah pada <i>knowledge</i> .	Terbatas pada hasil analisis yang terstruktur dan mudah diaplikasikan secara langsung di lingkungan bisnis/operasional.

Sumber : [53]

Berdasarkan dari tabel 3.1 yaitu tabel perbandingan *framework* tersebut, terlihat bahwa KDD merupakan *framework* yang lebih efektif dan sesuai untuk digunakan pada penelitian ini. Hal ini dibuktikan dengan fakta bahwa *framework* KDD bertujuan untuk mengekstraksi pola yang berguna, dan dapat dipahami dari kumpulan data. Selain itu juga, KDD memiliki tahapan yang lebih rinci dalam

proses pengolahan data, dimulai dari *data selection*, *data preprocessing*, *data transformation*, hingga penemuan pola serta interpretasi hasil seperti pada Gambar 3.1. Struktur tahapan yang sistematis ini membantu memperoleh pemahaman yang lebih mendalam terhadap data melalui proses analisis yang terarah dan logis.



Gambar 3.1 Framework KDD

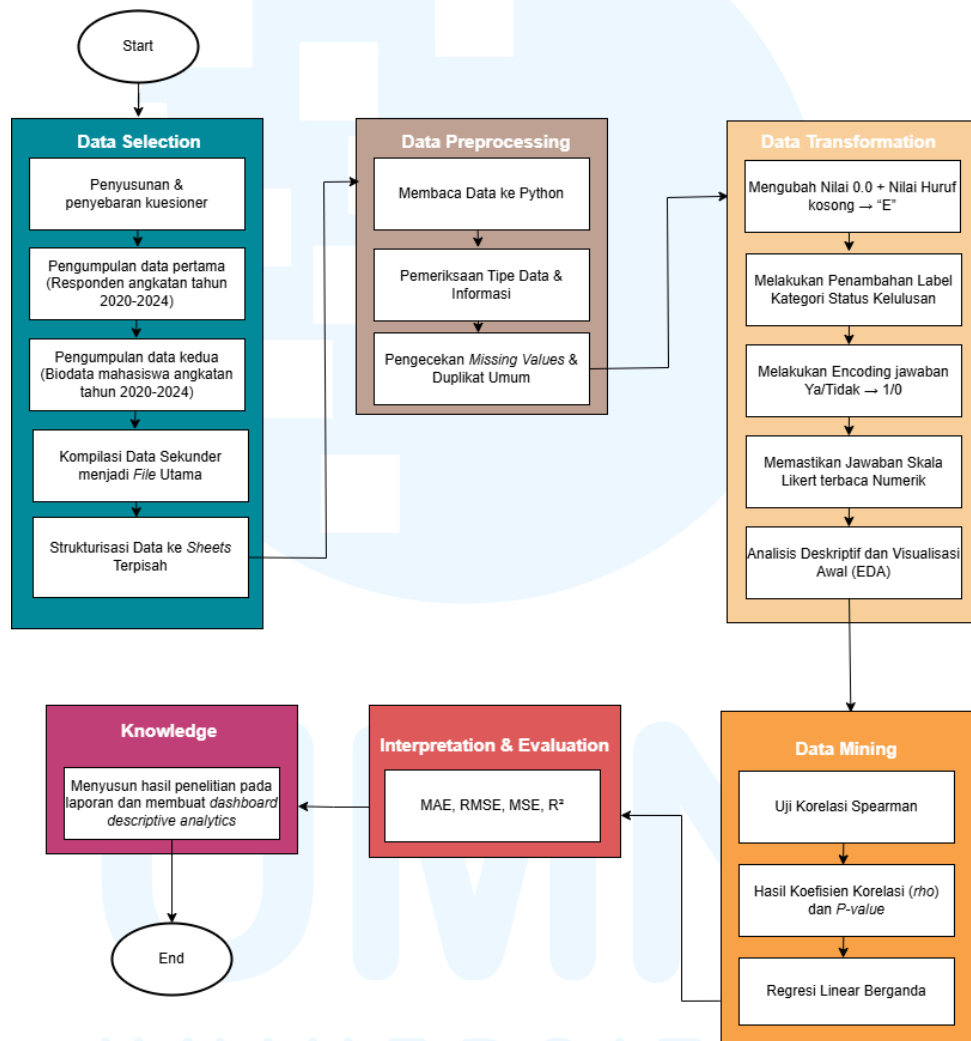
Sumber : [53]

Proses KDD dimulai dengan memahami konteks dan tujuan analisis. Kemudian, masalah utama dan variabel yang relevan diidentifikasi untuk menentukan jalan analisis. Pada tahap berikutnya, pemilihan dan preprocessing data difokuskan pada pemilihan data sesuai dengan persyaratan dan perbaikan kualitas seperti penanganan nilai yang hilang dan penghapusan duplikasi. Data juga diubah untuk menyederhanakan dan menekankan fitur yang paling penting. Pada tahap inti, algoritma statistik, data mining, dan analisis digunakan untuk menemukan pola atau hubungan tersembunyi dalam data. Hasil ini kemudian digunakan untuk menginterpretasikan dan mengevaluasi. Di sini, pola dinilai berdasarkan relevansi dan akurasi terhadap tujuan analisis. Knowledge representation adalah tahap terakhir, di mana pengetahuan yang diperoleh disajikan dalam bentuk laporan atau visualisasi yang mudah dipahami sehingga dapat digunakan sebagai dasar untuk pengambilan keputusan atau strategi pengembangan akademik. [53].

