

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Objek Penelitian**

Objek dalam penelitian ini adalah mahasiswa Program Studi Sistem Informasi di Universitas Multimedia Nusantara. Fokus utama penelitian ini adalah untuk menganalisis hubungan antara latar belakang sosial ekonomi mahasiswa dengan pencapaian akademik mereka. Data yang digunakan dalam penelitian ini diperoleh dari dua sumber utama, yaitu hasil survei langsung kepada mahasiswa serta data resmi dari pihak kampus. Dataset yang dikumpulkan mencakup informasi identitas mahasiswa, transkrip nilai akademik, serta data sosial ekonomi orang tua mahasiswa, seperti tingkat pendidikan, jenis pekerjaan, dan pendapatan keluarga. Terdapat tiga kelompok data yang dianalisis, yaitu data transkrip nilai mahasiswa angkatan 2010–2018, data transkrip nilai mahasiswa angkatan 2019–2023, dan data hasil survei sosial ekonomi.

Penelitian ini menggunakan metode Spearman Correlation untuk mengukur tingkat hubungan antara variabel sosial ekonomi dengan performa akademik mahasiswa. Metode ini dipilih karena mampu mengukur korelasi antar variabel ordinal maupun numerik secara non-parametrik, yang sesuai dengan karakteristik data yang diperoleh. Analisis dilakukan untuk melihat sejauh mana faktor sosial ekonomi, seperti pendapatan keluarga atau latar belakang pendidikan orang tua, memiliki korelasi terhadap nilai akademik mahasiswa. Hasil dari penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai pengaruh kondisi sosial ekonomi terhadap prestasi akademik mahasiswa di lingkungan pendidikan tinggi.

#### **3.2 Metode Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan mengadopsi kerangka kerja CRISP-DM. Pemilihan CRISP-DM didasarkan pada keunggulannya yang sistematis dan fleksibel, dimulai dari tahap pemahaman kebutuhan bisnis hingga penerapan model akhir [37]. Kerangka ini dianggap sesuai untuk penelitian

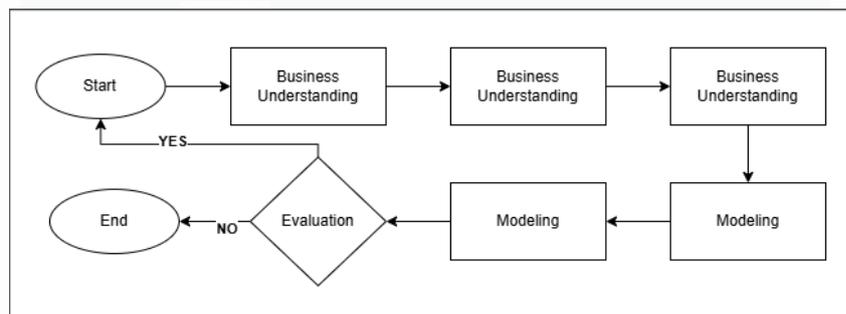
terkait prediksi harga rumah karena memungkinkan peneliti untuk mengidentifikasi masalah, melakukan proses pengolahan data, hingga mengevaluasi model secara bertahap dan terstruktur. Dengan alur yang runtut, CRISP-DM mendukung pengembangan model prediksi yang tidak hanya memiliki tingkat akurasi tinggi, tetapi juga mudah diterapkan di berbagai konteks.

CRISP-DM juga mengintegrasikan pemahaman terhadap aspek bisnis dengan analisis data secara mendalam, sehingga setiap tahapan dalam penelitian memiliki arah dan tujuan yang jelas [37]. Pemilihan kerangka kerja ini dilakukan setelah mempertimbangkan kelebihan dan kekurangan dari beberapa framework lainnya, seperti SEMMA dan KDD. Berikut ini disajikan perbandingan antara ketiga kerangka kerja tersebut untuk memperkuat alasan pemilihan CRISP-DM.

