

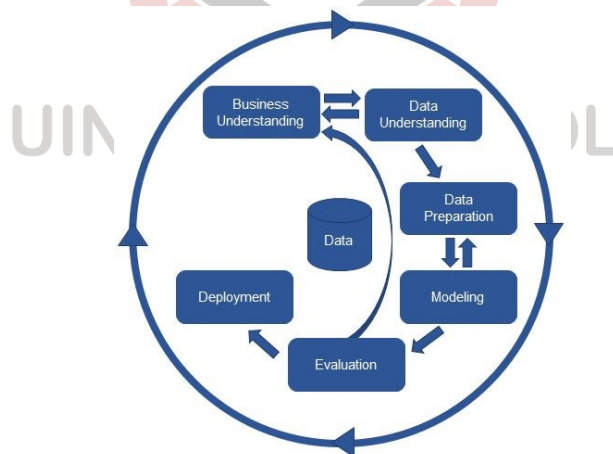
BAB III METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini termasuk dalam jenis penelitian kuantitatif deskriptif dengan pendekatan studi kasus. Metode kuantitatif umumnya digunakan untuk merangkum data, menghitung rata-rata, mengidentifikasi pola, membuat prediksi, serta menguji hubungan sebab-akibat. Tujuan penelitian ini adalah menganalisis data anggota KPRI MAN 2 Kota Payakumbuh dengan memanfaatkan algoritma Regresi Logistik untuk mengklasifikasikan tingkat kelayakan peminjaman.

B. Prosedur dan Teknik Analisis Data

Penelitian ini menerapkan metode Cross-Industry Standard Process for Data Mining (CRISP-DM) sebagai pendekatan utama dalam pengolahan data dan pengembangan model. CRISP-DM terdiri dari enam tahapan, sebagaimana ditunjukkan pada Gambar 3.1, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, serta tahap akhir berupa penyebaran hasil data mining (Amri, 2020).



Gambar 3. 1 Tahap CRIPS-DM

1. Pemahaman Bisnis (*Business Understanding*)

Tahap awal bertujuan untuk memahami permasalahan penelitian dan menetapkan tujuan yang ingin dicapai. Oleh karena itu, penelitian melakukan riset langsung di KPRI MAN 2 Kota Payakumbuh untuk merumuskan masalah utama, mengidentifikasi kebutuhan data, serta

menentukan kriteria keberhasilan yang relevan. Dalam proses ini, dilakukan identifikasi terhadap permasalahan yang dihadapi oleh koperasi, seperti kriteria penilaian kelayakan peminjaman yang belum terstruktur dengan baik (Apsari, 2024).

Langkah selanjutnya menentukan tujuan untuk Pengelompokan data yang diambil dari excel data tersebut hanya digunakan untuk menganalisis kelayakan pada koperasi MAN 2 payakumbuh. Peneliti mengidentifikasi variabel-variabel yang memengaruhi kelayakan peminjaman, seperti status anggota dan jumlah peminjaman. Setelah itu, dilakukan pemahaman teori yang cocok terkait penggunaan dalam analisis kelayakan peminjaman. Peneliti mengacu pada berbagai sumber seperti jurnal ilmiah dan laporan penelitian untuk membangun landasan teori yang kuat. Tujuannya adalah untuk mengetahui perkembangan terkini dalam analisis kelayakan peminjaman dan menentukan algoritma dan pengujian yang relevan. Berdasarkan hasil identifikasi ini, solusi yang akan ditawarkan adalah penerapan metode regresi logistik untuk menilai kelayakan peminjaman.