3.2 Tahapan Penelitian

Alur penelitian ini disusun berdasarkan tahapan *Knowledge Discovery in Database (KDD)* yang digunakan sebagai kerangka utama dalam proses

pengolahan dan analisis data. Setiap tahap dalam model KDD memiliki peran penting dalam memastikan bahwa data yang dikumpulkan dapat diolah secara sistematis hingga menghasilkan pengetahuan baru yang relevan dengan tujuan penelitian. Berikut Gambar 3.2 yang menunjukkan alur penelitian yang disusun berdasarkan tahapan KDD.



Gambar 3.2 Alur Penelitian Menggunakan KDD

a) Data Selection

Tahap ini dimulai dengan pengumpulan data sekunder dari dua sumber utama, yaitu data historis mahasiswa dan hasil kuesioner mahasiswa angkatan tahun 2020–2024 di Universitas Multimedia Nusantara. Setelah semua data dikumpulkan, proses penggabungan dari banyak file

berbeda menjadi satu file data utama dilakukan. Selanjutnya, struktur data disusun kembali ke dalam berbagai lembar, termasuk Biodata, Transkrip Nilai, Mata Kuliah, dan Responden. Ini memastikan bahwa data memiliki struktur yang rapi dan siap untuk dilanjutkan.

b) Data Preprocessing

Tahap ini dilakukan menggunakan Python untuk memastikan kelayakan dan kebersihan data. Proses dimulai dengan membaca file utama. Setelah itu, struktur data diperiksa, yang mencakup tipe data, informasi dataset, dan identifikasi nilai yang hilang. Pada tahap ini, juga dilakukan pengecekan kembali terhadap data duplikat atau ketidakkonsistenan lain yang berpotensi mengganggu analisis. Sebelum transformasi dilakukan, hasil preprocessing memastikan bahwa data sudah valid.

c) Data Transformation

Tahap ini berfokus pada penyesuaian nilai dan format data. Hal ini mencakup menambahkan label kategori status kelulusan, mengubah nilai 0.0 yang tidak memiliki huruf menjadi "E" menggunakan logika Python, mengonversi jawaban survei "ya/tidak" menjadi nilai biner 1/0, dan memastikan bahwa semua variabel berskala Likert dibaca sebagai tipe numerik sehingga memenuhi standar untuk analisis selanjutnya. Setelah transformasi, analisis deskriptif dan visualisasi awal dilakukan untuk mempelajari karakteristik dasar dataset melalui perhitungan statistik sederhana seperti nilai minimum, maksimum, rata-rata, dan median. Selain itu, dibuat juga visualisasi awal seperti *column chart*, *pie chart*, dan *bar chart* untuk menemukan pola, kecenderungan, dan kemungkinan outlier sebelum melanjutkan ke tahap pemodelan berikutnya.

d) Data Mining

Tahap ini merupakan bagian penting dari proses KDD, yang bertujuan untuk menemukan hubungan antara faktor-faktor eksternal dan prestasi akademik mereka. Uji korelasi Spearman digunakan untuk memulai analisis untuk mengevaluasi kekuatan dan arah hubungan antara variabel eksternal dan IPK. Hasilnya, yaitu nilai koefisien korelasi (ρ) dan nilai p-value, digunakan untuk mengukur signifikansi hubungan dan menentukan variabel mana yang dapat dilanjutkan ke regresi linear berganda.

e) Interpretation and Evaluation

Tahap ini menafsirkan hasil regresi linear berganda berdasarkan koefisien regresi dan metrik evaluasi model. Koefisien regresi digunakan untuk mengukur arah dan besarnya pengaruh setiap faktor eksternal terhadap IPK (misalnya, apakah pengaruh tersebut positif atau negatif). MAE, MSE, RMSE, dan R^2 digunakan untuk mengevaluasi kinerja model. Sementara MSE dan RMSE menunjukkan kesalahan kuadrat rata-rata dan besarnya kesalahan dalam skala IPK, MAE menunjukkan rata-rata selisih absolut antara nilai prediksi dan nilai aktual IPK. Nilai R^2 digunakan untuk menentukan seberapa baik model dapat menjelaskan variasi IPK.

f) Knowledge

Tahap *Knowledge* merupakan tahap akhir dalam proses KDD yang bertujuan merangkum seluruh hasil analisis menjadi pengetahuan yang dapat dimanfaatkan secara praktis. Data akademik mahasiswa disajikan melalui dashboard deskriptif menggunakan Microsoft Power BI untuk memberikan gambaran umum mengenai karakteristik dan kondisi akademik mahasiswa. Hasil analisis korelasi dan regresi disajikan melalui aplikasi analitik berbasis web menggunakan Streamlit sebagai bentuk *deployment* hasil pengolahan data, sehingga hasil analisis tidak

hanya disampaikan dalam laporan tertulis, tetapi juga dapat diakses dan dimanfaatkan secara langsung oleh pihak terkait.

