

BAB II

KERANGKA TEORITIS

A. Landasan Teori

1. Data Mining dan Clustering

Menurut Prastyadi dkk (2024,p.2) data mining adalah suatu proses ekstraksi pengetahuan atau informasi yang berharga dari suatu set data yang besar dan kompleks. Tujuan utama dari data mining adalah mengidentifikasi pola, hubungan, atau informasi yang mungkin tidak terlihat secara langsung dalam data, sehingga memberikan wawasan yang lebih dan bernilai.

Proses data mining melibatkan penggunaan berbagai teknik statistik, matematis, dan kecerdasan buatan untuk menganalisis data dengan cara yang sistematis dan otomatis. Hasil dari data mining dapat digunakan untuk mendukung pengambilan keputusan, mengidentifikasi tren pasar, meningkatkan efisiensi operasional, atau merumuskan strategi bisnis. (Prastyadi dkk, 2024, p.2)

Data mining menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisis terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut Menurut Daryl Pregibon disebutkan bahwa Data Mining adalah perpaduan dari Statistik, Artificial Intelligent dan Database (Gorunescu, 2011). Data Mining kemudian dikenal dengan nama Knowledge-Discovery in Databases (KDD) (Rahmadya dan Herlawati,2020, pp.1-2)

Tujuan utama dari data mining adalah mengidentifikasi pola, hubungan, atau pengetahuan yang berharga dan tersembunyi dalam suatu set data besar atau kompleks. Proses data mining bertujuan untuk menggali wawasan yang tidak dapat ditemukan secara langsung melalui pengamatan sederhana terhadap data. Beberapa tujuan khusus dari data mining meliputi:

a. Pencarian Pola atau Hubungan

Mengidentifikasi pola atau hubungan yang mungkin tidak terlihat secara langsung dalam data.

b. Prediksi dan Klasifikasi

Membangun model yang dapat memprediksi nilai yang belum diketahui atau mengklasifikasi data ke dalam kategori tertentu.

c. Segmentasi Pelanggan

Membagi pelanggan atau data ke dalam segmen atau kelompok berdasarkan karakteristik tertentu, seperti perilaku pembelian atau preferensi.

- d. Analisis Asosiasi
Mengidentifikasi hubungan atau asosiasi antara variable-variable dalam data, seperti hubungan antara produk yang sering dibeli bersama-sama.
- e. Deteksi Anomali
Mendeteksi pola yang tidak biasa atau anomaly dalam data, yang dapat mengindikasikan situasi atau kejadian yang tidak normal atau patut dicurigai.
- f. Optimasi Proses Bisnis
Meningkatkan efisiensi dan produktifitas dengan mengidentifikasi area-area dimana proses bisnis dapat dioptimalkan.
- g. Penentuan Profil Konsumen
Membangun profil konsumen berdasarkan data pembelian, preferensi, dan perilaku konsumen untuk mendukung strategi pemasaran yang lebih efektif
- h. Penentuan Tren Pasar
Mengidentifikasi tren dan pola perilaku pasar yang dapat digunakan untuk merumuskan strategi bisnis dan pemasaran.
- i. Analisis Risiko
Menganalisis risiko dalam keputusan bisnis, seperti risiko kredit dalam industri keuangan.
- j. Penemuan Pengetahuan Baru
Menemukan informasi atau pengetahuan baru yang dapat memberikan wawasan atau keuntungan kompetitif.
Data mining memiliki banyak manfaat dan dapat memberikan nilai tambah yang signifikan untuk berbagai industri dan organisasi. Berikut adalah beberapa manfaat utama dari data mining:
 - a. Pengambilan Keputusan yang Lebih Baik
Data mining membantu organisasi membuat keputusan yang lebih informasional dan terarah. Dengan menganalisis data historis, model prediktif, dan pola tersembunyi, pengambilan keputusan yang dapat didukung oleh wawasan yang lebih mendalam.
 - b. Penemuan Pola dan Tren
Data mining memungkinkan identifikasi pola dan tren yang tidak terlihat secara langsung oleh manusia. Ini dapat membantu organisasi dalam merespon perubahan pasar, mengidentifikasi peluang baru, atau menanggapi tantangan bisnis.

c. Prediksi Masa Depan

Dengan menggunakan teknik prediktif, data mining dapat membantu organisasi memprediksi kejadian masa depan berdasarkan data historis. Contohnya termasuk prediksi penjualan, perilaku pelanggan, atau peristiwa risiko.

d. Segmentasi Pelanggan

Melalui pengelompokan atau segmentasi pelanggan, data mining memungkinkan organisasi untuk memahami preferensi dan perilaku pelanggan secara lebih baik. Ini dapat digunakan untuk pengembangan produk, penargetan pemasaran, dan personalisasi layanan.

e. Efisiensi Operasional

Data mining membantu mengidentifikasi area dimana efisiensi operasional dapat ditingkatkan. Analisis ini dapat membantu organisasi mengoptimalkan proses bisnis, mengurangi biaya, dan meningkatkan produktivitas.

f. Pengelolaan Risiko

Dengan mendeteksi pola anomali atau melalui model risiko, data mining dapat membantu organisasi dalam pengelolaan risiko. Ini berlaku untuk institusi keuangan (deteksi penipuan), kesehatan (prediksi penyakit), dan lainnya.

g. Pemasaran yang Lebih Efektif

Melalui analisis data pelanggan, preferensi, dan perilaku pembelian, data mining dapat membantu perusahaan menyusun strategi pemasaran yang lebih efektif dan berfokus.

h. Peningkatan Layanan Pelanggan

Dengan memahami lebih baik kebutuhan dan preferensi pelanggan, organisasi dapat meningkatkan layanan pelanggan mereka, memberikan pengalaman yang lebih baik, dan meningkatkan retensi pelanggan.

Data mining memiliki berbagai penerapan di berbagai industri dan bidang.

