BAB II
KERANGKA TEORITIS

A. Landasan Teori

1. Data Mining

Menurut Kambe dkk. (2011, p. 6) data mining adalah proses menemukan pola atau informasi yang berguna dari kumpulan data dalam jumlah yang besar. Data mining sering disebut juga sebagai Knowledge Discovery in Databases (KDD) yang mencakup berbagai tahapan seperti seleksi data, pra-pemprosesan, transformasi, penambangan data, dan evaluasi hasil.

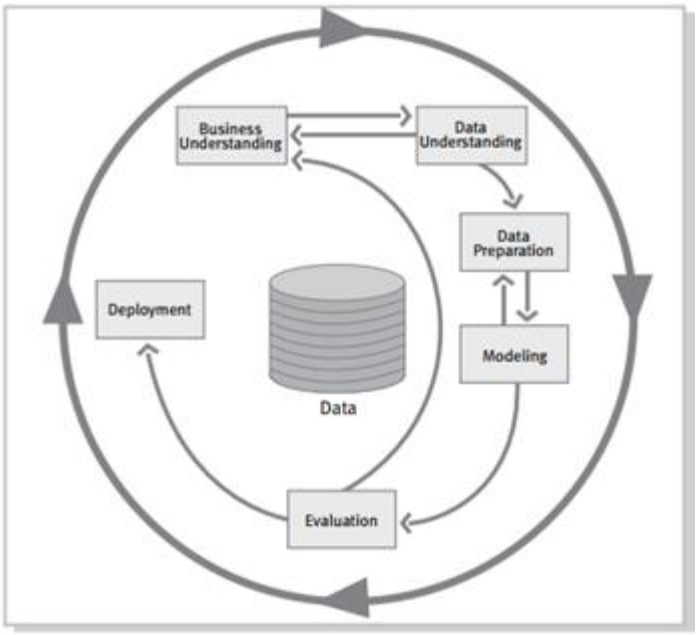
Berdasarkan buku Data Mining: Consepts and techniques, menurut (Kambe dkk, 2011, p. 83-117) data mining memiliki beberapa teknik utama sebagai berikut:

a) **Klasifikasi :** Menempatkan objek ke dalam kategori yang telah ditentukan. Algoritma yang digunakan adalah Decision Tree, Naïve Bayes, dan Neural Network;
b) ***Clustering*** **:** Mengelompokkan data berdasarkan kesamaan tanpa kategori yang telah ditentukan sebelumnya. Algorima yang digunakan adalah K-Means, DBSCAN, Hierarchical Clusteing;
c) **Asosiasi :** Memodelkan hubungan antara variabel untuk memperkirakan nilai masa depan. Algoritma yang digunakan adalah Linear Regression, ARIMA;
d) **Deteksi Anomali :** Mendeteksi data yang tidak biasa atau mencurigakan. Algortima yang digunakan adalah Isolation Forest, One-Class SVM

Berdasarkan teori di atas data mining adalah proses sistematis untuk menemukan pola yang berguna dalam data. Dengan berbagai teknik seperti klasifikasi, *Clustering*, dan asosiasi. Data mining saat ini telah menjadi alat yang penting dalam proses analisis data dan pengambilan keputusan.

2. CRISP-DM

Menurut Deshpande dkk. (2014, p. 17-18) dalam buku Predictive Analytics and Data Mining, CRISP-DM adalah pendekatan terstruktur dalam data mining yang memastikan praktik terbaik dan memungkinkan keterulangan proses, sehingga menjadi pilihan yang populer diberbagai industri.



Gambar 2.1 CRISP-DM

Metodologi ini terdiri dari enam tahapan utama yaitu :

a) ***Bussiness Understanding***

Tahapan ini adalah memahami kebutuhan pelanggan secara mendalam. Kegiatan yang dilakukan pada tahap ini adalah menentuka tujuan bisnis, menilai situasi ketersediaan sumber daya, menentukan tu pengumpulan data, dan menghasilkan rencana proyek.

b) ***Data Understanding***

Tahapan ini adalah tahap pemahaman data, yaitu mengidentifikasi, mengumpulkan, dan menganalisis kumpulan data yan dapat membantu untuk mencapai tujuan proyek. Kegiatan pada tahap ini adalah mengumpulkan data awal, menjelaskan data, menjelajahi data dan memverifikasi kualitas data.

c) ***Data Perparation***

Tahap ini sering disebut juga sebagai data mining, yaitu menyiapkan kumpulan data akhir untuk pemodelan. Kegiatan pada tahap ini adalah memperbaiki kualitas data agar sesuai dengan proses modeling yang akan dilakukan berikutnya.

d) ***Modeling***

Membuat dan menilai berbagai model berdasarkan beberapa teknik pemodelan yang berbeda. Pada tahap ini, ada empat tugas yaitu memilih teknik pemodelan, menghasilkan desai pengujian, membangun model, dan menilai model.

e) ***Evaluation***

Tahap evalusai ini melihat lebih luas model yang paling sesuai dengan bisnis dan yang harus dilakukan selanjutnya. Pada tahapan ini terdapat tiga kegiatan yang mewakili tahap evaluasi yaitu evaluasi hasil, proses peninjauan, dan penentuan tahap selanjutnya.

f) ***Deployment***

Tahap ini adalah tahap yang paling penting dari proses CRISP-DM dimana pada tahap ini memastikan bahwa model yang telah diuji dapat diimplementasikan dalam lingkungan bisnis nyata untuk memebrikan wawasan dan mendukung pengambikan keputusan.

