BAB III
METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif. Pendekatan kuantitatif digunakan karena data yang dianalisis berupa data numerik dan diolah secara sistematis menggunakan teknik analisis data. Menurut Sugiyono (2013), penelitian kuantitatif menekankan pada pengolahan data numerik yang objektif untuk menggambarkan dan menjelaskan suatu fenomena berdasarkan fakta yang dapat diukur secara empiris. Dengan demikian, hasil penelitian dapat digunakan untuk menarik kesimpulan secara sistematis dan terstruktur.

Jenis penelitian ini merupakan penelitian deskriptif karena bertujuan memberikan gambaran secara faktual dan mendalam mengenai pola keterlambatan pembayaran Uang Kuliah Tunggal (UKT) mahasiswa tanpa memberikan perlakuan atau manipulasi terhadap objek penelitian. Fokus penelitian adalah menganalisis karakteristik keterlambatan pembayaran berdasarkan data riwayat pembayaran mahasiswa.

Dalam penelitian ini, data yang digunakan berupa data historis pembayaran Uang Kuliah Tunggal (UKT) mahasiswa angkatan 2021, 2022, dan 2023 UIN Imam Bonjol Padang pada Tahun Akademik 2024/2025 yang mencakup Semester ganjil dan Semester genap. Data tersebut selanjutnya diolah menggunakan teknik data mining dengan algoritma *K-Means Clustering*. Atribut yang digunakan dalam proses *clustering* adalah Nominal UKT dan Selisih Keterlambatan Pembayaran, sedangkan atribut lainnya digunakan sebagai informasi pendukung dalam proses analisis dan interpretasi hasil.

Pada penelitian ini, analisis dilakukan dengan menggunakan pendekatan data mining, untuk mengelompokkan data pembayaran UKT berdasarkan kemiripan karakteristiknya. Proses *clustering* dilakukan menggunakan *Python* dengan pustaka *Scikit-learn*. Penentuan jumlah *cluster* dilakukan dengan menguji beberapa nilai *k* dan mengevaluasi hasil pengelompokan menggunakan

UIN IMAM BONJOL PADANG

metode *Silhouette Coefficient* untuk menentukan jumlah *cluster* yang optimal. Hasil pengelompokan selanjutnya dianalisis dan divisualisasikan untuk memperoleh informasi mengenai karakteristik dan pola keterlambatan pembayaran UKT mahasiswa yang dapat digunakan sebagai bahan evaluasi bagi pihak terkait.

B. Setting Penelitian

1) Lokasi Penelitian

Universitas Islam Negeri (UIN) Imam Bonjol Padang dipilih sebagai lokasi penelitian. Perguruan tinggi tersebut merupakan institusi pendidikan tinggi negeri yang berada di bawah Kementerian Agama Republik Indonesia. UIN Imam Bonjol Padang memiliki sistem administrasi akademik dan keuangan yang terintegrasi, termasuk dalam pengelolaan data mahasiswa dan pembayaran Uang Kuliah Tunggal (UKT). Pemilihan UIN Imam Bonjol Padang sebagai lokasi penelitian didasarkan pada ketersediaan data administrasi keuangan mahasiswa yang relevan dengan tujuan penelitian. Data pembayaran UKT yang tercatat secara periodik pada setiap semester menyediakan informasi numerik yang dapat diolah dan dianalisis menggunakan metode data mining, khususnya algoritma *K-Means Clustering*.

Selain itu, UIN Imam Bonjol Padang memiliki data pembayaran UKT mahasiswa dari berbagai program studi yang dapat digunakan untuk mendukung proses analisis pola pembayaran. Dengan kondisi tersebut, lokasi penelitian dinilai sesuai dengan kebutuhan data dan tujuan penelitian yang telah ditetapkan.

2) Jadwal Penelitian

Penelitian ini dilaksanakan dalam rentang waktu kurang lebih empat bulan. Jadwal penelitian disusun secara sistematis agar setiap tahapan penelitian dapat berjalan secara terencana, mulai dari identifikasi masalah, studi literatur, pengumpulan data, pemahaman data *(data understanding)*, persiapan data *(data preparation)*, proses *K-Means Clustering*, evaluasi hasil pengelompokan, *deployment* menggunakan *Streamlit*, hingga penyusunan laporan akhir. Rincian jadwal pelaksanaan penelitian disajikan pada Tabel 3.1.