Berikut adalah contoh penerapan data mining:

a. Pemasaran dan Analisis Pelanggan

Analisis data pelanggan dapat membantu perusahaan mengidentifikasi pola pembelian, preferensi pelanggan, dan perilaku konsumen. Ini memungkinkan perusahaan untuk merancang kampanye pemasaran yang lebih efektif dan menyediakan layanan yang lebih personal.

b. Keuangan dan Perbankan

Data mining digunakan untuk mendeteksi penipuan kredit, analisis risiko kredit, prediksi perilaku pasar keuangan, dan identifikasi tren dalam transaksi keuangan. Ini membantu Lembaga Keuangan dalam pengambilan keputusan yang lebih baik dan pengelolaan risiko.

c. Kesehatan dan Perawatan Medis

Data mining dapat membantu dalam analisis data kesehatan untuk prediksi penyakit, identifikasi pola penyebaran penyakit, dan pengembangan model diagnostik. Ini juga dapat digunakan untuk personalisasi perawatan pasien dan manajemen data klinis.

d. Manufaktur dan Produksi

Dalam industri manufaktur, data mining dapat digunakan untuk meningkatkan efisiensi operasional, merencanakan rantai pasokan dan memperbaiki kualitas produk. Analisis data dapat membantu mengoptimalkan proses produksi dan mengurangi biaya.

e. Sumber Daya Manusia

Data mining dapat digunakan dalam manajemen sumber daya manusia untuk rekrutmen yang lebih efektif, retensi karyawan, dan analisis kinerja karyawan. Ini membantu organisasi dalam pengelolaan sumber daya manusia secara lebih efisien.

f. Telekomunikasi

Dalam industri telekomunikasi, data mining digunakan untuk analisis pelanggan, manajemen kapasitas jaringan, dan deteksi kecurangan. Ini membantu penyedia layanan telekomunikasi dalam meningkatkan kualitas layanan dan kepuasan pelanggan.

g. Pendidikan

Data mining dapat membantu institusi pendidikan dalam meningkatkan efektivitas pengajaran dan membantu pengambilan keputusan berbasis data. Ini dapat digunakan untuk memahami pola belajar siswa, memprediksi keberhasilan akademis, dan mengidentifikasi area yang memerlukan perhatian khusus.

h. Riset Ilmiah

Dalam penelitian ilmiah, data mining digunakan untuk mengekstrak pola-pola dari data eksperimental dan analisis besar-besaran. Ini membantu peneliti dalam membuat hipotesis baru atau mengidentifikasi tren yang tidak terlihat secara langsung.

i. Pemerintahan dan Keamanan

Pemerintah dapat membangun data mining untuk analisis keamanan publik, deteksi aktivitas kriminal, dan prediksi tren keamanan. Ini juga dapat membantu dalam manajemen data pemerintah untuk pengambilan keputusan yang lebih baik.

j. E-commerce

Perusahaan e-commerce menggunakan data mining untuk rekomendasi produk yang lebih akurat kepada pelanggan, analisis pola pembelian, dan manajemen rantai pasokan. Ini dapat meningkatkan pengalaman belanja online dan kepuasan pelanggan.

Ada beberapa Teknik data mining yang digunakan untuk mengekstrak pola dan pengetahuan dari data. Berikut adalah beberapa Teknik utama:

a. Klasifikasi (Classification)

Teknik ini digunakan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik atau atribut tertentu. Contohnya termasuk pohon keputusan, jaringan saraf tiruan, dan algoritma klasifikasi lainnya.

b. Regresi (Regression)

Digunakan untuk memodelkan hubungan antara variable dependen dan independent. Regresi digunakan untuk memprediksi nilai kontinu berdasarkan hubungan linier atau non-linier antara variable-variable tersebut.

c. Klustering (Clustering)

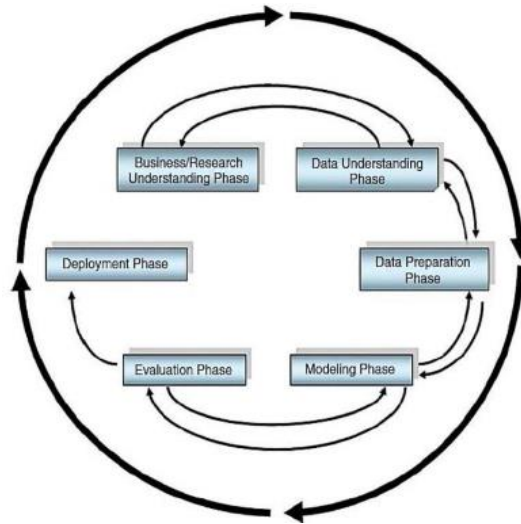
Klustering digunakan untuk mengelompokkan data kedalam kelompok atau klaster berdasarkan kesamaan fitur atau karakteristik tertentu. Algoritma K-Means dan hierarchical clustering adalah contoh teknik klustering.

d. Asosiasi (Association)

Teknik ini digunakan untuk menemukan hubungan atau asosiasi antara item atau variable dalam data. Algoritma terkenal seperti Apriori dan Eclat digunakan untuk menemukan aturan asosiasi dalam data transaksional. (Rahmadya dkk, 2020, pp.3-8)

2. CRISP-DM

Metodologi datamining CRISP-DM menjelaskan proses model proses secara hirarki, merupakan sekumpulan tugas yang dinyatakan dengan empat level abstrak (dari umum ke spesifik), tahap, tugas umum, tugas khusus dan contoh proses (proses instance). Pada level yang paling atas, proses datamining dikenal dengan empat tahap, masing-masing tahap terdiri dari beberapa tugas umum. Tugas umum digunakan sebagai pelengkap dan setabil mungkin. Lengkap artinya mencakup baik proses data mining secara umum dan semua kemungkinan aplikasi data mining. Stabil artinya model harus valid sebelum meramalkan pengembangan secara Teknik pemodelan yang baru (Mulaab, 2021, p.8).



Gambar 2. 1 Ilustrasi CRISP-DM

(Sumber: Mulaab, 2021,p.8)

Berikut penjelasan dari setiap tahapan diatas:

1. Business/Research Understanding

Pada tahapan ini kita mendefinisikan tujuan permasalahan yang dihadapi serta apa saja yang diperlukan untuk sebuah unit bisnis/riset secara keseluruhan. Kemudian kita juga menyiapkan strategi untuk mencapai tujuan tersebut.

2. Data Understanding

Selain mengumpulkan data, di tahap ini kita juga memahami data dengan melihat tipe data apa saja yang didapatkan. Di tahapan ini kita juga bisa membuat beberapa visualisasi untuk melihat hubungan atau pola antara beberapa variable dari data yang dimiliki.

3. Data Preparation

Pada tahapan ini kita membersihkan data, melakukan transformasi data, memilih fitur (*feature selection*) yang ingin dianalisis lebih lanjut.

4. Modeling Phase

Pada tahapan ini kita membuat model *machine learning* maupun model optimasi untuk menyelesaikan permasalahan yang ada. Biasanya semua model yang dibuat ditujukan untuk memprediksi data di masa depan.

5. Evaluation Phase

Merupakan tahapan di mana kita mengevaluasi performa model yang dihasilkan dengan membandingkan data hasil prediksi dengan data asli di lapangan. Jika performa yang dihasilkan kurang baik, maka kita bisa kembali ke tahapan-tahapan sebelumnya untuk menyempurnakan model.