3. K-Means

K-Means adalah suatu metode dalam analisis klaster yang digunakan untuk mengelompokan data ke dalam kumpulan-kumpulan homogen berdasarkan kemiripan karakteristik. Tujuan dari algortima K-Means adalah membagi data ke dalam sejumlah klaster, sehingga setiap titik data dalam suatu klaster memiliki kemiripan yang tinggi satu sama lain, sementara klaster yang berbeda memiliki karakteristik yang berbeda. Metode ini mencoba untuk meminimalkan varians intra-klaster dan memaksimalkan varian antar-klaster dengan perhitungan rata-rata dari masing-masing anggota klaster (Ma'ady dkk., 2024, p. 38).

Untuk menerapkan metode K-Means secara efektif, diperlukan pemahaman terhadap proses perhitungannya yang melibatkan beberapa tahapan utama. Tahapan-tahapan ini menjadi dasar dalam mengelompokan data secara sistematis dan konsisten ke dalam klaster yang sesuai. Dengan memahami proses perhitungan yang dilakukan dalam algoritma K-Means, pengguna dapat mengaplikasikan metode ini secara tepat dalam berbagai konteks analisis data. Adapun tahapan perhitungan algortima K-Means adalah sebagai berikut :

Langkah 1. Menentukan Nilai K atau Jumlah *Cluster*

Langkah utama yang paling penting dalam metode K-Means adalah membagi data sejumlah k *cluster*. Nilai k atau jumlah klaster ini wajib diinisiasi awal. Untuk mencari k, bisa berdasarkan dari nilai ambang yang ditentukan oleh pemangku kebijakan, atau jika tidak dapat menggunakan beberapa metode untuk menentukan nilai K yang paling optimal.

Pada contoh data di Tabel 2.1., akan dibagi ke dalam 2 kelompok atau 2 *cluster*. Oleh karena itu, nilai K pada percobaan ini adalah 2.

Langkah 2. Tentukan *Centroid*

Seperti yang diketahui sebelumnya bahwa *centroid* adalah nilai tengah dari suatu *cluster*. Nilai *centroid* bisa dipilih secara random, selama dalam rentang data yang akan dilakukan klsater. Cara termudah dalam menentukan *centroid* adalah mengambil nilai data set. Sebagai contoh pada data pelanggan tadi, ambil dua baris daya yang dijadikan untuk *centroid*. Banyaknya *centroid* berdasarkan banyaknya K atau banyaknya klaster yang telanh ditentukan pada Langkah 1.

Tabel 2.1 Data Pelanggan dan *centroid*-nya

No	Umur	Pendapatan (dalam USD)	
1	67	124.67	
2	22	150.773	
3	49	89.21	*Centroid culster 1*
4	45	171.565	
5	53	149.031	
6	35	144.848	
7	53	156.495	
8	35	193.621	
9	61	151.591	*Centroid culster 2*
10	28	174.646	

(Sumber : (Ma'ady *dkk.*, 2024)

Pada saat menentukan *centroid*, pastikan bukan data yang rentangnya berdekatan. Penentuan jumlah *centroid* sangan mempengaruhi beberapa kali iterasi yang perlu dilakukan untuk mendapatkan hasil klaster yang optimal.

Langkah 3. Menghitung jarank data dengan *centroid* masing-masing *cluster*

Langkah selanjutnya adalah menghitung jarak masing-masing data terhadap masing-masing *centroid* tiap klaster. Dalam menghitung jarang antar data, dapat menggunakan jarak euclidean, yang rumusnya dapat dilihat pada persamaan berikut.

$$d_{(i,n)} = \sqrt{\sum (d_i - n_i)^2}$$

Pada persamaan tersebut, *d* berarti *distance* dari data ke-*i* terhdap klaster ke-n. Di mana, apabila dalam satu data terdapat beberapa atribut, maka terdapat beberapa kali proses untuk mencari nilai kuadrat dari data ke-*i* yang dikurangi *centroid* $(d_i - n_i)$.

Sebagai contoh perhitungan dari persamaan tadi, dapat diimplementasikan pada data pelanggan terhadap masing-masing *centroid*-nya seperti yang tertera pada Tabel 2.1. Pada data nomor 1, yang memiliki nilai pada atribut umur = 67 dan nilai pada atribut pendapatan = 127.67. Dari Tabel 2.1. terdapat dua buah *centroid*, dimana *centroid* pertama memiliki nilai pada atribut umur = 49 dan nilai pada atribut pendapatan = 89,21. Dari No. 1 dan data *centorid* pertama, maka dapat diketahui bahwa jarak data 1 terhadap *centroid* klaster pertama dapat dihitung sebagaimana berikut :

A. Diketahui:
   a. Umur data No. 1 = 67
   b. Pendapatan data No. 1 = 124.67
   c. *Centroid cluster* 1 – umur = 49
   d. *Centroid cluster* 1 – pendapatan = 89.21
B. Ditanyakan
   Jarak data No. 1 terhadap *centroid cluster* 1.

C. Dijawab

$$d_{(i,n)} = \sqrt{\sum (d_i - n_i)^2}$$

$$d_{(1,1)} = \sqrt{(67 - 49)^2 + (124.67 - 89.21)^2}$$

$$d_{(1,1)} = \sqrt{324 + 1257.4116} = 39.767$$

Selanjutnya, perlu untuk menghitung jarak data No. 1 terhadap *centroid cluster* 2, maka perhitungannya adalah sebagaimana berikut :

A. Diketahui:
   a. Umur data No. 1 = 67
   b. Pendapatan data No. 1 = 124.67
   c. *Centroid cluster* 2 – umur = 61
   d. *Centroid cluster* 2 – pendapatan = 151.591
B. Ditanyakan
   Jarak data No. 1 terhadap *centroid cluster* 1.
C. Dijawab

$$d_{(i,n)} = \sqrt{\sum (d_i - n_i)^2}$$

$$d_{(1,1)} = \sqrt{(67 - 61)^2 + (124.67 - 151.691)^2}$$

$$d_{(1,1)} = \sqrt{36 + 724.740} = 27.582$$

Karena jarak data No. 1 terhadap *centroid cluster 2* (27.582) lebih kecil dibandingkan jarak data No. 1 terhadap *centroid cluster* 1 (38.767), maka data No. 1 diklasterkan ke *cluster* 2.