Tabel 3. 1 Perbandingan Framework Data Mining [37]

Aspek	CRISP-DM	SEMMA	KDD
<b>Fokus</b>	Proses analisis dari pemahaman bisnis hingga hasil.	Berorientasi pada eksplorasi data dan modeling.	Penemuan pola baru dari data yang besar.
<b>Tahapan</b>	<ol style="list-style-type: none"> <li>1. Business Understanding</li> <li>2. Data Understanding</li> <li>3. Data Preparation</li> <li>4. Modeling</li> <li>5. Evaluation</li> <li>6. Deployment</li> </ol>	<ol style="list-style-type: none"> <li>1. Sample</li> <li>2. Explore</li> <li>3. Modify</li> <li>4. Model</li> <li>5. Assess</li> </ol>	<ol style="list-style-type: none"> <li>1. Selection</li> <li>2. Preprocessing</li> <li>3. Transformation</li> <li>4. Data Mining</li> <li>5. Evaluation</li> </ol>
<b>Kelebihan</b>	Cocok untuk berbagai industri, terstruktur, fleksibel.	Fokus pada modeling, cocok untuk eksplorasi data.	Cocok untuk dataset besar dengan fokus penemuan pola baru
<b>Kekurangan</b>	Membutuhkan waktu lebih lama untuk setiap tahap.	Kurang memperhatikan tujuan bisnis dan tidak mencakup deployment	Kompleks untuk proyek kecil dan Kurang mendalam pada teknik implementasi

Berdasarkan Tabel 3.1, framework CRISP-DM dipilih karena memiliki pendekatan yang terstruktur sekaligus fleksibel, menjadikannya sangat sesuai untuk penelitian yang menganalisis hubungan antara faktor sosial ekonomi dan pencapaian akademik mahasiswa. Tahap *Business Understanding* membantu peneliti dalam merumuskan tujuan utama, yaitu memahami bagaimana latar belakang sosial ekonomi memengaruhi performa akademik mahasiswa. Selanjutnya, tahap *Data Understanding* memungkinkan eksplorasi lebih mendalam terhadap data mahasiswa, seperti informasi nilai, latar belakang orang tua, dan data survei, sehingga dapat diidentifikasi pola atau tren awal dalam hubungan antar variabel.



Gambar 3. 1 Alur Penelitian

Tahapan *Modeling* dan *Evaluation* memberikan ruang untuk menerapkan metode statistik yang sesuai, dalam hal ini korelasi Spearman, untuk mengukur kekuatan dan arah hubungan antar variabel sosial ekonomi dengan nilai akademik. Sementara itu, tahap *Deployment* bertujuan untuk menyajikan hasil analisis yang dapat dimanfaatkan oleh pihak kampus sebagai bahan pertimbangan dalam kebijakan akademik atau dukungan bagi mahasiswa. Meskipun tahapan dalam CRISP-DM cenderung lebih rinci dibanding framework lain seperti SEMMA dan KDD, pendekatan ini memungkinkan proses analisis yang lebih mendalam dan terarah. Oleh karena itu, CRISP-DM dianggap paling tepat dalam mengintegrasikan kebutuhan analisis data dengan tujuan penelitian secara keseluruhan. Berikut adalah tahapan-tahapan dalam framework CRISP-DM yang diterapkan dalam penelitian ini:

### **3.2.1 Business Understanding**

Tahap ini bertujuan untuk memahami latar belakang dan urgensi dari penelitian, yaitu pentingnya mengkaji hubungan antara faktor sosial ekonomi dengan pencapaian akademik mahasiswa di Universitas Multimedia Nusantara. Dalam konteks pendidikan tinggi, perbedaan latar belakang sosial ekonomi mahasiswa dapat berdampak pada kemampuan belajar, akses terhadap sumber daya pendidikan, serta hasil akademik mereka. Dengan meningkatnya perhatian terhadap kesetaraan pendidikan, pemahaman atas korelasi antara kondisi sosial ekonomi dan nilai akademik menjadi penting sebagai dasar perumusan kebijakan yang lebih adil dan tepat sasaran.

Penelitian ini menggunakan pendekatan korelasi Spearman untuk mengukur hubungan antara data sosial ekonomi seperti pekerjaan dan pendidikan orang tua dengan data akademik mahasiswa yang diperoleh dari transkrip nilai dan survei. Hasil analisis ini diharapkan dapat memberikan wawasan bagi pihak kampus dalam merancang strategi dukungan akademik yang lebih inklusif, serta membantu pihak-pihak terkait dalam memahami dinamika performa akademik berdasarkan latar belakang sosial. Dengan kerangka kerja CRISP-DM, proses penelitian dilaksanakan secara sistematis, dimulai dari pemahaman konteks kebutuhan hingga analisis hasil yang aplikatif dalam pengembangan kebijakan pendidikan.