6. Deployment Phase

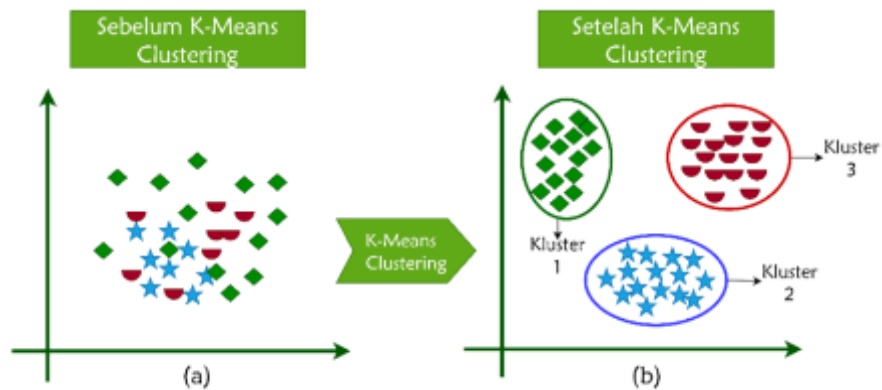
Tahapan dimana kita menjalankan model secara *live*. Pada tahapan ini *users* (pengguna) bisa memanfaatkan model yang dihasilkan. Contoh sederhananya adalah aplikasi yang ada di smartphones kita yang bis akita gunakan, maka aplikasi tersebut sudah *terdeploy* secara *live*.

Setelah semua proses dilewati, missal kita sudah berada di tahapan *deployment*, maka proses tidak berhenti dan selesai. Masih ada proses terus berlanjut yang disebut dengan *continuous monitoring* dan *improvement*. Dengan demikian, setiap implementasi model yang dihasilkan nanti terus dimonitor dan dipantau performanya dari waktu ke waktu. Jika performanya turun karena faktor eksternal yang berubah misalnya, maka perlu dilakukan proses *training* (pembelajaran) ulang menggunakan data-data terbaru atau bahkan merombak model dan memulai lagi dari awal (Hapsari dkk, 2021, pp.137-139).

3. Algoritma K-Means

Menurut Ibnu Daqiqil (2021:217) K-Means adalah salah satu "*unsupervised machine learning algorithms*" yang paling sederhana dan populer. Tujuan dari algoritma ini adalah untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variable K. Variabel K sendiri adalah jumlah kluster yang kita inginkan. Metode K-Means Clustering berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Karakteristik yang sama itu ditandai dengan jarak atau *distance* yang lebih dekat, mirip seperti KNN. Dengan kata lain, metode K-Means Clustering bertujuan untuk meminimalisir *objective function* yang diset dalam proses *clustering* dengan cara meminimalkan variasi antar data didalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya.

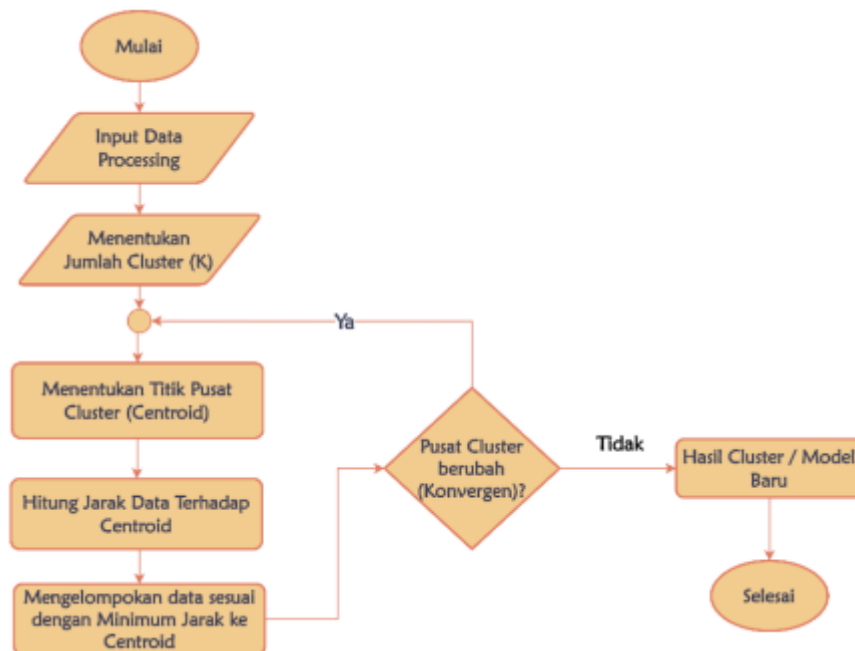
Putra dkk, (2023,p.85) menjelaskan Algoritma pengelompokan K-Means dalam *machine learning* adalah salah satu algoritma *unsupervised learning* (pembelajaran tanpa pengawasan) yang paling sederhana, namun memiliki hasil yang mendekati benar. Algoritma ini memiliki beberapa penerapan aplikasi, termasuk klasifikasi dokumen, segmmentasi gambar, sistem rekomendari, dll.



Gambar 2. 2 Ilustrasi Data Cluster

(Sumber: Putra dkk, 2023)

Algoritma K-Means merupakan algoritma non hirarki (*non-hierarchical*) yang berasal dari metode data *clustering*. Algoritma K-Means dapat melakukan partisi data ke dalam kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan kedalam set yang sama. Data yang memiliki karakteristik yang beda selanjutnya akan dikelompokkan kedalam set yang sama. Data yang memiliki karakteristik yang berbeda, selanjutnya akan dikelompokkan kedalam set berbeda. (Putra dkk, 2023, p.85)



Gambar 2. 3. Alur kerja Algoritma K-Means

(Sumber: Putra dkk, 2023)

Prinsip utama dari algoritma ini adalah Menyusun K buah partisi dari sekumpulan data. Tujuan K-Means untuk meminimalkan fungsi objektif pada set kelompok dalam proses pengelompokan. Pada umumnya Teknik ini berusaha meminimalkan variasi dalam suatu kelompok dan memaksimalkan variasi antar kelompok lainnya. Berikut ini merupakan penjelasan lebih lanjut terkait proses K-Means dari Gambar 2.2:

- Langkah 1 : Pilih nilai K untuk menentukan jumlah *cluster* yang akan dibentuk
- Langkah 2 : Menetapkan titik K secara acak yang akan bertindak sebagai pusat *cluster (centroid)*. Menentukan nilai *centroid* dengan cara mengambil dari nilai rata-rata (mean) semua nilai data pada setiap fiturnya. Jika M menyatakan jumlah data pada suatu kelompok, maka i menyatakan fitur ke- i dalam sebuah kelompok.

$$C_i = \frac{1}{M} \sum_{j=i}^M x_j$$

- Langkah 3 : Perhitungan jarak dari *centroid cluster*. Untuk mengukur jarak antar data dengan *centroid* menggunakan formula Euclidean distance.