2. Pemahaman Data (*Data Understanding*)

Tahap ini bertujuan untuk mengeksplorasi data yang akan digunakan. Aktivitas yang dilakukan meliputi pengumpulan, pengidentifikasian, dan pemahaman data yang akan digunakan pada penelitian ini (Hikmah dkk., 2022). Proses yang dilakukan meliputi:

a. Observasi

Observasi dilakukan dengan mengunjungi langsung kantor KPRI MAN 2 Kota Payakumbuh untuk memperoleh data historis terkait peminjaman anggota. Dalam proses observasi, peneliti mengamati secara langsung bagaimana data peminjaman anggota dicatat dan dikelola. Metode ini memungkinkan peneliti untuk memahami alur kerja dan proses administrasi yang diterapkan oleh koperasi. Observasi juga membantu dalam mengidentifikasi potensi kekurangan dalam pengelolaan data yang mungkin memengaruhi

analisis selanjutnya. Selain itu, data yang diperoleh melalui observasi ini akan menjadi landasan utama untuk membangun model regresi logistik.

b. Wawancara

Wawancara dilakukan dengan ketua koperasi untuk menggali informasi mendalam mengenai kriteria kelayakan peminjaman yang diterapkan. Melalui wawancara, peneliti dapat memperoleh penjelasan tentang faktor-faktor yang dianggap penting dalam menentukan apakah anggota layak mendapatkan pinjaman, seperti status anggota, riwayat pinjaman, jumlah pinjaman dan kemampuan pembayaran. Informasi ini akan membantu peneliti untuk menentukan variabel-variabel independen yang relevan dan digunakan dalam model regresi logistik. Wawancara juga memberikan wawasan kualitatif yang tidak selalu tersedia dalam data kuantitatif.

c. Studi Literatur

Studi literatur dilakukan dengan membaca dan menganalisis berbagai sumber informasi, seperti artikel jurnal dan penelitian terdahulu yang relevan dengan penggunaan regresi logistik. Peneliti juga mempelajari studi-studi sejenis yang telah menggunakan metode ini untuk menilai kelayakan peminjaman dalam konteks yang serupa. Studi literatur ini bertujuan untuk memahami teori dasar regresi logistik, memilih metode evaluasi yang sesuai, dan mendapatkan gambaran tentang variabel-variabel umum yang sering digunakan dalam penelitian terkait. Selain itu, dengan mempelajari literatur, peneliti dapat mengidentifikasi celah penelitian sebelumnya yang akan menjadi dasar pembenaran untuk melakukan studi ini.

3. Persiapan Data (*Data Preparation*)

Pada tahap ini, data yang telah dikumpulkan akan diproses agar siap digunakan dalam pemodelan. Proses ini melibatkan pembersihan data dari anomali, normalisasi data, transformasi atribut, pengkodean data kategorik, serta seleksi atribut yang relevan.

a. *Data Cleaning*

Data cleaning adalah langkah untuk menyusun dan memperbaiki data yang tidak tepat, memiliki format yang salah, atau berantakan. Tujuan dari data cleaning adalah menghapus kesalahan, ketidak konsistenan, dan ketidakakuratan yang mungkin ada dalam kumpulan data untuk memastikan kualitas data sebelum digunakan dalam analisis. (Macfud dkk., 2023).

Pada penelitian ini, langkah pertama adalah menggabungkan data untuk membentuk dataset yang dibutuhkan. Adapun atribut terdiri dari status anggota, jumlah simpanan, status pinjaman, jumlah pinjaman, dan jangka waktu sebagai variabel independen dan atribut kelayakan sebagai variabel dependen. Setelah itu, dilakukan proses melengkapi data dengan mengimputasi nilai-nilai yang hilang (*missing values*). Selanjutnya, dilakukan penanganan data outlier, yaitu data yang memiliki nilai yang tidak wajar dibandingkan dengan data lainnya. Data *outlier* ini diidentifikasi dan ditangani dengan cara menghapus atau menyesuaikan nilainya. Proses data cleaning ini penting untuk menghilangkan potensi noise dalam dataset, sehingga dapat meningkatkan keakuratan dan validitas hasil analisis.