Langkah 4. Letakan data di *cluster* dengan jarak terdekat

Perhitungan jarak data terhadap *centroid* suatu cluster dilakukan terhadap semua data pada tabel 2.1, yang hasilnya dapat dilihat pada tabel 2.2. Pada tabel 2.2, EDC berarti *euclidean centroid to cluster* atau jarak ecludien data tehadap centroid suatu cluster. Sedangkan cluster terbaik diambil dari nilai minimum EDC 1 antara EDC 2. Apabila nilai EDC 1 lebih kecil daripada nilai EDC 2, maka data dimasukan ke cluster 1. Sedangkan apabila nilai EDC 2 lebih kecil daripada nilai EDC 1, maka data dimasukan ke dalam cluster 2.

Setelah melakukan perhitungan jarak data terhadap masing-masing cluster, maka masing-masing data dapat dikelompokkan pada suatu cluster. Sehingga, pada Tabel 2.2, dapat dilihat bahwa masing-masing data dikelompokan pada cluster terbaik sesuai dengan nilai jarak minimum suatu data terhadap suatu centroid cluster.

Tabel 2.2 Hasil perhitungan jarak masing-masing data terhadap centroid antar cluster

No	Umur	Pendapatan	EDC1	EDC2	Cluster terbaik
1.	67	124.67	39.767	27.582	2
2.	22	150.773	67.224	39.009	2
3.	49	89.21	0.000	63.525	1
4.	45	171.565	82.452	25.592	2
5.	53	149.031	59.955	8.400	2
6.	35	144.848	57.372	26.860	2
7.	53	156.495	67.404	9.383	2
8.	35	193.621	105.345	49.442	2
9.	61	151.591	63.525	0.000	2
10.	28	174.646	87.979	40.256	2

(Sumber : (Ma'ady *dkk.*, 2024)

Tahapan penetuan cluster ini belum berakhir. Pada langkah berikutnya perlu untuk menghitung rata-rata setiap anggota cluster untuk mendapatkan nilai centroid baru.

Langkah 5. Hitung rata-rata data untuk masing-masing cluster dan menentukan centroid baru

Dari hasil pemetaan cluster terbaik pada Tabel 2.2, maka dapat diketahui masing-masing anggota cluster. Dari hasil tersebut, selanjutnya perlu untuk dihitung rata-rata (*means*) dari setiap anggota di satu cluster. Seperti contoh pada Tabel 2.3 menunjukan anggota cluster 1, sekaligus nilai rata-ratanya.

Tabel 2.3 Anggota cluster 1 dan rata-rata baru

Data No.	Umur	Pendapatan
3	49	89.21
Rata-rata	49	89.21

(Sumber : (Ma'ady *dkk.*, 2024)

Rata-rata pada tabel 2.3 menjadi centroid baru dari cluster 1. Selanjutnya, perlu dihitung rata-rata dari setiap anggota cluster 2, yang hasilnya dapat dilihat pada Tabel 2.4.

Tabel 2.4 Anggota cluster 2 dan rata-rata baru

Data No.	Umur	Pendapatan
1	67	124.67
2	22	157.773
4	45	171.565
5	53	149.031
6	35	144.848
7	53	156.495
8	35	193.621
9	61	151.591
10	28	174.646
Rata-rata	44.333	157.471

(Sumber : (Ma'ady *dkk.*, 2024)

Dari langkah ini, maka dapat diketahui bahwa, nilai centroid masing-masing cluster berubah menjadi seperti berikut :

Nilai centroid iterasi-1 :

A. Centroid cluster 1

a. Umur : 49

b. Pedapatan : 89.21

B. Centroid cluster 2

c. Umur : 44.333

d. Pedapatan : 157.471

Karena centroid pada inisiasi awal dan centorid setelah perhitugan penentuan anggota cluster pada iterasi pertama berubah, maka perlu melakukan perhitungan jarak antara data dengan centroid-nya (langkah-3), sampai menghitung rata-rata anggota masing-masing cluster (langkah-5), sampai nilai centroid tidak berubah.

Langkah 6. Mengulangi Langkah 3 – Langkah 5, sampai Centroid Tidak Berubah

Seperti yang diketahui dari langkah 5, bahwa tahapan mengelompokan data ke masing-masing cluster, berdasarkan jarak data terhadap centroid cluster-nya perlu dilakukan berulang sampai salah satu kondisi berikut ini terpenuhi :

a. Nilai centroid tidak berubah, atau
b. Tidak ada anggota cluster yang berpindah lagi, atau
c. Pada suatu kasus, karena data terlalu besar, maka iterasi dapat berlangsung terus-menerus. Oleh sebab itu bisa juga diberikan *threshold* untuk menentukan jumlah iterasi yang dapat dilakukan. Jika menentukan *threshold* nilai iterasi, maka apabila penentuan anggota cluster berulang sampai ke iterasi sesuai *threshold* tersebut, proses K-Means dianggap selesai.