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

Dimana:

$D(x_2, x_1)$ = Jarak antar dua titik (Euclidian Distance)

p = jumlah data

x_{2j} = titik data ke- p

x_{1j} = titik *centroid*

- Langkah 4 : Tetapkan setiap titik data, berdasarkan jaraknya dari titik yang dipilih secara acak (*centroid*), ke *centroid* terdekat, yang akan membentuk cluster yang telah ditentukan sebelumnya.

$$a_{ji} = \begin{cases} 1, & d = \min(x_j, C_i) \\ 0, & \text{lainnya} \end{cases}$$

- Langkah 5 : Tempatkan *centroid* baru dari setiap *cluster*.
- Langkah 6 : Ulangi Langkah 3 sampai Langkah 5 hingga nilai titik *centroid* tidak berubah (konvergen)
- Langkah 7 : Jika telah konvergen, maka model baru siap digunakan

Algoritma K-Means menjalankan iterasi awal dimana titik data ditempatkan secara acak ke dalam sebuah grup, yang titik pusatnya dikenal sebagai centroid. Metode Euclidian Distance dari setiap titik data ke centroid kemudian dihitung dan jika jarak suatu titik lebih tinggi dari centroid lain, maka titik tersebut dipindahkan ke centroid "lainnya". Algoritma K-Means akan menjalankan iterasi yang lain dan proses akan berlanjut hingga semua pengelompokan memiliki minimum dalam varian grup (Putra dkk, 2023, pp.85-87)

Contoh penerapan Algoritma K-Means Clustering (Azizah,2022,p.50-75) dengan tujuan untuk pemetaan kebutuhan obat sesuai prioritas dalam pengadaan berdasarkan total pemakaian dan jumlah transaksi, dengan tahapan sebagai berikut:

- (1) Langkah pertama menyiapkan dataset yang digunakan pada perhitungan dengan menetapkan variable data total pemakaian dan jumlah transaksi dari data pemakaian obat selama bulan Januari s.d Juni 2022 yang terdiri dari 213 data.
- (2) Langkah kedua yaitu menentukan jumlah cluster sebanyak 3 (tiga) kelompok yaitu, prioritas tinggi, prioritas sedang, dan prioritas rendah.
- (3) Langkah ketiga yaitu menentukan centroid awal secara acak. Data yang dipilih menjadi centroid awal yaitu data ke-7 data ke-10 dan data ke-27, dapat dilihat pada tabel berikut:

Tabel 2. 1 Penentuan Centroid Awal

No	Nama Obat	Total Pemakaian	Jumlah Transaksi
1	Acarbose 100 mg	6	2
2	Acetylcystein 200 mg	110	9
3	Acyclovir 5% krim	2	1
4	Acyclovir salep	5	2
5	Acyclovir tablet	9	3
6	Adalat Oros	60	3
7	Alloclair Gel	2	2
8	Allofar 100 mg	1	1
9	Allofar 300 mg	1	1
10	Allopurinol 100 mg	410	23
11	Allopurinol 300 mg	10	1
12	Ambroxol syrup	20	16
13	Amlodipine 10 mg	450	25
...
26	Batugin Elixir	1	1
27	Becom C	2500	187
...
213	Zinc 20 mg	190	12

Tabel 2. 2 Titik Awal Centroid:

Centroid 1	2	2
Centroid 2	410	23
Centroid 3	2500	187

- (4) Kemudian menghitung jarak setiap data input terhadap masing-masing pusat cluster menggunakan rumus jarak euclidean distance dengan rumus persamaan sebagai berikut:

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

dimana: $x = x_1, x_2, \dots, x_n$ dan y_1, y_2, \dots, y_n

yang merupakan banyaknya n atribut (kolom) antara 2 record.

Iterasi 1

Data ke-1 pada iterasi 1

Menghitung jarak data ke-1 dengan centroid 1

$$\begin{aligned} c_1 &= \sqrt{(6 - 2)^2 + (2 - 2)^2} \\ &= \sqrt{16 + 0} = 4 \end{aligned}$$

Menghitung jarak data ke-1 dengan centroid 2

$$\begin{aligned} c_1 &= \sqrt{(6 - 410)^2 + (2 - 23)^2} \\ &= \sqrt{163216 + 441} = 404,55 \end{aligned}$$

Menghitung jarak data ke-1 dengan centroid 3

$$\begin{aligned} c_1 &= \sqrt{(6 - 2500)^2 + (2 - 187)^2} \\ &= \sqrt{6220036 + 34225} = 2500,85 \end{aligned}$$

Data ke-2 pada iterasi 1

Menghitung jarak data ke-2 dengan centroid 1

$$\begin{aligned} c_2 &= \sqrt{(110 - 2)^2 + (9 - 2)^2} \\ &= \sqrt{11664 + 49} = 108,23 \end{aligned}$$

Menghitung jarak data ke-2 dengan centroid 2

$$\begin{aligned} c_2 &= \sqrt{(110 - 410)^2 + (9 - 23)^2} \\ &= \sqrt{90.000 + 196} = 300,33 \end{aligned}$$

Menghitung jarak data ke-2 dengan centroid 3

$$\begin{aligned} c_2 &= \sqrt{(110 - 2500)^2 + (9 - 187)^2} \\ &= \sqrt{5712100 + 31684} = 2396,62 \end{aligned}$$

- (5) Langkah selanjutnya adalah mengelompokkan data berdasarkan jarak minimum ke centroid

Tabel 2.3 Hasil Pengelompokan pada iterasi-1

No	Nama Obat	C1	C2	C3	Jarak Terdekat	Hasil
1	Acarbose 100 mg	4.00	404.55	2500.85	4.00	C1
2	Acetylcystein 200 mg	108.23	300.033	2396,62	108,23	C1
3	Acyclovir 5% krim	1.00	408.59	2504,92	1,00	C1
4	Acyclovir salep	3.00	405.54	2501,85	3,00	C1
5	Acyclovir tablet	7.07	350,57	2497,79	7,07	C1
6	Adalat Oros	58.01	408,54	2446,93	58,01	C1
7	Alloclair Gel	0.00	409,54	2504,84	0,00	C1
8	Allofal 100 mg	1.41	409,59	2505,91	1,41	C1
9	Allofar 300 mg	1.41	409,59	2505,91	1,41	C1
10	Allopurinol 100 mg	408.54	0,00	2096,42	0,00	C2
...
213	Zinc 20 mg	188.27	220,27	2316,62	188,27	C1

- (6) Kemudian melakukan penentuan centroid baru. Penentuan nilai centroid baru diperoleh dengan menghitung rata-rata (mean) dari semua data pada setiap cluster.
- (7) Ulangi langkah 3 s.d 5 sampai nilai centroid tidak berubah sehingga sudah tidak ada lagi data yang berpindah ke cluster lain sehingga diperoleh hasil cluster 1 sebanyak 197 data, cluster 2 sebanyak 13 data, dan cluster 3 sebanyak 3 data, berikut adalah hasil akhir perhitungan dan pengelompokan clusternya:

Tabel 2. 4 Hasil Akhir Cluster

No	Nama Obat	C1	C2	C3	Jarak Terdekat	Hasil
1	Acarbose 100 mg	4.00	404.55	2500.85	4.00	C1
2	Acetylcystein 200 mg	108.23	300.033	2396,62	108,23	C1
3	Acyclovir 5% krim	1.00	408.59	2504,92	1,00	C1
4	Acyclovir salep	3.00	405.54	2501,85	3,00	C1
5	Acyclovir tablet	7.07	350,57	2497,79	7,07	C1
6	Adalat Oros	58.01	408,54	2446,93	58,01	C1
7	Alloclair Gel	0.00	409,54	2504,84	0,00	C1
8	Allofal 100 mg	1.41	409,59	2505,91	1,41	C1
9	Allofar 300 mg	1.41	409,59	2505,91	1,41	C1
10	Allopurinol 100 mg	408.54	0,00	2096,42	0,00	C2
...
213	Zinc 20 mg	188.27	220,27	2316,62	188,27	C1

4. Unified Modeling Language (UML)


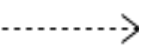

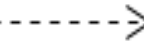

UML adalah sebuah bahasa yang berdasar pada grafik/gambar untuk emmvisualisasi, menspesifikasikan, membangun, dan pendokumentasian dari sebuah sistem pengembangan software berbasis OO (Object-Oriented). Salah satu dari tahapan SDLC adalah desain. Desain bertujuan agar software yang akan dibuat dapat memenuhi kebutuhan user dan handal. Oleh karena itu, desain menjadi tahaoan penting dalam proses pembuatan software. Dalam mendesain software, kita perlu mentransformasikan kebutuhan user, baik secara fungsional maupun non fungsional ke dalam model. Untuk memodelkan sesua, diperlukan bahasa pemodelan. Bahasa pemodelan dapat berupa pseudo-code, code, gambar, diagram atau deskripsi yang menggambarkan sebuah sistem. (Sumirat dkk,2023,pp.73-74)






Diagram-diagram yang didefinisikan oleh UML antara lain:

a. Use Case Diagram

Use case adalah komponen yang menggambarkan fungsional dalam sebuah sistem. Sehingga konsumen maupun pembuat saling mengenal dan mengerti alur sistem yang akan dibuat. (Sumirat dkk,2023,p.82)

Tabel 2. 5.Simbol Use Case Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actor</i>	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan <i>use case</i> .
2		<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (<i>independent</i>) akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri (<i>independent</i>).
3		<i>Generalization</i>	Hubungan dimana objek anak (<i>descendent</i>) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (<i>ancestor</i>).
4		<i>Include</i>	Menspesifikasikan bahwa <i>use case</i> sumber secara <i>eksplisit</i> .
5		<i>Extend</i>	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan.


NO	GAMBAR	NAMA	KETERANGAN
6		<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya.
7		<i>System</i>	Menspesifikasikan paket yang menampilkan sistem secara terbatas.
8		<i>Use Case</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor
9		<i>Collaboration</i>	Interaksi aturan-aturan dan elemen lain yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah dan elemen-elemennya (sinergi).
10		<i>Note</i>	Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi





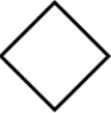
(Sumber: Sumirat dkk,2023,p.82)

b. Activity Diagram

Activity diagram menggambarkan berbagai alir aktivitas dalam sistem yang sedang dirancang, bagaimana masing-masing alir berawal, keputusan yang mungkin terjadi, dan bagaimana mereka berakhir. *Activity diagram* juga dapat menggambarkan proses parallel yang mungkin terjadi pada beberapa eksekusi. *Activity diagram* digunakan digunakan untuk menggambarkan langkah-langkah atau aktivitas pada suatu sistem. *Activity diagram* menggambarkan berbagai alir aktivitas dalam sistem yang sedang dirancang, mulai dari mana berasal, kemungkinan yang bisa terjadi dan bagaimana berakhirnya serta menggambarkan proses yang mungkin terjadi pada beberapa eksekusi. (Hasanah dkk,2020,pp.79-80)

Tabel 2. 6.Simbol Activity Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actifity</i>	Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain


2		<i>Action</i>	State dari sistem yang mencerminkan eksekusi dari suatu aksi
3		<i>Initial Node</i>	Bagaimana objek dibentuk atau diawali.
4		<i>Activity Final Node</i>	Bagaimana objek dibentuk dan dihancurkan
5		<i>Fork Node</i>	Satu aliran yang pada tahap tertentu berubah menjadi beberapa aliran
6		<i>Decision Node</i>	Mewakili titik percabangan bersyarat dengan satu input dan beberapa output, fungsinya mengarahkan alur berdasarkan kondisi tertentu.





(Sumber: Hasanah dkk,2020,p.82)

c. Sequence Diagram

Sequence Diagram atau diagram urutan adalah sebuah diagram yang digunakan untuk menjelaskan dan menampilkan interaksi antar objek-objek dalam sebuah sistem secara terperinci. Selain itu *sequence diagram* juga akan menampilkan pesan atau perintah yang dikirim, beserta waktu pelaksanaannya. Objek-objek yang berhubungan dengan berjalannya proses operasi biasanya diurutkan dari kiri ke kanan. Tujuan utama pembuatan diagram *sequence* adalah untuk mengetahui urutan kejadian yang dapat menghasilkan output yang diinginkan. Selain itu, tujuan dari diagram *sequence* ini mirip dengan *activity diagram*, seperti menggambarkan alur kerja dari sebuah aktivitas, serta dapat menggambarkan aliran data dengan lebih detail, termasuk data atau perilaku yang diterima atau dikirimkan. Berikut simbol *sequence diagram* yang dapat dilihat pada tabel 2.3 (Sumirat dkk,2023,pp.86-87)

Tabel 2. 7.Sequence Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actor</i>	Komponen ini menggambarkan seorang pengguna (user0 yang berada di luar sistem dan sedang berinteraksi dengan sistem. Dalam <i>sequence diagram</i> , aktor biasanya

NO	GAMBAR	NAMA	KETERANGAN
			digambarkan dengan simbol <i>stick figure</i> .
2		<i>Activation</i>	Komponen activation box ini merepresentasikan waktu yang dibutuhkan suatu objek untuk menyelesaikan tugasnya. Semakin lama waktu yang diperlukan, maka secara otomatis activation box-nya juga akan menjadi lebih panjang.
3		<i>LifeLine</i>	Komponen ini digambarkan dengan bentuk garis putus-putus yang terhubung dengan objek, sepanjang <i>lifeline</i> terdapat <i>activation</i> .
4		<i>Boundary</i>	Komponen ini berisi kumpulan kelas yang menjadi interface atau interaksi antara satu atau lebih aktor dengan sistem, seperti <i>form entry</i> dan <i>form cetak</i>
5		<i>Control</i>	Komponen ini merupakan suatu objek berisi logika aplikasi yang tidak memiliki tanggung jawab kepada entitas, contohnya adalah kalkulasi dan aturan bisnis yang melibatkan berbagai objek.