b. *Data Transformasi*

Transformasi data adalah langkah penting dalam proses analisis untuk memastikan data siap digunakan oleh algoritma regresi logistik. Pada penelitian ini, transformasi dilakukan dengan beberapa tahapan. Pertama, variabel kategorik seperti Status Anggota dan Status Pinjaman diubah menjadi representasi numerik menggunakan metode label encoding. Misalnya, untuk Status Anggota, kategori "Biasa" diberi nilai 1, sedangkan "Luar Biasa" diberi nilai 0. Hal serupa dilakukan pada Status Pinjaman, di mana kategori "New Order" diberi nilai 1 dan "Repeat Order" diberi nilai 0. Kedua, variabel dependen Kelayakan yang menjadi label dalam

model dikonversi menjadi data numerik biner, yaitu "Layak" diberi nilai 1 dan "Tidak Layak" diberi nilai 0. Transformasi ini memastikan format data konsisten dengan pemeriksaan untuk memastikan tidak ada nilai yang salah atau tidak sesuai, seperti huruf pada kolom numerik.

c. *Data Splitting*

Data Splitting adalah langkah penting dalam proses analisis untuk membagi dataset menjadi beberapa bagian, sehingga dapat digunakan secara efektif untuk pelatihan dan evaluasi model. Dalam penelitian ini, dari total 337 data, 273 data digunakan sebagai data *training*, sedangkan 64 data digunakan sebagai data *testing*. Untuk meningkatkan validasi model, proses pembagian data training dilakukan menggunakan teknik *K-Fold Cross-Validation*. Teknik ini membagi data training menjadi lima bagian (*fold*) yang kurang lebih memiliki jumlah data yang sama, yaitu sekitar 54–55 data per *fold*. Langkah-langkah pembagian data menggunakan *5-Fold Cross-Validation* adalah sebagai berikut:

1) Pembagian Data

Dataset training dibagi menjadi lima *fold*. Pada setiap iterasi, satu *fold* akan digunakan sebagai data validasi, sementara empat *fold* lainnya digunakan untuk melatih model.

2) Iterasi

Proses pelatihan dan validasi dilakukan sebanyak lima kali, dengan setiap *fold* bergiliran menjadi data validasi.

3) Evaluasi Kinerja

Setelah lima iterasi selesai, performa model dari setiap iterasi (misalnya, akurasi, *precision*, *recall*, dan *F1-score*) dirata-rata untuk mendapatkan hasil evaluasi yang lebih representatif.

Dengan menggunakan *5-Fold Cross-Validation*, model dapat memanfaatkan data *training* secara optimal karena setiap data akan digunakan sebagai data validasi satu kali dan data pelatihan empat

kali. Teknik ini juga membantu mengurangi risiko *overfitting* dan memberikan hasil evaluasi yang lebih andal sehingga menghasilkan model *Logistic Regression* yang lebih *robust* dan *generalizable* ketika diujikan pada 64 data *testing*.

Data *training* berfungsi untuk membantu model mempelajari pola dan hubungan antara variabel independen dengan variabel dependen. Data *testing*, di sisi lain, digunakan untuk mengevaluasi kemampuan model dalam membuat prediksi terhadap data baru yang tidak pernah dilihat sebelumnya selama proses pelatihan. Proses pembagian data dilakukan dengan menggunakan data tahun 2023 dan 2024 sebagai data *training* lalu data Januari sampai Mei 2025 sebagai data *testing*.

4. Pemodelan

Pemodelan algoritma menggunakan regresi logistik adalah teknik penting dalam klasifikasi data mining untuk prediksi dan nilai pengaruh variabel, terutama dalam konteks penilaian kelayakan peminjaman anggota KPRI. Algoritma ini memiliki pendekatan berbasis probabilitas untuk menangani data dan sering digunakan untuk mengevaluasi performa model dalam berbagai aplikasi. Berikut ini merupakan alur atau skema yang digunakan dalam penerapan regresi logistik.

Adapun tahapan dari alur skema metode adalah sebagai berikut:

- a. *Input Library*: Tahap pertama adalah mengimpor pustaka yang akan digunakan untuk analisis, seperti *pandas*, *numpy*, *matplotlib*, dan *scikit-learn*, yang memfasilitasi pengolahan data, pemodelan, dan visualisasi.
- b. *Input Dataset*: Memasukkan dataset historis peminjaman anggota KPRI yang akan digunakan untuk pelatihan dan pengujian model.
- c. *Proses Preprocessing*: Melakukan preprocessing data seperti data cleaning untuk menghapus data tidak lengkap atau tidak relevan, transformasi data untuk mengubah format variabel menjadi sesuai, dan data splitting untuk menentukan dataset menjadi data latih dan data uji.