Pada percobaan sebelumnya, terjadi perubahan nilai centroid masing-masing cluster dari inisiasi awal dengan iterasi pertama. Maka dilakukan iterai kedua, yang dimulai dengan menghitung jarak masing-masing data terhadap centroid terbaru. Perlu diperhatikan bahwa pada iterasi kedua ini, menggunakan centroid terbaru, bukan nilai centroid yang ditentukan pada langkah 2. Hasil perhitungan jarak data terhadap centroid terbaru dapat pada iterasi kedua dilihat pada Tabel 2.5. Dalam menghitung jarak data, tetap menggunakan rumus sebelumnya yang ada pada langkah tiga.

Tabel 2.5 Hasil perhitungan jarak data pada iterasi kedua

No	Umur	Pendapatan	EDC1	EDC2	Cluster terbaik
1.	67	124.67	39.767	39.871	1
2.	22	150.773	67.224	23.316	2
3.	49	89.21	0.000	68.420	1
4.	45	171.565	82.452	14.110	2
5.	53	149.031	59.955	12.097	2
6.	35	144.848	57.372	15.699	2
7.	53	156.495	67.404	8.721	2
8.	35	193.621	105.345	37.335	2
9.	61	151.591	63.525	17.674	2
10.	28	174.646	87.979	23.701	2

(Sumber : (Ma'ady *dkk.*, 2024)

Penentuan centroid dari iterasi-2 yang didapat dari perhitungan rata-rata anggota masing-masing cluster

Tabel 2.6 Anggota cluster dari iterasi 2 dan perhitungan centroid baru

Data No.	Umur	Pendapatan
Anggota Cluster 1		
1.	67	124.67
3.	49	89.21
Rata-rata	58	106.94
Anggota Cluster 2		
2.	22	150.773
4.	45	171.565
5.	53	149.031
6.	35	144.848
7.	53	156.495
8.	35	193.621
9.	61	151.591
10.	28	174.646
Rata-rata	41.5	161.571

(Sumber : (Ma'ady *dkk.*, 2024)

Nilai centroid iterai-1 :

A. Centroid cluster 1

a. Umur : 49

b. Pedapatan : 89.21

B. Centroid cluster 2

a. Umur : 44.333

b. Pedapatan : 157.471

Nilai centroid iterasi-1 dan nilai centroid iterasi-2 ternyata berubah. Karena nilai centroid berubah, dan data No.1 bergeser dari cluster 2 ke cluster 1, maka dibutuhkan perulangan langkah 3 sampai langkah 5, atau iterasi ketiga. Hasil dari iterasi ketiga dilihat pada Tabel 2.7.

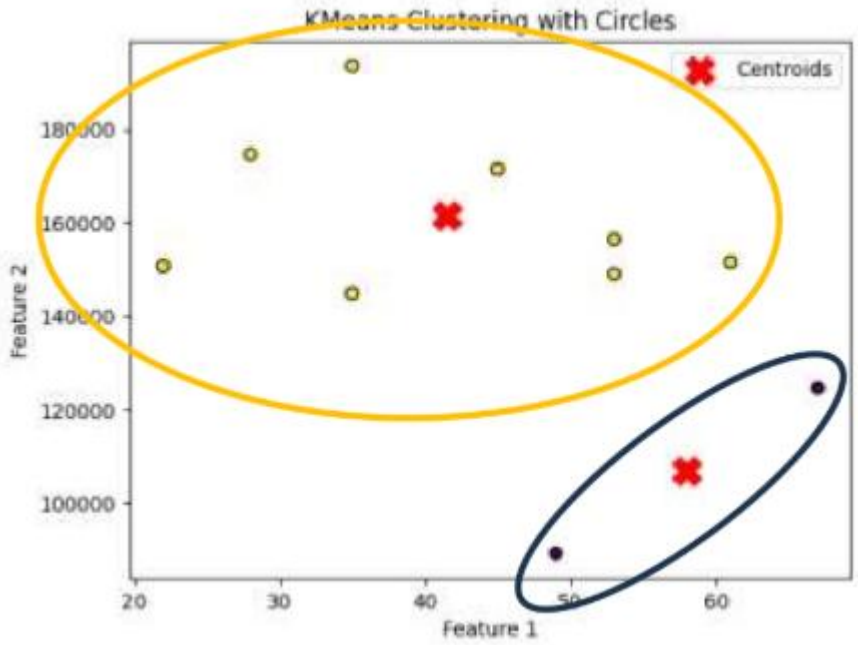
No	Umur	Pendapatan	EDC1	EDC2	Cluster terbaik
1.	67	124.67	19.883	44.855	1
2.	22	150.773	56.722	22.290	2
3.	49	89.21	19.883	72.749	1

No	Umur	Pendapatan	EDC1	EDC2	Cluster terbaik
4.	45	171.565	65.920	10.589	2
5.	53	149.031	42.387	17.015	2
6.	35	144.848	44.340	17.942	2
7.	53	156.495	49.807	12.571	2
8.	35	193.621	89.681	32.702	2
9.	61	151.591	44.751	21.906	2
10.	28	174.646	74.005	18.794	2

(Sumber : (Ma'ady *dkk.*, 2024)

Meskipun terdapat perubahan nilai baik dari EDC 1 dan EDC 2 dari Tabel 2.6 dan 2.7, namun pada tahap perulangan iterasi ini, lebih baik berfokus pada atribut Cluster terbaik. Pada masing-masing data dari No. 1 sampai No. 10, tidak ada perubahan cluster terbaik dari Tabel 2.6 dan Tabel 2.7. Ini mengindikasikan bahwa tidak ada anggota cluster yang berpindah. Jika anggota cluster 1 dan anggota cluster 2, logikasnya, apabila dihitung rata-rata anggotanya, maka nilai centroidnya pun akan tetap. Untuk itu proses penentuan cluster menggunakan K-Means, pada data pelanggan ini berhenti pada iterasi ketiga.