NO	GAMBAR	NAMA	KETERANGAN
6		<i>Entity</i>	Komponen ini merupakan bagian dari sistem yang berisi kumpulan kelas berupa entitas-entitas yang membentuk gambaran awal sistem dan menjadi landasan untuk menyusun basis data
7		<i>Recursive</i>	Komponen ini menggambarkan pengiriman pesan yang dikirim untuk dirinya sendiri.
8		<i>Message</i>	Komponen ini untuk menggambarkan komunikasi antar objek. <i>Message</i> biasanya muncul secara berurutan pada <i>lifeline</i> . Komponenn <i>message</i> ini direpresentasikan dengan anak panah. Inti dari sebuah diagram urutan terdapat pada komponen <i>lifeline</i> dan <i>message</i> ini.

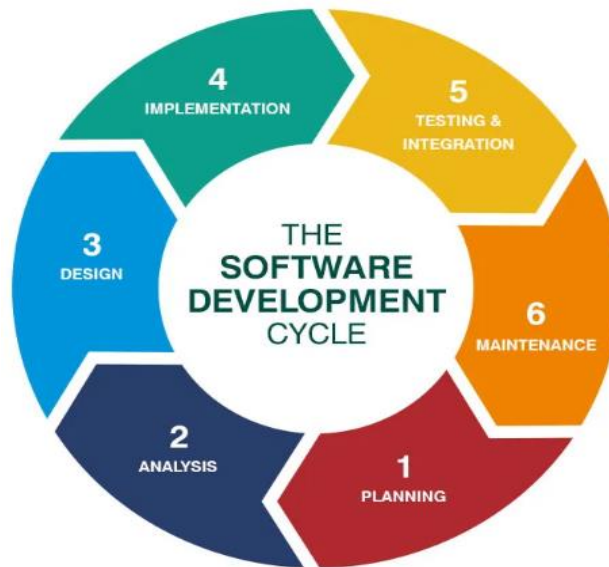
(Sumber: Sumirat dkk,2023,pp.87-88)

5. System Development Life Cycle (SDLC)

Menurut Wijayanto (2024,pp.35-37) SDLC adalah model konseptual dalam proyek manajemen yang menjelaskan tahapan kerja dalam pengembangan sistem informasi. Model ini mencakup studi kelayakan dari tahap awal hingga pemeliharaan aplikasi. Tujuan SDLC adalah menghasilkan sistem berkualitas tinggi yang memenuhi kebutuhan pelanggan dalam hal waktu, biaya, efektivitas, dan efisiensi. Akibatnya, sistem ini menyediakan rencana lengkap untuk mengembangkan, memelihara, dan menggantikan perangkat lunak tertentu.

Metode SDLC akan membantu pengembang merancang dan mengembangkan serta mengidentifikasi masalah yang mungkin muncul selama proses pengembangan dan memantau mereka menyelesaikannya. Dengan kata lain, SDLC adalah tahapan kerja yang bermanfaat untuk menghasilkan suatu sistem yang memenuhi persyaratan klien. Ini terdiri dari kerangka yang berisi Langkah-langkah yang diambil dalam proses pengembangan software. Didalam penerapannya, Model SDLC digunakan untuk merancang sistem yang terstruktur (Perancangan Sistem

Tradisional), kemudian SDLC juga memiliki beberapa tahapan seperti gambar dibawah ini:



Gambar 2. 4.Tahapan SDLC

(Sumber: Wijayanto, 2024)

Berikut ini adalah penjelasan dari ke enam tahapan SDLC:

a. Planning

Dalam fase ini, elemen studi yang berkaitan dengan feasibility study atau kelayakan pengembangan sistem lebih ditekankan. Menentukan prioritas teknologi, menentukan tujuan dan ruang lingkup pengembangan, menciptakan dan mengkonsolidasikan tim pengembang dan menentukan dan menilai strategi.

b. Analysis

Sebuah analisis sistem akan dilakukan mengenai ncara software akan dijalankan, Hasilnya dapat mencakup kelebihan dan kekurangan sistem, fungsinya, dan perbaikan.

c. Design

Saat ini tugasnya adalah membuat prototipe dan output lainnya, termasuk dokumen yang berisi desain, pola, dan bagian yang diperlukan untuk proyek.

d. Implementation

Tahapan implementasi sistem, dimana Anda dapat menerapkan rancangan yang sudah Anda buat pada tahapan sebelumnya.

e. Testing

Tahapan ini adalah tahapan untuk dapat melakukan uji coba untuk memastikan bahwa operasinya berjalan lancar.

f. Maintenance

Meskipun software sudah jadi, tetap harus dijaga dan diawasi agar error dapat diperbaiki secara cepat untuk tujuan perawatan agar data pengguna tetap aman. (Wijayanto, 2024, pp.35-37)

6. Metode Prototype

Metode Prototype atau sering disebut juga dengan prototyping merupakan sebuah metode pengembangan sistem yang didasarkan pada konsep working model. Penelitian lain pun mengatakan prototype didefinisikan sebagai alat yang memberikan ide bagi pembuat maupun pemakai potensial tentang cara sistem berfungsi dalam bentuk lengkapnya dan proses untuk menghasilkan sebuah prototype disebut prototyping. (Yusniar, 2020, pp.14-15)



Gambar 2. 5. Metode Prototype

(Sumber: Yusniar, 2020)

Gambar diatas menjelaskan bagaimana tahapan-tahapan yang dilakukan oleh metode prototype tersebut. Pada tahap pertama dilakukannya proses pengumpulan kebutuhan seperti data-data terkait dalam penelitian dan kebutuhan sistem yang akan dikembangkan. Pada tahap kedua akan dilakukannya proses perancangan dan membuat prototype system. Pada tahap terakhir yaitu dilakukannya pengujian terhadap system yang dibuat dan kemudian dilakukan evaluasi.