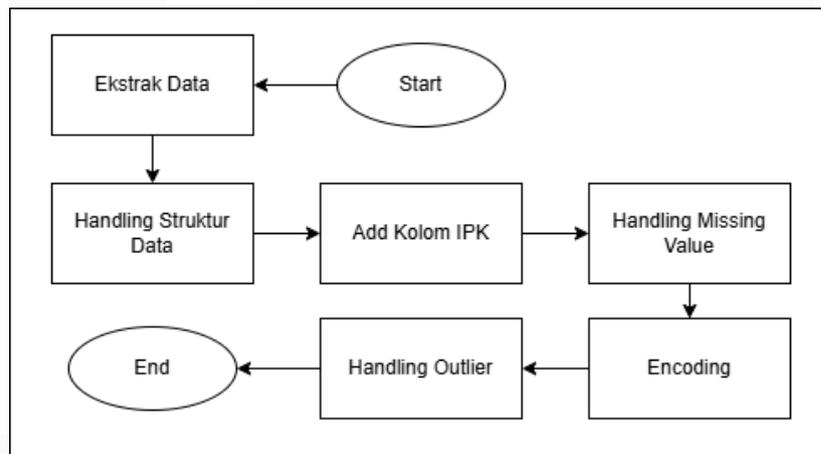
### **3.2.2 Data Understanding**

Pada tahap Data Understanding, dilakukan eksplorasi awal untuk memahami kualitas dan kelengkapan data sebelum proses analisis lebih lanjut. Data yang digunakan dalam penelitian ini terdiri dari dua jenis, yaitu data primer dan data sekunder. Data primer diperoleh melalui survei sosial ekonomi yang disebarkan kepada mahasiswa Universitas Multimedia Nusantara menggunakan Google Form ([Link Survey Google Form](#)) selama periode 2 Mei 2025 hingga 30 Mei 2025, dengan total 65 responden. Sementara itu, data sekunder berasal dari Lab Fakultas Teknik Informatik yaitu data historis sistem

internal yang berjumlah 961 baris dan 426 kolom, berupa transkrip nilai akademik mahasiswa dari tahun 2019 hingga 2024.

Eksplorasi data dilakukan melalui analisis statistik deskriptif untuk meninjau distribusi nilai akademik dan variasi kondisi sosial ekonomi, serta untuk mengidentifikasi data yang hilang atau tidak konsisten. Visualisasi data juga digunakan untuk mengenali pola awal antar variabel, seperti hubungan antara kelompok sosial ekonomi dengan capaian akademik. Pemahaman ini menjadi dasar dalam merancang proses pembersihan dan transformasi data yang diperlukan untuk mendukung akurasi model analisis korelasi.

### 3.2.3 Data Preparation



Gambar 3. 2 Diagram Alur Data Preparation

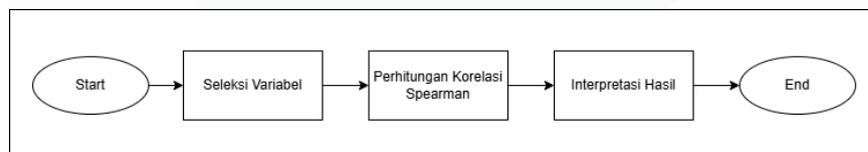
Tahap Data Preparation diawali dengan proses ekstraksi data dari dua sumber utama, yaitu data survei dan data transkrip nilai mahasiswa. Setelah data diperoleh, dilakukan handling struktur data khusus pada data transkrip nilai menggunakan Google Apps Script untuk menyusun format data menjadi lebih terstruktur. Selanjutnya, ditambahkan kolom IPK ke dalam data transkrip nilai menggunakan Python sebagai variabel utama dalam analisis.

Langkah berikutnya adalah penanganan missing value yang diterapkan hanya pada data transkrip nilai untuk memastikan tidak ada data kosong yang mengganggu proses analisis. Setelah itu, dilakukan proses encoding pada kedua jenis data (survei dan transkrip nilai) untuk mengubah variabel kategorikal

menjadi bentuk numerik yang dapat dibaca oleh model machine learning. Tahapan terakhir adalah handling outlier yang dilakukan pada data transkrip nilai menggunakan Python untuk mendeteksi dan mengatasi nilai-nilai ekstrem, sehingga kualitas dan akurasi analisis dapat lebih terjaga.

### 3.2.4 Modeling

Pada tahap modeling, dilakukan penerapan metode analisis statistik untuk mengukur hubungan antara variabel sosial ekonomi dengan performa akademik mahasiswa. Metode yang digunakan adalah Korelasi Spearman karena mampu menangani data ordinal maupun numerik dengan distribusi yang tidak harus normal. Proses modeling diawali dengan seleksi variabel yang dianggap relevan berdasarkan tahap feature selection, seperti latar belakang pendidikan orang tua, jenis profesi yang dijalani, serta tingkat pendapatan keluarga, dan prestasi akademik mahasiswa yang diukur melalui IPK atau nilai rata-rata dari transkrip akademik.