Dari hasil perhitungan tiga iterasi tersebut, maka dapat diketahui bahwa anggota cluster 1 adalah data No. 1 dan No. 3. Sedangkan sisanya adalah anggota dari cluster 2. Apabila divisualisasikan, maka letak centroid dan persebaran anggota cluster dapat dilihat pada Gambar 4.3. Pada Gambar 4.3, cluster 1 ditampilkan pada lingkaran hitam, dan cluster 2 ditampilkan pada lingkaran kuning. Centroid masing-masing cluster digambarkan dengan tanda silang merah.



Gambar 2.2 Hasil klasterisasi data pelanggan menggunakan K-Means

4. Sistem Pendukung Keputusan

Sistem Pendukung Keputusan (SPK) atau Decision Support System (DSS) adalah sebuah sistem yang dimaksudkan untuk mendukung para pengambil keputusan manajerial dalam situasi keputusan semi terstruktur. SPK juga dimaksudkan untuk menjadi alat bantu bagi para pengambil keputusan untuk memperluas kapabilitas di dalam pengambilan keputusan. SPK ditujukan untuk keputusan-keputusan yang memerlukan penilaian atau pada keputusan – keputusan yang sama sekali tidak dapat didukung oleh algoritma (Turban, Aronson dan Liang, 2005).

Menurut Keen (1978, p. 32) membagi keputusan berdasarkan keharusan keputusan dibuat dan cakupan keputusan tersbut, yaitu:

a) Keputusan terstruktur : sebuah keputusan terstruktur dapat merupakan keputusan yang dihasilkan oleh program komputer, keputusan terstruktur diambil untuk memecahkan masalah yang pernah terjadi sebelumnya;
b) Keputusan tidak terstruktur : keputusan yang diambil untuk memecahkan masalah baru atau sangat jarang terjadi, sehingga perlu dipelajari secara hati-

hati. komputer tetap dapat membantu pembuat keputusan, tetapi hanya dapat memberikan sedikit dukungan;

c) Keputusan semi terstruktur : merupakan keputusan diantara keputusan terstruktur dan tidak terstruktur.

Sistem pendukung keputusan bertujuan untuk menyediakan informasi, membimbing, memberikan prediksi serta mengarahkan kepada pengguna informasi agar dapat melakukan pengambilan keputusan dengan lebih baik.

Adapun komponen-komponen sistem pendukung keputusan menurut (Rahmansyah dan Lusinia, 2016, p. 8) adalah sebagai berikut :

a) Data Manegement

   Merupakan komponen yang mengelola data internal maupun external yang dibutuhkan untuk pengambilan keputusan. Di dalamnya biasanya terdapat database yang menyimpan data historis, data transaksi, maupun data refernsi dari luar organisasi.

b) Model Management

   Merupakan komponen yang berisi model matematis, analitis, maupun algoritma yang digunakan untuk memproses data. Model ini bisa berupa model statis, simulasi, optimasi atau bahkan algoritma kecerdasan buatan.

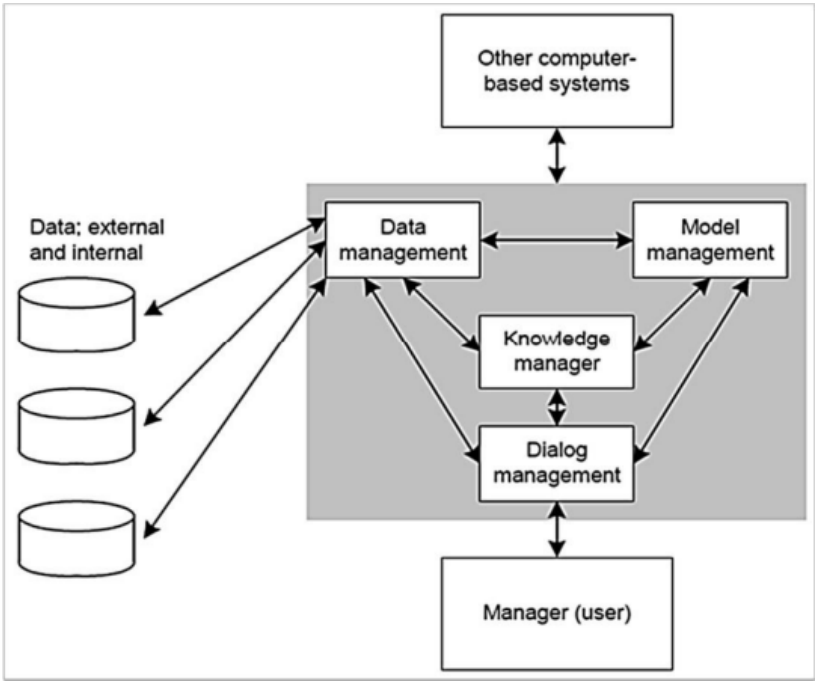
c) Communication

   Komponen ini berupa user interface yang menjadi penghubung antara pengguna dengan sistem. Fungsinya agar pengguna dapat dengan mudah mengakses data, menjalankan model, dan memahami hasil analisis.

d) Knowlage Managemen

   Komponen ini menyimpan dan mengelola pengetahuan, pengalaman, dan aturan-aturan yang relevan untuk mendukung keputusan. Berperan melengkapi data dan model dengan wawasan atau best practice yang sudah terbukti efektif.