UIN IMAM BONJOL PADANG

Tabel 3. 1 Jadwal Penelitian

Kegiatan	Bulan 1				Bulan 2				Bulan 3				Bulan 4			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Identifikasi Masalah																
Studi Literatur																
Pengumpulan Data																
Data *Understanding*																
Data *Preparation*																
Proses *Clustering K-Means*																
Evaluasi Hasil																
Deployment *(Streamlit)*																
Penyusunan Laporan																

3) Bahan atau Materi Penelitian

Bahan atau materi penelitian merupakan data dan informasi yang digunakan dalam proses analisis guna mencapai tujuan penelitian. Pada penelitian ini, bahan penelitian berupa data administratif mahasiswa yang diperoleh dari unit terkait di lingkungan Rektorat UIN Imam Bonjol Padang yang mengelola sistem pembayaran Uang Kuliah Tunggal (UKT). Adapun bahan atau materi yang digunakan dalam penelitian ini meliputi:

a. Data historis pembayaran Uang Kuliah Tunggal (UKT) mahasiswa angkatan 2021, 2022, dan 2023 yang mencakup informasi status dan riwayat pembayaran pada Semester Ganjil dan Semester Genap Tahun Akademik 2024/2025.

UIN IMAM BONJOL PADANG

b. Data pendukung berupa atribut identitas akademik mahasiswa yang digunakan untuk mendukung proses analisis dan interpretasi hasil *clustering*.
c. Literatur pendukung berupa buku dan jurnal ilmiah yang berkaitan dengan data mining, CRISP-DM, teknik *clustering*, algoritma *K-Means*, normalisasi data, dan metode evaluasi *Silhouette Coefficient*.

4) Alat/Instrumen Penelitian

Alat/instrumen penelitian merupakan perangkat yang digunakan untuk mendukung pelaksanaan penelitian, khususnya dalam proses pengolahan, analisis, dan penyajian data. Penelitian ini menggunakan data sekunder berupa data administrasi pembayaran UKT mahasiswa, sehingga tidak menggunakan instrumen pengukuran langsung seperti kuesioner atau tes. Adapun alat yang digunakan dalam penelitian ini terdiri atas perangkat keras *(hardware)* dan perangkat lunak *(software)* sebagai berikut:

1) Perangkat keras (*Hardware*)

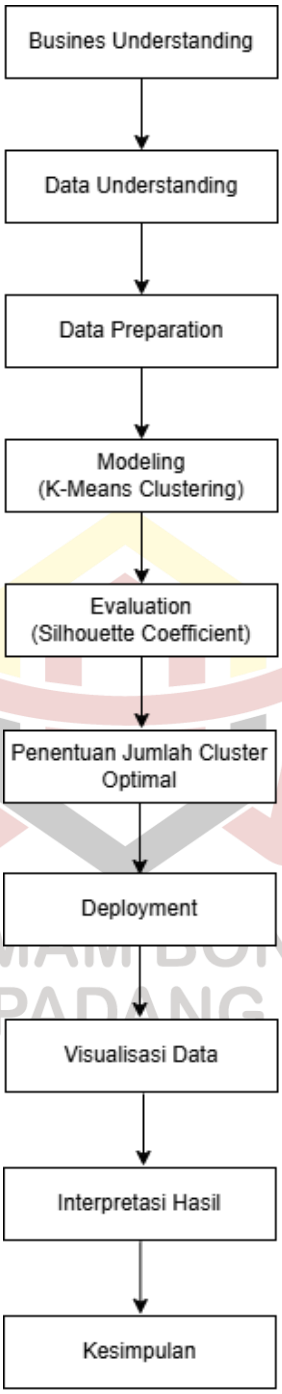
a. Laptop Acer TravelMate P214-53 / Desktop Acer
b. Processor Intel Core i3
c. RAM 8 GB
d. Penyimpanan HDD / SSD ≥ 256 GB
e. Mouse

2) Perangkat Lunak (*Software*)

a. Sistem Operasi Windows 10 / 11 (64-bit)
b. Microsoft Excel → *preprocessing* & rekap data
c. Draw.io
d. Python → pengolahan data, implementasi algoritma *K-Means*, dan evaluasi hasil *clustering*. *Scikit-learn, NumPy, Pandas, Matplotlib*.
e. *Google Colab* → *Platform* berbasis *cloud* untuk menulis, menjalankan, dan mendokumentasikan kode *Python*.
f. Streamlit digunakan untuk mengembangkan *dashboard* visualisasi hasil *clustering*.
g. Microsoft Word → penyusunan laporan penelitian

UIN IMAM BONJOL PADANG

C. Alur Penelitian



UIN IMAM BONJOL
PADANG

Gambar 3. 1 Kerangka Kerja Penelitian

D. Teknik Pengumpulan Data/Prosedur Penelitian

1) Teknik Pengumpulan Data

Teknik pengumpulan data merupakan metode yang digunakan untuk memperoleh informasi yang akan menjadi dasar analisis dalam penelitian (Sugiyono, 2013). Teknik ini berperan penting dalam memastikan bahwa data yang diperoleh sesuai dengan kebutuhan penelitian dan dapat mendukung tujuan yang ingin dicapai.