Gambar 3. 3 Diagram Alur Modeling

Setelah variabel terpilih, dilakukan perhitungan koefisien korelasi Spearman untuk masing-masing pasangan variabel sosial ekonomi dan nilai akademik. Hasil korelasi kemudian dianalisis untuk mengetahui kekuatan dan arah hubungan (positif atau negatif). Korelasi dengan nilai koefisien mendekati +1 atau -1 menunjukkan hubungan yang kuat, sedangkan nilai mendekati 0 menunjukkan hubungan yang lemah atau tidak ada hubungan signifikan.

### 3.2.5 Evaluation

Tahap evaluasi dalam penelitian ini bertujuan untuk menilai validitas dan signifikansi hasil analisis korelasi Spearman yang diperoleh pada tahap modeling. Evaluasi dilakukan dengan menguji signifikansi statistik menggunakan p-value untuk memastikan bahwa korelasi yang ditemukan tidak

terjadi secara kebetulan, dengan tingkat kepercayaan yang umum digunakan yaitu 0,05. Selain itu, kekuatan korelasi juga dianalisis berdasarkan nilai koefisien Spearman untuk mengklasifikasikan hubungan antar variabel dalam kategori sangat lemah hingga sangat kuat. Aspek ini penting untuk memahami seberapa kuat pengaruh faktor sosial ekonomi terhadap pencapaian akademik mahasiswa.

Selain menguji signifikansi dan kekuatan korelasi, tahap evaluasi juga mencakup pemeriksaan konsistensi hasil analisis dengan karakteristik data dan temuan eksplorasi awal. Hal ini bertujuan memastikan bahwa hasil korelasi yang diperoleh relevan dan sesuai dengan pola data yang ada. Jika diperlukan, analisis lanjutan secara multivariat dapat dilakukan untuk melihat pengaruh simultan beberapa variabel sosial ekonomi terhadap performa akademik. Keseluruhan proses evaluasi ini menjadi dasar yang kuat untuk menyimpulkan hubungan antara latar belakang sosial ekonomi mahasiswa dan prestasi akademik, sekaligus memastikan bahwa temuan penelitian dapat dipercaya dan memberikan kontribusi bagi pengembangan kebijakan akademik yang lebih inklusif.

### **3.2.6 Deployment**

Tahap deployment dilakukan dengan menggunakan Streamlit sebagai platform untuk menampilkan hasil analisis secara interaktif melalui web, serta GitHub sebagai repositori penyimpanan kode dan data secara terbuka. Dengan Streamlit, hasil analisis yang sebelumnya hanya berupa output statis kini dapat disajikan dalam bentuk dashboard yang intuitif dan mudah diakses. Aplikasi ini memungkinkan pengguna untuk memilih jenis data yang ingin dianalisis, mengunggah file transkrip nilai mahasiswa, menjalankan analisis korelasi Spearman, serta secara otomatis menampilkan hasil analisis dalam bentuk tabel, kesimpulan tertulis, dan visualisasi berupa heatmap yang menggambarkan kekuatan hubungan antar variabel.

Penggunaan Streamlit dalam tahap deployment mempermudah interaksi pengguna non-teknis, seperti dosen, pengambil kebijakan, atau pihak administrasi universitas, untuk mengeksplorasi dan memahami hasil penelitian tanpa harus menjalankan kode secara langsung. Di sisi lain, GitHub digunakan untuk mendokumentasikan seluruh kode pemrograman, struktur folder, serta file pendukung lain yang dibutuhkan, sehingga meningkatkan transparansi dan memungkinkan kolaborasi atau replikasi oleh peneliti lain. Dengan pendekatan ini, proses deployment tidak hanya menyampaikan hasil analisis secara informatif, tetapi juga mendorong keterbukaan data dan keberlanjutan penelitian di masa depan.

### **3.3 Teknik Pengumpulan Data**

Teknik pengumpulan data dalam penelitian ini menggunakan dua metode utama, yaitu dokumentasi dan penyebaran kuesioner. Metode dokumentasi digunakan untuk memperoleh data sekunder berupa nilai akademik mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara dari angkatan 2010 hingga 2023. Data ini mencakup Indeks Prestasi Kumulatif (IPK) dan nilai mata kuliah, yang diperoleh secara resmi dari pihak kampus sebagai representasi pencapaian akademik mahasiswa.