3.3 Teknik Pengumpulan Data

Penelitian ini menggunakan desain studi kuantitatif dengan memanfaatkan data historis yang diintegrasikan dengan data kuesioner. Dua sumber utama dalam penelitian ini saling melengkapi selama periode pengumpulan data. Sumber data pertama diperoleh melalui penyebaran kuesioner yang dibuat menggunakan Google Forms yang dirancang untuk menggali faktor-faktor sosial dan akademik yang memengaruhi capaian prestasi mahasiswa. Data tambahan berasal dari data historis mahasiswa Program Studi Sistem Informasi di Universitas Multimedia Nusantara. Data ini digunakan untuk secara faktual untuk memperoleh gambaran umum tentang profil akademik mahasiswa.

3.3.1 Populasi dan Sample

Penelitian ini melibatkan 1023 mahasiswa dari program studi Sistem Informasi Universitas Multimedia Nusantara dari angkatan 2020–2024. Mahasiswa Sistem Informasi Universitas Multimedia Nusantara dengan angkatan tersebut, yang mengisi kuesioner biodata akademik melalui Google Form, digunakan sebagai sampel. Profil demografis responden, kondisi keluarga, gaya hidup, faktor kesehatan, aktivitas atau pekerjaan sampingan, dan faktor-faktor di kampus yang dianggap dapat memengaruhi prestasi akademik mereka dikumpulkan dalam kuesioner tersebut.

3.3.2 Periode Pengambilan dan Persiapan Data

Periode pengambilan data dalam penelitian ini dirancang secara sistematis dan menyeluruh untuk menjamin kelengkapan, keakuratan, dan validitas informasi yang relevan. Berikut adalah langkah-langkah pengambilan data:

1) Pengumpulan Sumber Data

Dua sumber data utama dikumpulkan secara bersamaan. Sumber data pertama, kuesioner, telah dibagikan kepada siswa aktif dari 31 Agustus 2025 hingga 28 September 2025. Kuesioner ini menilai faktor eksternal seperti keluarga, pilihan pendidikan, ekonomi dan pekerjaan, kesehatan

dan gaya hidup, akademik dan pendidikan, serta fasilitas kampus. Tabel 3.2 menggambarkan struktur data kuesioner yang digunakan.

Tabel 3.2 Gambaran Data Responden

Nama Atribut	Nama Lengkap
	Jenis Kelamin
	Nomor Induk Mahasiswa
	Email Kampus
	Program Studi
	Tahun Angkatan
	IPK Terakhir
	Tempat Tinggal Sekarang
	Pekerjaan Ayah & Ibu & Wali
	Pendapatan Ayah & Ibu & Wali
	Uang Saku Sekarang
	Beberapa Pertanyaan Lainnya

Secara bersamaan, diperoleh pula data historis akademik mahasiswa pada tanggal 25 September 2025 dari beberapa sumber file yang menyediakan catatan lengkap riwayat perkuliahan mahasiswa. Data akademik ini berisi informasi mengenai riwayat pengambilan mata kuliah, nilai, kehadiran, serta beberapa atribut demografis mahasiswa. Untuk memberikan gambaran awal mengenai struktur data akademik yang menjadi dasar proses pengolahan, berikut ditampilkan tabel atribut yang terdapat pada dataset tersebut pada Tabel 3.3.

Tabel 3.3 Gambaran Data Akademik

Nama Atribut	Nomor Induk Mahasiswa (NIM)
	Nama Mahasiswa
	Angkatan
	Semester Ambil
	Kode Mata Kuliah
	Nama Mata Kuliah
	Kelas
	Nilai Angka
	Nilai Huruf
	IPS
	Jumlah Hadir
	Tanggal Lahir