Dalam penelitian ini, pengumpulan data dilakukan melalui studi dokumentasi dengan memanfaatkan dokumen yang telah tersedia dan tersimpan pada instansi terkait. Data diperoleh melalui penelaahan terhadap dokumen tersebut tanpa melibatkan pengukuran maupun interaksi secara langsung dengan responden. Data dalam penelitian ini merupakan data historis pembayaran Uang Kuliah Tunggal (UKT) mahasiswa UIN Imam Bonjol Padang yang diperoleh secara langsung dari bagian atau unit administrasi keuangan kampus. Data tersebut mencakup riwayat pembayaran mahasiswa angkatan 2021, 2022 dan 2023 pada Tahun Akademik 2024/2025 untuk Semester ganjil dan Semester Genap.

Dataset yang diperoleh berisi beberapa atribut penting terkait pembayaran UKT, antara lain:

UIN IMAM BONJOL PADANG

a. Angkatan
b. Fakultas
c. Program Studi
d. Periode
e. Metode Pembayaran
f. Channel
g. Nominal
h. Tanggal Jatuh Tempo
i. Tanggal Pembayaran
j. Selisih Keterlambatan
k. Status Pembatalan

Data yang dikumpulkan digunakan sebagai dasar dalam proses pengolahan dan analisis menggunakan metode *clustering K-Means* untuk mengidentifikasi pola keterlambatan pembayaran UKT secara objektif.

2) Prosedur Penelitian

Prosedur penelitian menggambarkan tahapan yang dilaksanakan secara terarah untuk mencapai tujuan penelitian. Penelitian ini menerapkan kerangka kerja CRISP-DM *(Cross Industry Standard Process for Data Mining)* sebagai acuan dalam proses data mining. Kerangka kerja ini digunakan untuk mengarahkan proses penelitian mulai dari pemahaman permasalahan, pemahaman dan persiapan data, pemodelan, evaluasi, hingga penyajian hasil analisis (Hasanah dkk., 2021). Adapun tahapan prosedur penelitian yang dilakukan adalah sebagai berikut:

a. *Business Understanding*

Tahap *Business Understanding* dilakukan untuk mengidentifikasi permasalahan dan menetapkan tujuan analisis terhadap data historis pembayaran Uang Kuliah Tunggal (UKT) mahasiswa UIN Imam Bonjol Padang. Data pembayaran UKT yang telah terdokumentasi digunakan untuk memperoleh informasi mengenai pola pembayaran mahasiswa melalui pendekatan data mining. Pada tahap ini ditetapkan penggunaan algoritma *K-Means Clustering* untuk mengelompokkan data berdasarkan kemiripan karakteristik pembayaran sebagai dasar dalam menganalisis pola keterlambatan pembayaran UKT.

b. *Data Understanding*

Tahap *Data Understanding* dilakukan untuk memahami karakteristik data yang digunakan dalam penelitian. Data yang digunakan berupa data historis pembayaran UKT mahasiswa angkatan 2021, 2022, dan 2023 pada Tahun Akademik 2024/2025 Semester Ganjil dan Semester Genap yang diperoleh dari unit administrasi keuangan UIN Imam Bonjol Padang. Pada tahap ini dilakukan pemeriksaan struktur dan tipe data, kelengkapan data dan *missing value*, analisis statistik deskriptif awal, serta peninjauan atribut yang relevan untuk proses *clustering.* Atribut utama yang digunakan dalam proses *clustering* adalah Nominal UKT dan Selisih Keterlambatan, sedangkan atribut lainnya digunakan sebagai informasi pendukung dalam interpretasi hasil.