Dalam setiap metode tertentu memiliki kelebihan dan kekurangannya termasuk metode prototype tersebut. Adapun kelebihan dan kekurangan yang dimiliki metode prototype adalah sebagai berikut:

Tabel 2. 8.Kelebihan dan Kekurangan Metode Prototype

Kelebihan	Kekurangan
Pengguna (User) berperan aktif dalam pengembangan sistem	Kualitas aplikasi belum teruji dan belum mencantumkan pemeliharaan jangka panjang.
Waktu yang digunakan lebih efisien	Algoritma dan Bahasa yang digunakan sederhana
Adanya komunikasi antara user dengan pengembang	Teknik rancangan tidak baik dilihat dari hubungan pelanggan dengan computer
Pengembang dapat bekerja lebih baik	

(Sumber: Yusniar, 2020)

7. Definisi Kredit dan Kredit Usaha Rakyat

Menurut Agusfianto dkk (2022,p.67) secara sederhana, kredit diartikan sebagai penyaluran dana dari pihak pemilik dana kepada pihak yang memerlukan dana. Penyaluran tersebut didasarkan atas kepercayaan yang diberikan oleh pemilik dana kepada pengguna dana, dimana kredit berasal dari Bahasa latin “*Credere*” yang berarti kepercayaan. Artinya pihak yang memberikan kredit percaya kepada pihak yang menerima kredit bahwa kredit yang diberikan pasti akan dibayar.

Menurut Wico Tarigan kk (2024,p.84) KUR adalah bentuk pinjaman yang diberikan oleh bank kepada usaha kecil dan mikro dengan tingkat suku bunga yang lebih rendah yang bertujuan untuk mendukung pengembangan usaha kecil dan mikro untuk meningkatkan kesejahteraan masyarakat yang umumnya memiliki persyaratan yang lebih mudah diakses oleh pelaku usaha kecil.

B. Tinjauan Pustaka

Penelitian rujukan merupakan acuan yang dibutuhkan seorang peneliti untuk melakukan penelitian. Penelitian rujukan pada penelitian ini diambil berdasarkan kesamaan metode yaitu dengan Algoritma K-Means, banyak penelitian yang menggunakan metode ini dalam berbagai kasus. Antara lain adalah:

(1) Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia (Fadhillah dkk, 2021)

Penerapan metode K-Means mudah untuk diimplementasikan dan dijalankan sehingga dapat mempermudah penulis dalam penyusunan penelitian.

- (2) Penerapan Metode K-means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru** oleh (Nanda Ayu dkk, 2022)

Penelitian bertujuan untuk menentukan strategi promosi Universitas berdasarkan data penerimaan mahasiswa baru dengan menggunakan metode KDD (Knowledge Discovery in Databases) dalam proses analisa data.

- (3) Penerapan Metode K-Means pada Data Penduduk Miskin Per Kecamatan Kabupaten Blitar** oleh (Dede dkk, 2022)

Penelitian bertujuan untuk mengelompokkan data penduduk miskin dari beberapa kecamatan dengan mementingkan aspek: jumlah keluarga miskin, penduduk sedang sekolah, sector pertanian, kepadatan penduduk, pelayanan PLN

- (4) Klasifikasi Data Nasabah Kredit Pinjaman Menggunakan Data Mining Dengan Metode K-Means pada Mega Central Finance** oleh (Rike dkk, 2022)

Penelitian bertujuan untuk menentukan nasabah yang layak menerima pinjaman dengan 3 karakteristik nasabah: aktif, pasif, dan repeat order.

- (5) Komparasi K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman Financial Technology** oleh (Akmal Fikri dkk, 2023)

Penelitian menggunakan 2 cluster dengan atribut: jumlah rekening pemberi pinjaman, jumlah dana yang diberikan, jumlah penerima pinjaman, jumlah rekening penerima pinjaman aktif, outstanding pinjaman.

- (6) Penerapan Metode K-Means untuk Menganalisis Minat Nasabah Asuransi** oleh (Juniar dkk, 2021)

Peneliti membahas permasalahan umum nasabah yang tidak percaya kepada pihak asuransi terhadap proses klaim dan lamanya proses terbitnya polis asuransi. Menggunakan 3 cluster produk: Asuransi Kebakaran, Asuransi Kecelakaan, Asuransi Kesehatan. Sampel menggunakan 55 data nasabah.

- (7) Penerapan Data Mining Untuk Mengelompokkan Data Anggota Koperasi Simpan Pinjam (KSP) Credit Union (CU) Damai Sejahtera Medan dengan Menggunakan Metode K-Means** oleh (Monang dkk, 2021)

Peneliti menggunakan metode observasi dalam pengumpulan data dengan 3 atribut: loyalitas anggota, simpanan bulanan, pembayaran tepat waktu.

- (8) Penerapan Metode K-Medoids Guna Pengelompokan Data Usaha Mikro, Kecil dan Menengah (UMKM) Bidang Kuliner di Kota Yogyakarta** oleh (Utaminingsih dkk, 2024)

Penelitian bertujuan untuk mengelompokkan data UMKM bidang kuliner dengan memperhatikan aspek/variable:

- a. Pendidikan
- b. Kegiatan Usaha
- c. Tujuan Pemasaran
- d. Agunan
- e. Sarana Media Elektronik
- f. Modal bantuan pemerintah
- g. Pinjaman KUR
- h. Omset
- i. Kepemilikan asuransi Kesehatan
- j. Tenaga kerja laki-laki
- k. Tenaga kerja perempuan
- l. Lama pendirian usaha

(9) Pengelompokan Transaksi Pelanggan Kartu Kredit Menggunakan Algoritma Menggunakan Algoritma Clustering K-Means oleh (Bambang dkk, 2023)

Penelitian berfokus pada pengelompokkan nasabah kartu kredit berdasarkan perilakunya dengan menggunakan 2 cluster: minimum payment dan credit limit

(10) Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta) oleh (Nisriina dkk, 2022)

Penelitian berfokus pada pengelolaan buku di Perpustakaan LPP Yogyakarta yang belum optimal yang menimbulkan banyaknya ketersediaan buku dengan jumlah peminjam tidak seimbang sehingga dimungkinkan terdapat kekosongan apabila jumlah eksemplar tidak diperbanyak.