	Asal Daerah
	Asal Sekolah
	Pendidikan Ayah & Ibu

2) Kompilasi dan Transformasi Data Awal

Setelah periode pengumpulan kuesioner selesai, data primer dari responden dan data historis akademik mahasiswa kemudian dikompilasi pada tanggal 29 September 2025 untuk membentuk satu kesatuan dataset yang siap diolah. Proses kompilasi ini mencakup dengan penggabungan beberapa file sumber yang berbeda, baik yang berasal dari kuesioner maupun dari arsip akademik ke dalam satu file yang sama. Penyatuan kedua sumber data ini dilakukan menggunakan tool Microsoft Excel, sebagaimana ditunjukkan pada Gambar 3.3.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	NIR	NAMA	ANGKATAN	SEMESTER	KODE MK	NAMA MK	RELAS	NILAI ANGKA	NILAI HURUF	IPS	JUMLAH_HADIR	TANGGAL_LAHIR	ASAL DAERAH	ASAL SEKOLAH	PENDIDIKAN AYAH	PENDIDIKAN IBU
2	42528	YEREMIA SI	2020	2011	IS 100	Dasar-Dasar Pemrograman	H	63 C+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
3	42528	YEREMIA SI	2020	2011	IS 100	Management Information Systems	A	72 B	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
4	42528	YEREMIA SI	2020	2011	UM 122	English 1	AG	76 B+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
5	42528	YEREMIA SI	2020	2011	EA 101	Accounting Fundamentals	A	79 B+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
6	42528	YEREMIA SI	2020	2011	EM 190	Introduction to Business Manag	A	79 B+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
7	42528	YEREMIA SI	2020	2011	IS 251	Organizational Behavior	A	84 A-	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
8	42528	YEREMIA SI	2020	2011	IS 155	Introduction to System and IT	A	90 A	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
9	42528	YEREMIA SI	2020	2021	IS 255	Algorithm and Data Structure	A	72 B	3	3	12	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
10	42528	YEREMIA SI	2020	2021	UM 152	Religion	F	75 B	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
11	42528	YEREMIA SI	2020	2021	IS 240	Probability and Statistic	A	76 B+	3	3	15	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
12	42528	YEREMIA SI	2020	2021	IF 141	Discrete Mathematics	A	82 A-	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
13	42528	YEREMIA SI	2020	2021	IS 201	Corporate Business Processes	A	82 A-	3	3	13	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
14	42528	YEREMIA SI	2020	2021	IS 220	Human and Computer Interaction	A	82 A-	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
15	42528	YEREMIA SI	2020	2021	UM 223	English 2	Z	88 A	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
16	42528	YEREMIA SI	2020	2021	IS 541	Database Systems	A	86 B	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
17	42528	YEREMIA SI	2020	2021	IS 652	Seminars on Information System	A-BL	72 B	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
18	42528	YEREMIA SI	2020	2021	IS 534	Mobile Application Development	A	74 B	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
19	42528	YEREMIA SI	2020	2021	IS 388	Data Analysis	C-BL	79 B+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
20	42528	YEREMIA SI	2020	2021	UM 163	Clives	AP	82 A-	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
21	42528	YEREMIA SI	2020	2021	IS 321	IS Analysis & Design	A	84 A-	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
22	42528	YEREMIA SI	2020	2021	UM 321	English 3	T	87 A	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
23	42528	YEREMIA SI	2020	2021	IS 311	Security and Computer Network	C	93 A	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
24	42528	YEREMIA SI	2020	2021	IS 429	Big Data Analytics	A	61 C+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
25	42528	YEREMIA SI	2020	2021	IS 512	IS Project Management	B	68 B	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
26	42528	YEREMIA SI	2020	2021	UM 162	Pancasila	C	73 B	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
27	42528	YEREMIA SI	2020	2021	IS 411	Data Modelling	B	75 B+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
28	42528	YEREMIA SI	2020	2021	IS 556	Web Design and Development	F	80 B+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
29	42528	YEREMIA SI	2020	2021	IS 404	Technopreneurship	B	82 A-	3	3	13	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
30	42528	YEREMIA SI	2020	2021	IS 513	E-Business	C	85 A	3	3	15	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
31	42528	YEREMIA SI	2020	2021	UM 142	Indonesian Language	M	86 A	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
32	42528	YEREMIA SI	2020	2021	IS 431	Visual Programming	C	58 C	3	3	13	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
33	42528	YEREMIA SI	2020	2021	IS 104	Technology and Innovation Mana	A	76 B+	3	3	14	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	
34	42528	YEREMIA SI	2020	2021	IS 660	Information Systems Research M	C	77 B+	3	3	15	04/16/2002 00:00:00	Tangerang	SMAS STRADA ST AQI S1	SMA	

<

>

Biodata mhs SI TA 2020-2024

Transkrip Mhs SI TA 2020-2024

MataKuliah

Responden

+

Ready

Accessibility: Investigate

Average: 73336.65638

Count: 34720

Sum: 2546175373

Gambar 3.3 Proses Kompilasi Beberapa File Menjadi Satu File Sumber

Selanjutnya dilakukan proses transformasi struktural terhadap data historis akademik, sebagaimana disajikan pada Gambar 3.4, menunjukkan proses transformasi struktural yang dilakukan pada data historis akademik sebelum tahap pembaruan dan perbaikan lanjutan. Pada tahap ini, format data di nilai mata kuliah diubah dari berbaris (di mana setiap baris menunjukkan satu mata kuliah) menjadi berkolom (di mana setiap baris menunjukkan satu mahasiswa secara keseluruhan). Maka dari itu informasi tentang semua mahasiswa, termasuk nilai

angka, nilai huruf, semester pengambilan, dan atribut akademik lainnya, dapat disusun dalam struktur yang konsisten dengan transformasi ini.