UIN IMAM BONJOL PADANG

c. *Data Preparation*

Tahap *Data Preparation* dilakukan untuk mempersiapkan data sebelum digunakan dalam proses *clustering*. Atribut yang digunakan dalam pembentukan *cluster* adalah Nominal UKT dan Selisih Keterlambatan. Kedua atribut tersebut dinormalisasi menggunakan metode *Min-Max Scaling* agar berada pada rentang skala yang sama sebelum digunakan sebagai input pada algoritma *K-Means*. Adapun tahapan yang dilakukan pada proses *data preparation* meliputi:

a. Seleksi atribut
b. Pembersihan data *(data cleaning)*
c. Transformasi data
d. Normalisasi data

Langkah-langkah tersebut dilakukan untuk memastikan dataset berada dalam format dan skala yang sesuai sebelum digunakan pada tahap pemodelan.

d. Modeling

Tahap *Modeling* dilakukan dengan menerapkan algoritma *K-Means Clustering* terhadap data yang telah melalui tahap *Data Preparation*. Proses *clustering* dilakukan menggunakan atribut Nominal UKT dan Selisih Keterlambatan yang telah dinormalisasi. Pada tahap ini diuji empat alternatif jumlah *cluster,* yaitu *k*=2, *k*=3, *k*=4, dan *k*=5. Algoritma *K-Means* dijalankan pada setiap alternatif nilai *k* hingga mencapai kondisi konvergen dan menghasilkan label *cluster* pada setiap data. Hasil dari masing-masing alternatif selanjutnya digunakan pada tahap *Evaluation.*

e. Evaluation

Tahap *Evaluation* dilakukan untuk mengevaluasi kualitas hasil *clustering* dari setiap alternatif jumlah *cluster* yang telah diuji pada tahap *Modeling.* Evaluasi dilakukan menggunakan *Silhouette Coefficient* pada *k*=2, *k*=3, *k*=4, dan *k*=5. Nilai *Silhouette Coefficient* dari setiap alternatif kemudian dibandingkan, dan jumlah *cluster* dengan nilai *Silhouette Coefficient* tertinggi ditetapkan sebagai jumlah *cluster* optimal di antara kandidat yang

UIN IMAM BONJOL PADANG

diuji. Hasil *cluster* optimal selanjutnya digunakan dalam analisis karakteristik dan pola pembayaran UKT.

f. Deployment

Tahap *Deployment* dilakukan dengan menyajikan dan menginterpretasikan hasil *clustering* yang telah diperoleh. Pada tahap ini dilakukan analisis karakteristik masing-masing *cluster,* pola waktu pembayaran berdasarkan kategori lebih awal, tepat waktu, dan terlambat, serta *profiling* berdasarkan Angkatan sebagai informasi pendukung. Hasil analisis kemudian disajikan dalam bentuk visualisasi dan *dashboard* berbasis *Streamlit* yang memuat ringkasan hasil *clustering*, distribusi *cluster*, karakteristik *cluster*, pola pembayaran, *profiling*, dan detail data hasil *clustering*. *Dashboard* yang dikembangkan bersifat *prototipe* sebagai media penyajian hasil analisis dan belum terintegrasi secara langsung dengan sistem pembayaran UKT universitas. Adapun bentuk visualisasi yang disajikan pada *dashboard* ini meliputi:

a. Grafik distribusi jumlah data pada setiap *cluster*
b. *Scatter plot* hasil *clustering* untuk memperlihatkan pola persebaran data berdasarkan *cluster* yang terbentuk.
c. Grafik pola ketepatan waktu pembayaran UKT berdasarkan kategori pembayaran lebih awal, tepat waktu, dan terlambat.
d. Ringkasan karakteristik masing-masing *cluster* mahasiswa

E. Teknik Analisis Data

Teknik analisis data merupakan serangkaian langkah yang digunakan untuk mengolah dan menganalisis data pembayaran Uang Kuliah Tunggal (UKT) mahasiswa angkatan 2021, 2022, dan 2023 UIN Imam Bonjol Padang Tahun Akademik 2024/2025 Semester ganjil dan Semester genap untuk memperoleh pola pengelompokan berdasarkan kemiripan karakteristik pembayaran. Analisis dilakukan secara kuantitatif menggunakan teknik data mining dengan algoritma *K-Means Clustering*. Proses analisis dilakukan dengan bantuan bahasa pemrograman *Python* menggunakan beberapa pustaka pendukung, seperti *Pandas*,

UIN IMAM BONJOL PADANG

NumPy, *Scikit-learn*, dan *Matplotlib*. Hasil analisis selanjutnya disajikan dalam bentuk *dashboard* visualisasi menggunakan *Streamlit* (Dhewayani dkk., 2022).