Untuk lebih jelas memahami model konseptual SPK, perhatikan gambar 2.3 berikut :



Gambar 2.3 Model Konseptual

5. Evaluasi dan Peningkatan Kompetensi Guru

Evaluasi kompetensi guru merupakan suatu prose sistematis untuk menilai sejauh mana seorang guru telah mencapai standar kompetensi yang ditetapkan dalam dunia pendidikan. Menurut (Anif *dkk.*, 2020) dalam Jurnal Menajemen Pendidikan, evaluasi terhadap kompetensi guru perlu dilakukan untuk memastikan bahwa pelatihan dan program peningkatan kualitas yang diberikan benar-benar efektif. Menurut (Anif *dkk.*, 2020) beberpa langkah yang dapat dilakukan untuk meningkatkan kompetensi guru antara lain :

a) Menigkatkan kualitas materi pelatihan agar lebih sesuai dengan kebutuhan guruu dilapangan
b) Memilih instruktur yang kompeten, memiliki sertifikasi serta pengalaman mengajar yang baik
c) Menyediakan fasilitas pelatihan yang memadai agar proses pembelajaran dalam pelatihan dapat berjalan efektif
d) Menggunakan metode pembelajaran yang lebih inovatif
e) Melakukan evaluasi dan tindak lanjut setelah pelatihan untum memastikan bahwa pelatihan benar benar berdampak pada kompetensi guru.

Peningkatan kompetensi gruu bukan hanya tanggung jawab individu guru itu sendiri, tetapi juga melibatkan peran pemerintah dan lembaga pendidikan dalam menyediakan pelatihan yang efektif dan berkualitas (Anif *dkk.*, 2020).

6. Unified Modeling Language (UML)

Menurut Sumirat dkk. (2023, p. 73)mengatakan UML (Unified Modeling Language) adalah “Sebuah bahasa yang berdasar pada grafik/gambar untuk memvisualisasi, menypesifikasikan, membangun, dan pendokumentasian dari sebuah sistem pengembangan software berbasis OO (Object-Oriented).”. Merujuk pada apa yang dijelaskan oleh teori di atas UML adalah bahasa yang biasa digunakan untuk mempresentasikan analisa desain serta spesifikasi dalam proses pemrograman berorientasi pada objek. Sehingga sistem dapat mudah dikembangkan sesuai dengan alur yang disepakati.

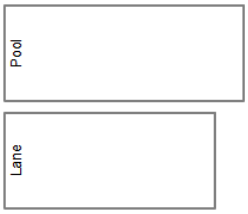
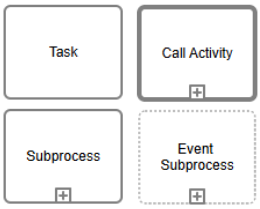
Berdasarkan definisi di atas, maka dapat disimpulkan bahwa UML adalah suatu bahasa yang sudah menjadi standar pada visualisasi dan juga pendokumentasian dalam melakukan spesifikasi pada sistem. Menurut:

a) BPMN Diagram

BPMN diagram adalah diagram yang digunakan untuk menggambarkan proses bisnis dengan notasi standar yang mudah dipahami oleh berbagai pihak, mulai dari analisis bisnis hingga pengembangan sistem. Dengan simbol-simbol :

Tabel 2.7 Simbol BPMN Diagram

	:	Event: sebuah event direpresentasikan dengan lingkaran. Events dapat berupa Start, Intermediate, atau End
	:	Gateway : merupakan elemen yang digunakan untuk mengontrol aliran proses berdasarkan kondisi atau aturan tertentu

	:	Data : Mempresentasikan informasi yang digunakan atau dihasilkan dalam suatu proses bisnis
	:	Swimlane: merupakan elemen yang digunakan untuk menggambarkan partisipan dan tanggung jawab dalam suatu proses bisnis, terdiri dari pool dan lane
	:	Activity: merepresentasikan pekerjaan (task) yang harus diselesaikan. Ada empat macam activity, yaitu task, looping task, sub process, dan looping subprocess.

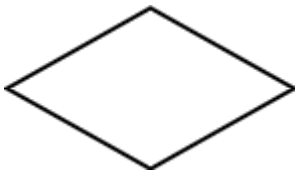
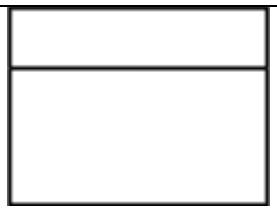
(Sumber : Sumirat dkk., 2023)

b) Activity Diagram

Activity diagram adalah diagram yang biasanya digunakan untuk melakukan pemodelan alur kerja atau proses dalam suatu sistem. Diagarm ini menunjukan urutan aktivitas atau tindakan yang dilakukan pada sistem. Activity diagram memiliki kegunaan dalam memvisualisasikan alur proses bisnis, menjelaskan langkat-langkah dalam suatu sistem dan memahami bagaimana aktivitas berinteraksi dan bergantung satu sama lain. Dengan simbol-simbol:

Tabel 2.8 Simbol Activity Diagram

	:	Status awal: bagaimana objek dibentuk atau diawali

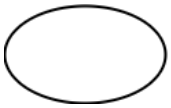
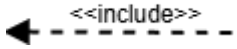
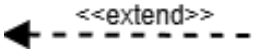
	:	Status akhir: bagaimana objek dibentuk dan diakhiri
	:	Aktivitas: Memperlihatkan bagaimana masing masing kelas saling berinteraksi satu sama lain
	:	Pencabangan: dimana ada polohan aktivitas yang lebih dari satu
	:	Penggabungan: dimana yang mana lebih dari satu aktivitas digabungkan menjadi satu
	:	Swimlane: memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi

(Sumber : Sumirat dkk., 2023)

c) Use Case Diagram

Use case diagram adalah diagram yang digunakan untuk menggambarkan proses interaksi antara pengguna (aktor) dengan sistem. Diagaram ini berfokus pada fungsionalitas yang disediakan oleh sistem dari sudut pandang pengguna. Use case diagram memiliki kegunaan untuk memahami kebutuhan fungsional sistem, mengidentifikasi aktor yang berinteraksi dengan sistem dan menjelaskan hubungan antara aktor dan fungsi yang dilakukan sistem. Dengan simbol-simbol:

Tabel 2.9 Simbol Usecase Diagram

	:	Aktor : Mewakili peran orang, sistem yang lain atau alat ketika berkomunikasi dengan sistem
	:	Use Case : Abstraksi dan interaksi antara sistem dan aktor
	:	Asosiasi : Penggambaran dari penghubung antara aktor dangan use case
	:	Generalisasi : Menunjukan spesialisasi aktor untuk dapat berpartisipasi dengan use case
	:	Menunjukan bahwa suatu use case seluruhnya merupakan fungsionalitas dari use case lainnya
	:	Menunjukan bahwa suatu use case merupakan tambahan fungsional dari use case lainnya jika suatu kondisi terpenuhi

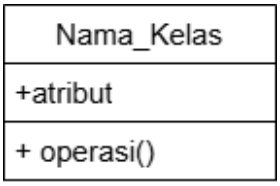
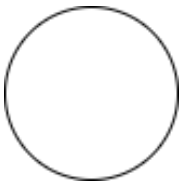
(Sumber : Sumirat dkk., 2023)

d) Class Diagram

Class diagram adalah diagram struktural yang digunakan untuk melakuka pemodelan kelas-kelas dalam sistem, atribut-atriutnya, metode-metodenya, serta hubungan antar kelas. Diagram ini memiliki kegunaan

untuk memvisualisasikan struktur sistem dalam bentuk kelas dan hubungannya, membantu memahamin bagaimana objek-objek yang ada dalam sistem berinteraksi, merancang database atau struktur data dalam sistem dan menjadikan dasara implementasi kode dalam pemrograman berorientasi objek. Dengan simbol-simbol:

Tabel 2.10 Simbol Class Diagram

	:	Kelas : menggambarkan kelas pada struktur sistem
	:	Antar muka : menggambarkan konsep interface dalam pemrograman
	:	Asosiasi : menggambarkan relasi secara umum
	:	Asosiasi berarah : menggambarkan relasi antar kelas
	:	Generalisasi : menunjukan relasi antar kelas dengan makna generalisasi - spesialisasi

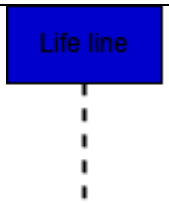
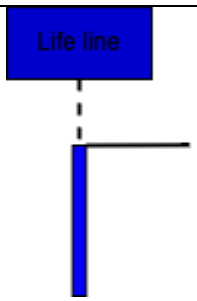
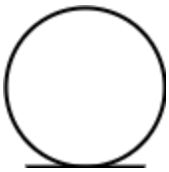
(Sumber : Sumirat dkk., 2023)

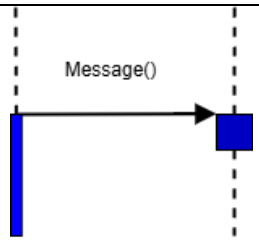
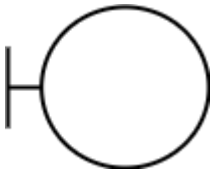
e) Sequence Diagram

Sequence diagram adalah diagram perilaku yang digunakan untuk melakuakn pemodelan interaksi antar objek dalam sistem secara

berurutan berdasarkan waktu. Diagram ini memiliki kegunaan untuk memvisualisasikan alur pesan atau interaksi antar objek dalam suatu skenario, memahami urutan eksekusi dalam sistem dan memantu mengidentifikasi kesalahan atau ketidakkonsistenan dalam alur logika sistem. Dengan simbol-simbol:

Tabel 2.11 Simbol Sequence Diagram

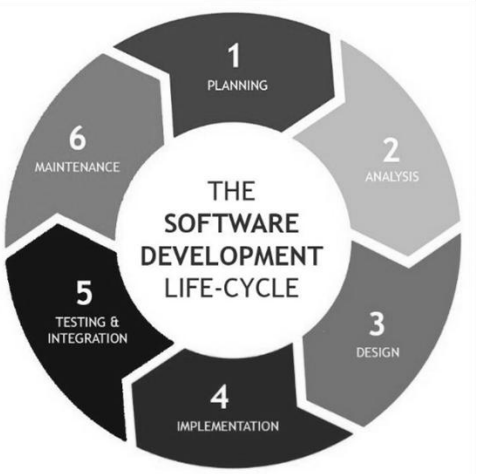
	:	Aktor : peran yang dilakukan oleh entitas untuk berinteraksi
	:	Garis hidup : mewakili peserta individu dalam interaksi
	:	Kontak aktifitas : menggambarkan periode dimana suatu elemen melakukan operasi
	:	Entitas : menggambarkan hubungan kegiatan yang akan dilakukan

	:	*Message* : menunjukkan pengiriman pesan
	:	*Boundary* : menggambarkan hubungan kegiatan yang akan dilakukan
	:	*Controler* : menggambarkan penghubung antara bounderi dengan tabel

(Sumber : Sumirat dkk., 2023)