Item	3D Game Design & Dev	Account Planning & Mgt	Account	Accounting Funda	Advanced Exp Data	Advertising Business	Algorithm and Data	Algorithms & Data	Algorithms and Pro	Animation I	Applied	Art, Computing & C	Artificial Intell	ES&T Commu	Exp Data
4250					79	79		72							
42500					79	66	63	75						64	
42502					81			76						52	
42505					86	81		85							
42507					85	69		75							
42512					83			74							
42520					52			60						32	
42522					87	63		87							
42722					79			77							
42730					84			86						31	
42771					86			79							
42774					89			76							
42838						77		87							
42941								76							
42953								89						31	
42990						66		83							
42995								89							
42934					31			88						31	
42982	52				63			61					62	30	
42988					75	69	35	63							
42996					89			74						86	
43037								83							
43056						80		85							
43227								89							
43238						76		76							
43286						81		85							
43270								82							
43300						79		88							
43380								77							
43393								76							
43424								82						85	
43450	45							75							
43544						79	86	79						84	
43632															
43639						77		72							
43654						86	87	89					7		
43656							79	76							
43677								8							
43681						69		89							
43689						75		77						31	

Gambar 3.4 Proses Transformasi Format Data Nilai Mata Kuliah

3) Pembaruan, Transformasi Lanjutan, dan Finalisasi *Dataset*

Dataset awal yang telah berhasil dikompilasi kemudian memasuki tahap pembaruan pada tanggal 24 Oktober 2025. Tahap ini dilakukan untuk menambahkan variabel baru yang sebelumnya belum tercakup dalam dataset awal, namun diidentifikasi sebagai variabel yang memiliki relevansi penting dalam penelitian, yaitu total Satuan Kredit Semester (SKS) serta Indeks Prestasi Kumulatif (IPK) mahasiswa. Kedua variabel ini dianggap signifikan karena mampu memberikan gambaran yang lebih lengkap mengenai performa akademik mahasiswa, serta berpotensi menjadi indikator yang memengaruhi status kelulusan maupun perkembangan studi mereka. Setelah proses pembaruan, pengolahan dan penanganan data lanjutan dilakukan secara lebih mendalam pada tanggal 25 Oktober 2025. Pada tahap ini, seluruh data yang telah diperbarui kembali ditata ulang dan diverifikasi untuk memastikan bahwa struktur dataset sudah selaras dengan kebutuhan analisis. Berikut Tabel 3.4 menyajikan gambaran atribut-atribut yang tersedia dalam dataset setelah pembaruan dilakukan.

Tabel 3.4 Gambaran Data Akademik yang Sudah Diperbarui

Nama Atribut	Nomor Induk Mahasiswa (NIM)
	Nama Mahasiswa
	Angkatan
	Semester Ambil
	Kode Mata Kuliah
	Nama Mata Kuliah
	Kelas
	Nilai Angka
	Nilai Huruf
	IPS
	Jumlah Hadir
	Tanggal Lahir
	Asal Daerah
	Asal Sekolah
	Pendidikan Ayah & Ibu
	SKS
	IPK

3.4 Variable Penelitian

3.4.1 Variabel Independen

Variabel independen dalam penelitian ini mencakup berbagai faktor eksternal yang diyakini memiliki pengaruh signifikan terhadap kinerja akademik mahasiswa. Faktor-faktor tersebut meliputi faktor keluarga, faktor pilihan studi, faktor ekonomi dan pekerjaan, faktor kesehatan dan gaya hidup, faktor akademik dan dosen, serta faktor fasilitas kampus. Setiap faktor dianggap berperan dalam membentuk motivasi, kebiasaan belajar, dan kemampuan mahasiswa dalam menyelesaikan tugas akademik, sehingga dapat memberikan gambaran mengenai perbedaan pencapaian akademik antarindividu. Pengukuran variabel independen dilakukan melalui instrumen kuesioner pada periode 31 Agustus hingga 28 September 2025. Sebagian besar variabel ini diukur menggunakan skala Likert.

3.4.2 Variabel Dependen

Variabel dependen dalam penelitian ini adalah Indeks Prestasi Kumulatif (IPK) mahasiswa, yang digunakan sebagai indikator utama untuk menilai keberhasilan akademik selama masa studi. IPK merepresentasikan hasil kumulatif dari seluruh proses pembelajaran, tugas, dan evaluasi akademik yang dijalani mahasiswa, sehingga menjadi ukuran yang objektif untuk menilai performa akademik secara keseluruhan. Data *outcome* ini diambil dari catatan resmi akademik mahasiswa hingga semester terakhir pengambilan data (25 September 2025).

3.5 Teknik Analisis Data

Dalam analisis data, pemilihan perangkat lunak yang tepat menjadi aspek penting untuk mendukung efisiensi serta keakuratan proses pengolahan data. Selain itu, hasil analisis yang telah diperoleh kemudian divisualisasikan menggunakan Power BI untuk memudahkan interpretasi pola performa akademik mahasiswa dan faktor-faktor eksternal yang memengaruhinya. Jupyter Notebook dan RStudio merupakan dua *tools* yang sering digunakan dalam berbagai proyek analitik. Berikut Tabel 3.2 yang menampilkan perbandingan kedua tools tersebut.