- d. Pemodelan dengan Algoritma Regresi Logistik: Menggunakan algoritma regresi logistik untuk membangun model yang dapat memprediksi kelayakan peminjaman berdasarkan data historis dan memodelkan hubungan antara variabel dependent dan variabel independent.
- e. Evaluasi Model: Melakukan evaluasi model dengan metrik seperti uji wald, *confusion matrix*, akurasi, *precision*, *recall*, *F1-score*, nilai *binary cross entropy* (BCE) dan visualisasi *ROC Curve* untuk menilai performa model dalam memprediksi kelayakan peminjaman.
- f. Output Hasil Pemodelan Output yang dihasilkan meliputi visualisasi hasil prediksi, seperti grafik *ROC Curve*, nilai BCE dan metrik akurasi yang menunjukkan performa model.

Hasil dari tahapan ini memberikan informasi tentang kemampuan model regresi logistik dalam melakukan klasifikasi kelayakan peminjaman. Model yang optimal akan digunakan untuk pengujian lebih lanjut.

5. Evaluasi

Evaluasi model dalam penelitian ini dilakukan untuk menilai performa model regresi logistik dalam menilai kelayakan peminjaman anggota KPRI MAN 2 Kota Payakumbuh. Evaluasi menggunakan metrik seperti *confusion matrix* untuk menganalisis distribusi prediksi benar (*true positive*, *true negative*) dan kesalahan (*false positive*, *false negative*), serta menghitung metrik akurasi, *precision*, *recall*, dan *F1-score* guna mendapatkan gambaran performa model secara menyeluruh.

Binary Cross Entropy (BCE) juga digunakan sebagai fungsi kerugian untuk mengukur seberapa besar selisih antara probabilitas prediksi model dan nilai aktual pada data latih. BCE memberikan penalti yang tinggi ketika model melakukan prediksi yang sangat meleset dari nilai sebenarnya, dan menjadi indikator penting dalam proses pelatihan untuk memastikan model mempelajari hubungan antarvariabel secara optimal. Dengan

kombinasi metrik evaluasi ini, performa model dalam memprediksi kelayakan peminjaman dapat dinilai secara menyeluruh dan akurat.

Selain itu, ROC (*Receiver Operating Characteristic*) Curve digunakan untuk mengevaluasi kemampuan model dalam membedakan antara dua kelas. *Area Under Curve* (AUC) dari ROC Curve menjadi indikator utama, di mana nilai AUC yang lebih besar menunjukkan kemampuan model yang lebih baik dalam melakukan klasifikasi. Hasil evaluasi melalui *confusion matrix* dan ROC Curve dibandingkan untuk mengukur keakuratan, sensitivitas, dan spesifisitas model dalam mengklasifikasikan data kelayakan peminjaman. Evaluasi ini memberikan pemahaman yang komprehensif mengenai performa model regresi logistik yang digunakan dalam penelitian ini.

6. Penyebaran Hasil

Dalam tahap ini, model regresi logistic diintegrasikan ke dalam aplikasi berbasis *website*, yang nantinya dapat digunakan oleh pengelola koperasi dalam proses pengambilan keputusan peminjaman. Model yang telah dievaluasi dan menunjukkan performa yang baik akan diterapkan dalam sistem berbasis data yang memungkinkan koperasi melakukan analisis peminjaman secara lebih cepat dan objektif. Selain itu, sistem ini dapat menampilkan hasil klasifikasi peminjam secara *real-time*, memberikan rekomendasi berdasarkan pola data historis.

Untuk mendukung kemudahan penggunaan, sistem ini divisualisasikan melalui google collab dengan bahasa pemrograman *Python*, lalu menggunakan streamlit untuk membangun aplikasi web yang *responsive* agar dapat diakses oleh pengguna tanpa latar belakang teknis. Dengan adanya *deployment* ini, koperasi dapat meningkatkan efisiensi operasional, meminimalkan risiko gagal bayar, dan membuat keputusan berbasis data yang lebih akurat dan objektif.