Sementara itu, data primer dikumpulkan melalui penyebaran kuesioner online kepada mahasiswa Universitas Multimedia Nusantara. Kuesioner ini dirancang untuk menggali informasi mengenai latar belakang sosial ekonomi responden, termasuk tingkat pendidikan orang tua, jenis pekerjaan orang tua, dan pendapatan keluarga. Teknik ini memungkinkan peneliti untuk mengidentifikasi pola hubungan antara faktor sosial ekonomi dan prestasi akademik mahasiswa.

### **3.4 Variable Penelitian**

#### **3.4.1 Variable Dependent**

Variabel dependen dalam penelitian ini adalah pencapaian akademik mahasiswa, yang diukur menggunakan nilai akademik mahasiswa berdasarkan transkrip nilai/IPK. Variabel ini dipilih karena menjadi indikator utama

performa akademik yang dapat mencerminkan keberhasilan belajar mahasiswa selama masa studi mereka.

### **3.4.2 Variable Independent**

Variabel bebas dalam penelitian ini adalah faktor sosial ekonomi mahasiswa, yang mencakup tingkat pendidikan orang tua, jenis pekerjaan orang tua, serta penghasilan keluarga. Tingkat pendidikan orang tua mencerminkan latar belakang pendidikan keluarga mahasiswa, jenis pekerjaan orang tua menggambarkan status pekerjaan serta potensi pendapatan keluarga, sedangkan pendapatan keluarga menunjukkan kemampuan ekonomi dalam mendukung kebutuhan pendidikan mahasiswa. Variabel-variabel ini dipilih karena dianggap memiliki pengaruh signifikan terhadap pencapaian akademik mahasiswa, mengingat kondisi sosial ekonomi keluarga dapat memengaruhi akses dan kualitas sumber daya pendidikan yang diterima selama proses belajar.

## **3.5 Teknik Analisis Data**

Dalam penelitian ini, teknik analisis data difokuskan pada penggunaan bahasa pemrograman yang mampu mengolah data sosial ekonomi dan performa akademik mahasiswa secara efektif dan efisien. Pemilihan bahasa pemrograman jatuh pada Python, karena keunggulannya dalam pengolahan data berskala besar serta dukungan terhadap berbagai pustaka (library) analisis statistik dan visualisasi yang sangat lengkap, seperti Pandas, NumPy, SciPy, Seaborn, dan Matplotlib [38]. Python dikenal luas dalam dunia data science karena sintaksnya yang sederhana namun powerful, sehingga mempermudah peneliti dalam mengelola data, melakukan eksplorasi, membangun model, serta menyajikan hasil secara visual dan terstruktur.

Tabel 3. 2 Perbandingan Bahasa Pemrograman [38]

Aspek	R	Phyton
Tujuan	Data science dan statistik	Research dan data science
Terbit	1993	1991
Ekosistem	19.000 Package	> 30.000 Package
Kelebihan	<ul style="list-style-type: none"> <li>• Visualisasi Grafik Menarik</li> <li>• Analisis statistik yang baik</li> <li>• Kecepatan pemrosesan</li> </ul>	<ul style="list-style-type: none"> <li>• Komunitas Besar</li> <li>• Proses data cepat</li> <li>• Mudah untuk deployment</li> </ul>
Kekurangan	<ul style="list-style-type: none"> <li>• Komunitas lebih kecil dibanding Python</li> <li>• Efisiensi memori rendah</li> </ul>	<ul style="list-style-type: none"> <li>• Pemrosesan lambat untuk data besar</li> </ul>

Dibandingkan dengan bahasa pemrograman lain seperti R, Python memiliki ekosistem yang lebih luas dengan lebih dari 30.000 paket yang mendukung analisis data dan machine learning, serta unggul dalam integrasi dan deployment aplikasi berbasis web seperti Streamlit. Python juga didukung oleh komunitas global yang besar dan aktif, memudahkan peneliti dalam mencari solusi atas permasalahan teknis selama proses analisis [38]. Dengan kecepatan pemrosesan data yang tinggi, fleksibilitas, serta dukungan pustaka statistik dan visualisasi yang lengkap, Python menjadi pilihan utama dalam penelitian ini untuk mengolah data sosial ekonomi dan prestasi akademik mahasiswa secara efisien dan akurat.