Tabel 3.5 Perbandingan Tools Analisis Data

Aspek	Jupyter Notebook	R Studio
Dukungan bahasa	Mendukung berbagai bahasa seperti Python, R, Julia, dan lainnya, sangat fleksibel.	Fokus pada bahasa R, dengan fitur yang mendalam untuk statistik dan analisis.
Kemudahan Penggunaan	User-friendly untuk pemula, terutama dengan visualisasi yang cepat dan berbasis web.	Lebih teknis, cocok untuk pengguna dengan pengalaman statistik ataupun R.
Eksekusi Kode	Menjalankan kode secara interaktif dalam cell, dan cocok untuk eksplorasi data.	Eksekusi kode yang terintegrasi dengan alat debugging khusus untuk R.

Kolaborasi	Format notebook (.ipynb) mudah dibagikan akan tetapi tetap membutuhkan open source untuk dibuka.	Mendukung integrasi Git langsung, memudahkan kolaborasi berbasis kontrol versi.
Visualisasi	Mendukung library visualisasi seperti Matplotlib dan Plotly dengan output langsung di notebook.	Kuat untuk visualisasi berbasis R dengan library seperti ggplot2 dan shiny.
Integrasi	Kurang dalam integrasi versi (Git), namun kompatibel dengan banyak library.	Integrasi Git kuat dengan fitur seperti commit dan history yang tersedia di dalam aplikasi.

Sumber : [54], [55]

Berdasarkan hasil dari Tabel 3.2, Jupyter Notebook dipilih karena fleksibilitasnya dalam mendukung berbagai bahasa pemrograman, eksekusi kode secara interaktif, dan kemampuan visualisasi data secara cepat. Library populer seperti Pandas, NumPy, Matplotlib, dan Seaborn memudahkan pengolahan serta analisis data dari berbagai variabel. Setiap langkah analisis dapat terdokumentasi secara rapi dalam format notebook (.ipynb), sehingga mempermudah replikasi dan validasi hasil.

3.5.1 Penanganan Data dan Transformasi Awal

Proses penanganan data dilakukan secara komprehensif menggunakan bahasa pemrograman Python dengan bantuan pustaka Pandas dan NumPy pada tahap data preprocessing dan transformation. Tahap ini mencakup pembersihan data, penanganan nilai kosong, serta pembentukan variabel turunan yang dibutuhkan untuk analisis lanjutan. Dalam menangani *missing values*, beberapa langkah diterapkan. Pertama, baris yang mengandung nilai kosong pada kolom-kolom yang tidak memberikan kontribusi signifikan terhadap variabel lainnya seperti mata kuliah Pre-Final Project atau Pre-Activities dihapus (*drop rows*) menggunakan fungsi `dropna()`. Langkah ini dilakukan karena penghapusan data tersebut tidak memengaruhi analisis variabel utama, sehingga integritas dataset tetap terjaga.

Kedua, untuk kolom NILAI_HURUF yang kosong namun kolom NILAI_ANGKA bernilai 0.0, data tersebut diidentifikasi sebagai indikasi kegagalan akademik dan dilakukan imputasi sederhana dengan mengisi nilai huruf “E”. Pendekatan ini dipilih karena sifat data yang kategorikal dan memiliki makna akademik spesifik, sehingga metode imputasi lain seperti *multiple imputation* tidak relevan. Selain itu, mahasiswa yang menunjukkan ketidakterusan pengambilan mata kuliah pada semester berikutnya secara otomatis diklasifikasikan ke dalam kategori “Dropout” melalui pembentukan variabel baru bernama KELULUSAN_STATUS.

3.5.2 Transformasi Variabel dan Eksplorasi Data Awal

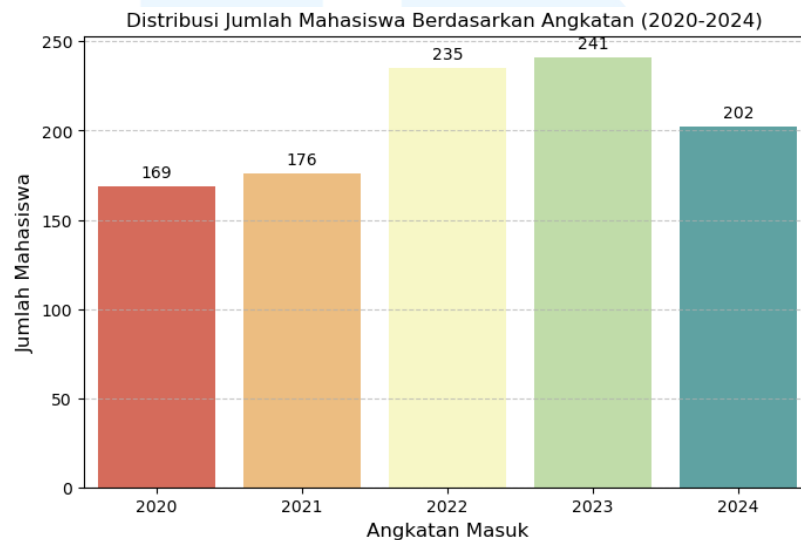
Tahap transformasi dan eksplorasi awal data dilakukan untuk memastikan seluruh variabel siap digunakan dalam analisis korelasi dan klasifikasi. Variabel prediktor dari hasil kuesioner yang menggunakan skala Likert diperlakukan sebagai data ordinal yang dapat diinterpretasikan secara kuantitatif kontinu, sehingga hubungan antarvariabel dapat dianalisis secara terukur. Tidak diterapkan proses normalisasi atau standardisasi, karena analisis yang digunakan bersifat non-parametrik dan tidak bergantung pada distribusi data.