Analisis Kelayakan Peminjaman Anggota Koperasi

Status Anggota

Status Pinjaman

Jumlah Simpanan (Juta)

Jumlah Pinjam (Juta)

Jangka Waktu (Bulan)

Gambar 3. 2 Tampilan Awal Aplikasi Web

Wireframe aplikasi prediksi kelayakan peminjaman ini dirancang untuk memberikan antarmuka yang sederhana, interaktif, dan mudah dipahami oleh pengguna, khususnya pengelola koperasi. Tampilan awal aplikasi terdiri dari halaman input data, di mana pengguna dapat mengisi variabel dan menekan tombol “Cek Kelayakan”, seperti ditunjukkan pada Gambar 3.2.

Analisis Kelayakan Peminjaman Anggota Koperasi

Status Anggota

Status Pinjaman

Jumlah Simpanan (Juta)

Jumlah Pinjam (Juta)

Jangka Waktu (Bulan)

Hasil Kelayakan:

Probabilitas:

Gambar 3. 3 Tampilan Hasil Aplikasi Web

Setelah data diproses, sistem akan menampilkan hasil prediksi kelayakan, nilai probabilitas, serta penjelasan interpretatif dalam bentuk teks, sebagaimana terlihat pada Gambar 3.3. Desain antarmuka ini dibuat untuk memudahkan pihak koperasi dalam melakukan analisis secara mandiri, tanpa memerlukan keahlian teknis di bidang data science.

Tampilan wireframe ini telah ditunjukkan dan mendapatkan persetujuan dari pihak KPRI MAN 2 Kota Payakumbuh, dengan masukan untuk menjaga kesederhanaan layout dan memprioritaskan kejelasan hasil prediksi.

C. *Setting Penelitian*

1. Lokasi Penelitian

Penelitian dilakukan di KPRI MAN 2 Kota Payakumbuh, yang merupakan koperasi aktif dengan layanan peminjaman dana. Data yang digunakan berasal dari catatan historis aktivitas keuangan anggota Januari 2023 – Mei 2025. Penelitian dimulai dari pengamatan, observasi, dan wawancara langsung pada lokasi penelitian.

2. Waktu Penelitian

Penelitian ini dijadwalkan selama 5 bulan dimulai dari bulan desember 2024.

Tabel 3. 1 Waktu Penelitian

Kegiatan/Waktu	Bulan 1				Bulan 2				Bulan 3				Bulan 4				Bulan 4			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pemahaman Bisnis																				
Pemahaman Data																				
Persiapan Data																				
Pemodelan																				
Evaluasi																				
Penyebaran Hasil																				
Dokumentasi																				

D. Bahan Penelitian

Dataset yang digunakan dalam penelitian ini adalah data historis peminjaman anggota KPRI MAN 2 Kota Payakumbuh dari Januari 2023 hingga Mei 2025. Dataset ini terdiri atas 337 record dengan lima atribut utama:

1. Status Anggota
2. Jumlah Simpanan
3. Status Pinjaman
4. Jumlah Pinjaman
5. Jangka Waktu

Untuk proses pengolahan data, dataset dibagi menjadi dua bagian, yaitu data training dan data testing. Data training terdiri atas 273 record yang mencakup data historis peminjaman selama tahun 2023 dan tahun 2024. Sementara itu, data testing mencakup data historis peminjaman pada periode Januari hingga Mei 2025 dengan jumlah 64 record. Pembagian ini bertujuan untuk memastikan bahwa model yang dibangun dapat diuji menggunakan data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi lebih representatif.

E. Alat Penelitian

Alat dan instrumen yang digunakan untuk menyelesaikan penelitian ini terdiri dari perangkat lunak dan perangkat keras, yaitu:

1. Perangkat Lunak

Perangkat Lunak atau *Software* yang digunakan pada penelitian ini:

- a. Sistem Operasi Windows 11
- b. Bahasa pemrograman *python*
- c. *Browser*

2. Perangkat Keras

Perangkat Lunak atau *Software* yang digunakan pada penelitian ini:

- a. Processor Intel Pentium
- b. RAM (Random Access Memory) 4 GB
- c. *System type* 64-bit