Tabel 2. 9. Tinjauan Pustaka

NO	PENELITI/TAHUN	JUDUL	JURNAL SUMBER	KONTRIBUSI
1	(Fadhillah Azmi Tanjung, Agus Perdana Windarto, M. Fauzan, 2021)	Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia	Jurasik, Vol 6 No 1 (2021) Link: https://tunasbangsa.ac.id/ejurnal/index.php/jurasik/article/view/271/250	<ul style="list-style-type: none"> • Penjelasan perhitungan yang mudah dipahami • Penentuan Data • Penentuan permasalahan dan tujuan penelitian

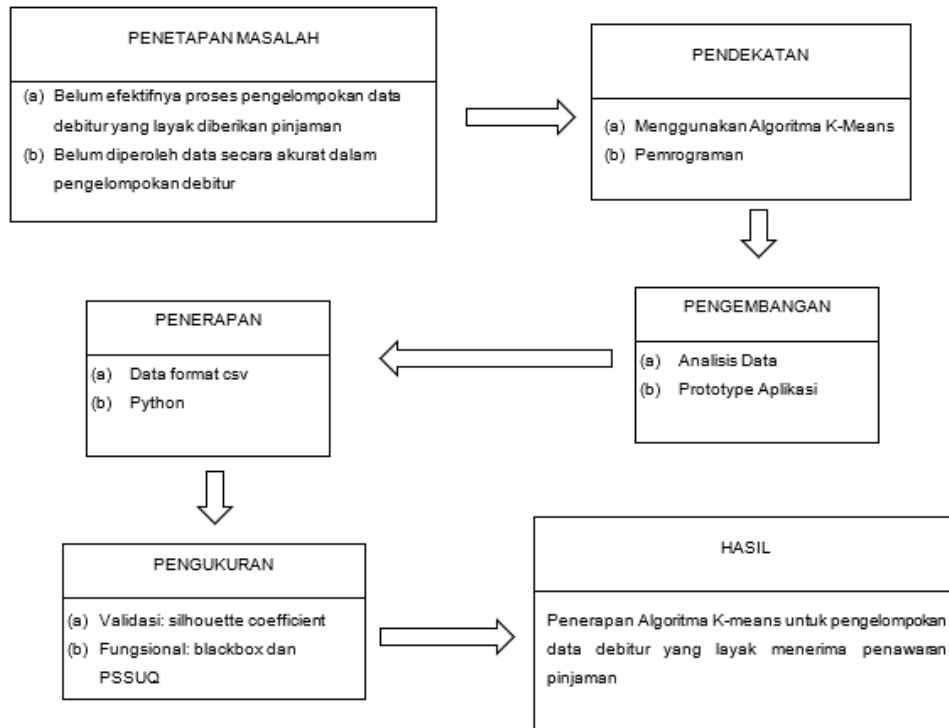
NO	PENELITI/TAHUN	JUDUL	JURNAL SUMBER	KONTRIBUSI
2	(Nanda Ayu Rahmalinda, Arief Jananto 2022)	Penerapan Metode K-means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru	Jurnal Tekno Kompak Vol 16, No 2 (2022) Link: https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1971/954	<ul style="list-style-type: none"> • Penentuan atribut • Metode analisis data • Konversi atribut
3	(Dede Nor Alfiansyah, Vinna Rahmayanti Setyaning Nastiti, Nur Hayatin,2022)	Penerapan Metode K-Means pada Data Penduduk Miskin Per Kecamatan Kabupaten Blitar	Repositor, Vol 4 No. 1 Februari 2022 Link: https://ejournal.u mm.ac.id/index.php/repositor/article/view/31083/14000	<ul style="list-style-type: none"> • Penentuan variable • Penentuan permasalahan
4	(Rike Limia Budiarti, Gina Cendana, 2022)	Klasifikasi Data Nasabah Kredit Pinjaman Menggunakan Data Mining Dengan Metode K-Means pada Mega Central Finance	Jurnal Akademia, Vol 14 No 2 (2022) Link: https://ojs.unh.ac.id/index.php/akademika/article/view/866	<ul style="list-style-type: none"> • Penentuan tujuan penelitian • Penentuan informasi data nasabah • Penentuan variable
5	(Juniar Hutagalung, Fifin Sonata, 2021)	Penerapan Metode K-Means untuk	Jurnal Media Informatika Budidarma Vol.	<ul style="list-style-type: none"> • Tujuan penelitan • Rumusan

NO	PENELITI/TAHUN	JUDUL	JURNAL SUMBER	KONTRIBUSI
		Menganalisis Minat Nasabah Asuransi	5 No. 3 (2021) Link: https://www.ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/3113	permasalahan
6	(Monang Situmorang, Ahmad Fitri Boy, Amrullah, 2021)	Penerapan Data Mining Untuk Mengelompokkan Data Anggota Koperasi Simpan Pinjam (KSP) Credit Union (CU) Damai Sejahtera Medan dengan Menggunakan Metode K-Means	Jurnal CyberTech Vol. 1 No. 2 (2021) Link: https://ojs.trigunadharmo.ac.id/index.php/jct/article/view/4567	<ul style="list-style-type: none"> • Tujuan penelitian • Rumusan permasalahan
7	(Utaminingsih Linarti, Anita Rahmawati, Anna Hendri Soleliza Jones, Lisna Zahrotun, 2024)	Penerapan Metode K-Medoids Guna Pengelompokan Data Usaha Mikro, Kecil dan Menengah (UMKM) Bidang Kuliner di Kota Yogyakarta	Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI) Vol 7, No. 1 (2024) Link: http://ejournal.sisfokomtek.org/index.php/jikom/article/view/2194	<ul style="list-style-type: none"> • Penentuan variable • Transformasi data

NO	PENELITI/TAHUN	JUDUL	JURNAL SUMBER	KONTRIBUSI
8	(Bambang Ary Nugroho, Indrarini Dyah Irawati, Aldo Lionel Saonard, 2023)	Pengelompokan Transaksi Pelanggan Kartu Kredit Menggunakan Algoritma Menggunakan Algoritma Clustering K-Means	e-Proceeding of Applied Science: Vol. 9 No. 1 (2023) Link: https://openlibrarpublication.telkomuniversity.ac.id/index.php/appliedscience/article/view/19376?btwaf=22815573	<ul style="list-style-type: none"> • Tujuan penelitian • Rumusan permasalahan
9	(Nisriina Nuur hasanah, Agus Sidiq Purnomo, 2022)	Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta)	Jurnal Teknologi Dan Sistem Informasi Bisnis Vol.4 No.2 Juli 2022 Hal. 300-311 Link: https://jurnal.unidha.ac.id/index.php/jteksis/article/view/499/300	<ul style="list-style-type: none"> • Tujuan penelitian • Metode penelitian diawali dengan pengumpulan data, processing data dan data integration.
10	(Akmal Fikri, Benedika Ferdian Hutabarat, Ulfa Khaira, 2023)	Komparasi K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman Financial Technology	Jurnal Ilmiah Media Sisfo Vol.17 No.2 (2023) Link: https://ejournal.unama.ac.id/index.php/mediasisfo/article/view/1373	<ul style="list-style-type: none"> • Tujuan Penelitian • Atribut Penelitian

C. Kerangka Pemikiran

Metode K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster yang bisa dijadikan metode pemecahan masalah penentuan debitur penerima pinjaman. Berikut adalah kerangka pemiksiran untuk memecahkan masalah penelitian ini yang dapat dilihat pada Gambar 2.6.



Gambar 2. 6.Kerangka Pemikiran

D. Hipotesis Penelitian

Berdasarkan permasalahan diatas, maka diduga penggunaan Algoritma K-Means dapat menentukan pengelompokkan debitur yang menerima penawaran pinjaman KUR.