1) Analisis Menggunakan Algoritma *K-Means Clustering*

Analisis *clustering* dilakukan menggunakan atribut Nominal UKT dan Selisih Keterlambatan yang telah melalui tahap persiapan dan normalisasi data. Proses *K-Means* dilakukan dengan menguji empat alternatif jumlah *cluster*, yaitu k=2, k=3, k=4, dan k=5. Adapun tahapan analisis menggunakan algoritma *K-Means* adalah sebagai berikut:

a)* Menyiapkan Data *Clustering

Data yang digunakan dalam proses *clustering* adalah atribut Nominal UKT dan Selisih Keterlambatan yang telah dinormalisasi menggunakan metode *Min-Max Scaling*. Data hasil normalisasi selanjutnya digunakan sebagai input dalam proses *K-Means Clustering.*

b) Menentukan Kandidat Jumlah *Cluster* (*k*)

Jumlah *cluster* yang diuji dalam penelitian ini terdiri atas empat alternatif, yaitu k=2, k=3, k=4, dan k=5. Model *K-Means* dibentuk untuk setiap nilai k, kemudian hasil pengelompokan masing-masing kandidat dievaluasi menggunakan *Silhouette Coefficient.*

c) Inisialisasi Titik Pusat Awal (*Centroid*)

Inisialisasi titik pusat awal *(centroid)* dilakukan secara otomatis menggunakan algoritma *K-Means* pada pustaka *Scikit-learn*. Centroid awal digunakan sebagai acuan dalam proses pengelompokan data pada setiap kandidat jumlah *cluster*.

d) Perhitungan Jarak Menggunakan *Euclidean Distance*

Setelah proses inisialisasi titik pusat awal *(centroid)* dilakukan, jarak setiap data terhadap *centroid* dihitung menggunakan *Euclidean Distance* sebagaimana mengacu pada Persamaan (2.1). Data kemudian dialokasikan ke dalam *cluster* berdasarkan *centroid* dengan jarak terdekat.

UIN IMAM BONJOL PADANG

e) Pembaruan *Centroid*

Setelah setiap data memperoleh keanggotaan *cluster*, nilai *centroid* diperbarui berdasarkan rata-rata data pada masing-masing *cluster* sebagaimana mengacu pada Persamaan (2.2).

f) Iterasi Proses *Clustering*

Proses perhitungan jarak, penentuan keanggotaan *cluster*, dan pembaruan *centroid* dilakukan secara berulang hingga mencapai kondisi konvergen. Hasil akhir proses *K-Means* berupa label *cluster* pada setiap data untuk masing-masing kandidat jumlah *cluster*.

2) Evaluasi Hasil *Clustering*

Hasil *clustering* pada setiap kandidat jumlah *cluster* dievaluasi menggunakan *Silhouette Coefficient*. Evaluasi dilakukan terhadap hasil *K-Means* pada k=2, k=3, k=4, dan k=5 dengan mengacu pada Persamaan (2.4) sampai Persamaan (2.7). Nilai *Silhouette Coefficient* dari masing-masing kandidat dibandingkan, kemudian kandidat dengan nilai tertinggi ditetapkan sebagai jumlah *cluster* optimal di antara kandidat yang diuji. Hasil *cluster* optimal selanjutnya digunakan untuk analisis karakteristik dan pola pembayaran UKT.

3) Analisis dan Interpretasi Hasil *Clustering*

Hasil *clustering* dengan jumlah *cluster* optimal selanjutnya dianalisis secara deskriptif untuk mengetahui karakteristik masing-masing *cluster* dan pola pembayaran UKT mahasiswa. Analisis dilakukan terhadap jumlah anggota *cluster*, rata-rata Nominal UKT, rata-rata Selisih Keterlambatan, serta distribusi kategori pembayaran lebih awal, tepat waktu, dan terlambat. Selain itu, dilakukan *profiling* berdasarkan Angkatan sebagai atribut pendukung untuk memberikan gambaran pola pembayaran pada masing-masing angkatan.

4) *Deployment* Model

Hasil analisis *clustering* selanjutnya disajikan melalui dashboard berbasis web menggunakan *Streamlit. Dashboard* digunakan sebagai media visualisasi dan interpretasi hasil analisis yang memuat ringkasan dataset, hasil *clustering*, nilai *Silhouette Coefficient*, distribusi anggota *cluster*, *scatter plot* hasil

UIN IMAM BONJOL PADANG

clustering, karakteristik masing-masing *cluster*, pola waktu pembayaran, *profiling* berdasarkan Angkatan, serta detail data hasil *clustering*. *Dashboard* yang dikembangkan bersifat *prototipe* dan belum terintegrasi secara langsung dengan sistem pembayaran UKT universitas.

UIN IMAM BONJOL
PADANG