Eksplorasi awal dilakukan melalui pemeriksaan distribusi IPK, IPS, serta proporsi mahasiswa per angkatan dan status kelulusan. Dari hasil eksplorasi, ditemukan adanya variasi nilai ekstrem yang masih tergolong wajar secara akademik, seperti mahasiswa dengan IPK sangat tinggi atau rendah pada semester tertentu. Pola *missing values* yang terdeteksi juga tidak sepenuhnya acak, melainkan terkait dengan kondisi akademik mahasiswa, misalnya tidak melanjutkan semester berikutnya atau belum menempuh seluruh mata kuliah. Oleh karena itu, proses penanganan dilakukan secara kontekstual, yaitu dengan menilai kelogisan data berdasarkan status dan riwayat studi mahasiswa.

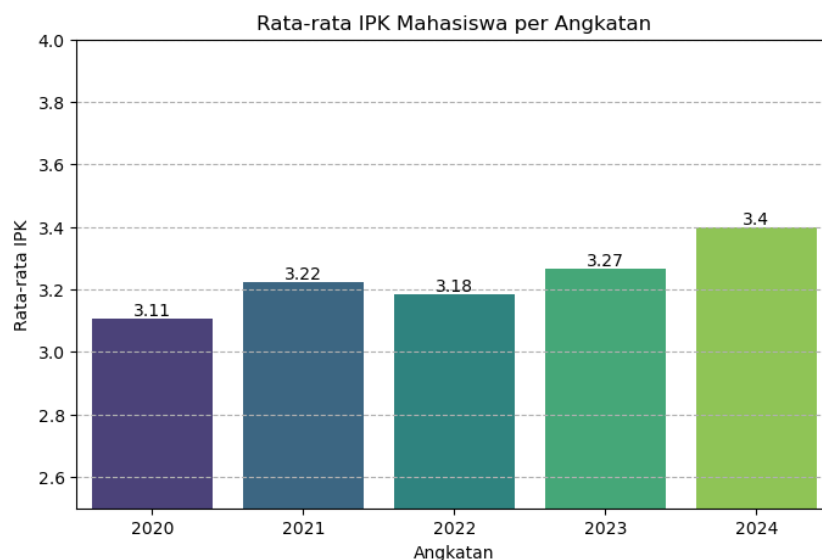
3.5.3 Analisis Deskriptif, dan Inferensial

Analisis deskriptif digunakan untuk meringkas serta memahami pola hubungan antara faktor-faktor eksternal dan performa akademik mahasiswa berdasarkan data historis yang menggambarkan perkembangan studi dari waktu

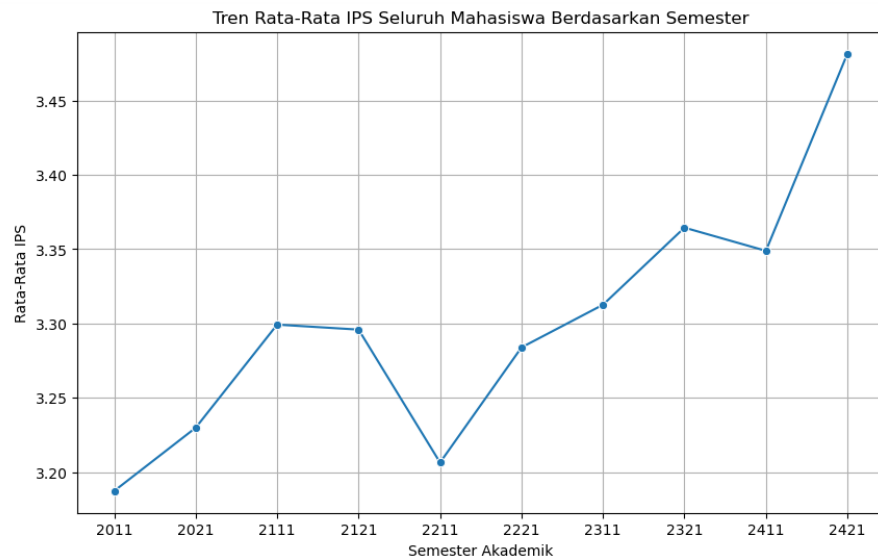
ke waktu. Tahap eksplorasi awal dilakukan melalui visualisasi deskriptif berupa *bar chart*, *line chart*, dan *pie chart*. yang menampilkan informasi seperti distribusi mahasiswa per angkatan, rata-rata IPK tiap angkatan, tren rata-rata IPS per semester, serta proporsi status kelulusan sebagaimana ditunjukkan pada Gambar 3.5 hingga Gambar 3.8. di bawah ini.



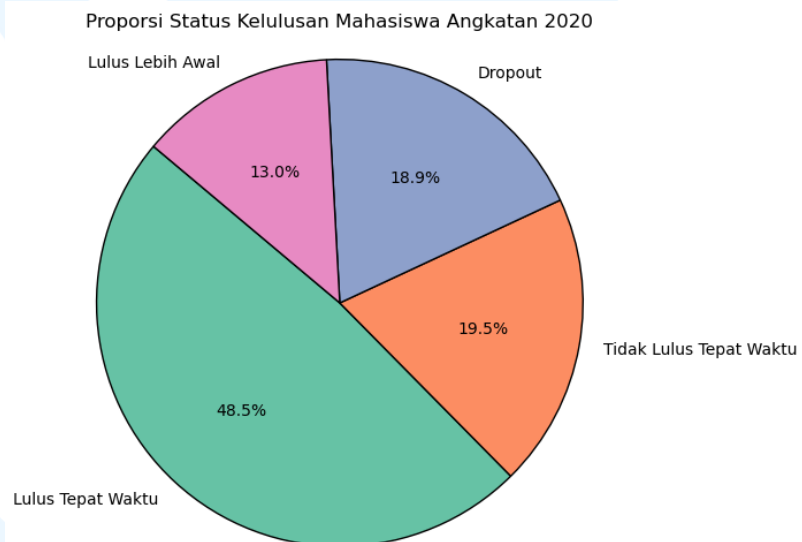
Gambar 3.5 Visualisasi Distribusi Mahasiswa



Gambar 3.6 Visualisasi Rata-Rata IPK



Gambar 3.7 Visualisasi Tren Rata-rata IPS per semester



Gambar 3.8 Visualisasi Proporsi Kelulusan

Selain itu, dilakukan juga analisis terhadap sepuluh mata kuliah dengan jumlah peserta terbanyak serta identifikasi mata kuliah wajib dengan nilai rata-rata tertinggi dan terendah untuk memperoleh pemahaman lebih dalam mengenai tingkat kesulitan dan variasi performa akademik antar mata kuliah. Hasil eksplorasi menunjukkan adanya variasi signifikan pada distribusi jumlah mahasiswa antar angkatan serta pola penurunan IPS pada beberapa semester

menengah yang mengindikasikan meningkatnya tingkat kesulitan akademik pada fase tertentu.

Tahap inferensial dilanjutkan dengan Uji Korelasi Spearman's Rho untuk menilai arah dan kekuatan hubungan antara variabel eksternal seperti kondisi ekonomi keluarga, dukungan sosial, serta kepuasan terhadap program studi dengan capaian IPK mahasiswa. Kinerja hubungan dievaluasi melalui koefisien korelasi (ρ) dan nilai signifikansi (p-value), di mana hubungan dianggap signifikan apabila $p\text{-value} < 0,05$. Selanjutnya dilakukan analisis regresi linear berganda untuk menilai besaran pengaruh faktor-faktor eksternal terhadap IPK secara numerik. Model regresi dievaluasi menggunakan R^2 untuk melihat proporsi variasi IPK yang dapat dijelaskan model, serta MAE, MSE, dan RMSE untuk menilai tingkat kesalahan prediksi. Interpretasi koefisien regresi dilakukan untuk mengetahui arah dan besar pengaruh masing-masing variabel terhadap performa akademik mahasiswa.

Selain analisis korelasi, hasil data akademik juga digunakan untuk membentuk kategori risiko performa studi, yang dikembangkan berdasarkan logika Python pada variabel `KELULUSAN_STATUS`. Mahasiswa diklasifikasikan menjadi tiga kelompok, yaitu Risiko Rendah (lulus tepat waktu atau lebih cepat), Risiko Tinggi (dropout atau tidak lulus tepat waktu), dan Aktif bagi mahasiswa yang masih menjalani masa studi yang tertampil pada Gambar 3.9.

	NIM	ANGKATAN	KELULUSAN_STATUS
0	42528	2020	Lulus Tepat Waktu
1	42544	2020	Lulus Lebih Awal
2	42599	2020	Lulus Tepat Waktu
3	42602	2020	Lulus Tepat Waktu
4	42609	2020	Lulus Tepat Waktu
...
1018	139882	2024	Dropout
1019	140188	2024	Masih Aktif
1020	140996	2024	Masih Aktif
1021	141000	2024	Masih Aktif
1022	143518	2024	Dropout

1023 rows x 3 columns

```
array(['Lulus Tepat Waktu', 'Lulus Lebih Awal', 'Dropout',  
      'Tidak Lulus Tepat Waktu', 'Masih Aktif'], dtype=object)
```

Gambar 3.9 Kategori Status Kelulusan Mahasiswa

Klasifikasi ini memperkuat interpretasi hasil korelasi dan menjadi dasar utama dalam pembangunan *dashboard descriptive analytics* untuk menampilkan pola dan tren performa akademik mahasiswa secara komprehensif. Dashboard Power BI ini kemudian digunakan sebagai media utama untuk menyajikan temuan penelitian secara deskriptif, sehingga mempermudah proses penarikan kesimpulan terhadap pola hubungan antarvariabel.