7. Pengembangan Sistem SDLC

Menurut Smith (1992, p. 172) SDLC adalah kerangka kerja yang digunakan untuk memastikan bahwa perangkat lunak dikembangkan secara efisien dan sesuai dengan kebutuhan bisnis serta standar kualitas yang berlaku. Dengan menerapkan SDLC, organisasi dapat meminimalkan risiko proyek, mengoptimalkan biaya dan memastikan bahwa perangkat lunak yang dihasilkan memiliki kualitas tinggi serta sesuai dengan kebutuhan pengguna. SDLC mencakup beberapa tahap sebagaimana gambar 2.4 berikut :



Gambar 2.4 Tahapan Software Development Life Cycle

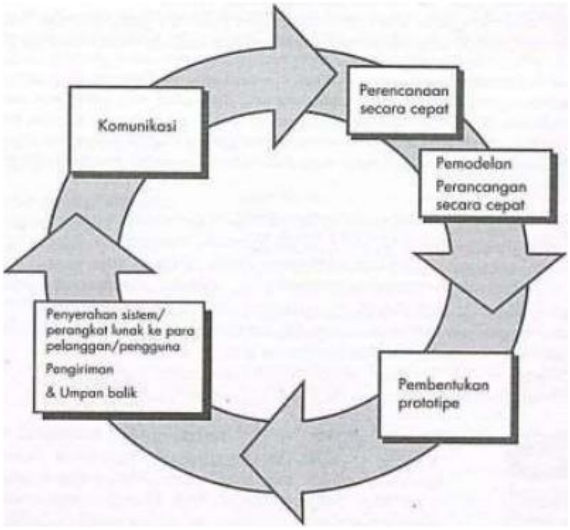
a) **Planing (Perencanaan)** : menentukan tujuan proyek, ruang lingkup, dan sumber daya yang dibutuhkan;
b) ***Requirement Analysis*** **(Analisis Kebutuhan)** : Mengidentifikasi kebutuhan pengguna dan spesifikasi sistem;
c) ***Design*** **(Perancangan)** : Membuat blueprint sistem termasuk arsitektur dan desai UI/UX;
d) ***Implementation*** **(Implementasi)** : Mengembangkan kode program berdsarkan desain yang telah dibuat;
e) ***Testing*** **(Pengujian)** : Memeriksa kualitas perangkat lunak untuk menemukan dan memperbaiki bug;
f) ***Deployment*** **(Penerapan)** : Menginstall dan menjalankan perangkat lunak dalam lingkungan produksi
g) ***Maintenance*** **(Pemeliharaan)** : Memperbaiki bug dan melakukan peningkatan sistem setelah rilis.

SDLC memastikan bahwa pengembangan perangkat lunak dilakukan dengan cara yang sistematis, efisien, dan menghasilkan produk berkualitas tinggi.

8. Prototyping Model

Metode prototyping adalah metode yang dimulai dengan pengumpulan kebutuhan pengguna, dalah hal ini pengguna dari perangkat yang dikembangkan adalah karyawan divisi helpdesk. Kemudian membuat sebuah rancangan kilat yang pada langkah selanjutnya akan dilakuka proses evaluasi kembali sebelum proses produksi dilakukan secara benar. Prototyping bukanlah merupakan sesuatu

yang lengkap, akan tetapi sesuatu yang perlu dan harus dievaluasi dan dilakukan modifikasi kembali. Segala perubahan dapat terjadi pada saat prototype dibuat untuk memenuhi kebutuhan pengguna dan pada saat yang sama memungkinkan pengembangan untuk lebih mamahami kebutuhan pengguna secara lebih baik (Pressman, 2012, hlm. 50).



Gambar 2.5 Prototyping model Sumber

Pembuatan prototype dimulai dengan dilakukannya komunikasi antar tim pengembang perangkat lunak dengan para pelanggan. Tim pengembangan perangkat lunak akan melakukan beberapa pertemuan-pertemuan dengan para pemangku kebijakan untuk mendefinisikan sasaran keseluruhan untuk perangkat lunak yang akan dikembangkan mengidentifikasi spesifikasi kebutuhan apapun yang saat ini diketahui dan menggambarkan dimana area-area definisi lebih jauh pada iterasi selanjutnay merupakan keharusan, iterasi pembuatan prototype direncanakan dengan cepan dan pemodelan (dalam bentuk “rancangan cepat”) dilakukan. Suatu rancangan cepat berfokus pada representasi semua aspek perangkat lunak yang akan terlihat oleh pengguna akhir misalnya rancangan antar muka pengguna atau format tampilan. (Pressman, 2012, p. 50)

Rancang cepat (quick design) akan memulai kontruksi pembuatan prototype, prototype kemudian akan diserahkan kepada para stakeholder dan kemudian akan melakukan evaluasi – evaluasi tertentu terhadap prototype yang telah dibuat sebelumnya, kemudian akhirnya akan memberikan umpan balik yang akan digunakan untuk memperhalus spesifikasi kebutuhan. Iterasi akan terjadi saat prototype diperbaiki untuk memenuhi kebutuhan dari para stakeholder,

sementara pada saat yang sama memungkinkan kita untuk lebih memahami kebutuhan apa yang kita kerjakan pada iterasi sebelumnya.

9. Database DBMS

Menurut Ling dan Lee, 2008, p. 11) *Database Management System* (DBMS) adalah perangkat lunak yang dirancang untuk mengelola, menyimpan, mengorganisir, dan mengambil data dalam basis data dengan cara yang efisien dan aman. DBMS bertindang sebagai perantara anatar pengguna dan basis data, memungkunkan manipulasi data melalui bahasa *query* seperti SQL (*Structured Query Language*). Fungsi utama dari DBMS menurut (Ling dan Lee, 2008, p. 25-33) adalah sebagai berikut :

a) Penyimpanan dan Pengambilan Data yaitu memungkinkan pengguna untuk menyimpan data dengan struktur yang sistematis serta mengambil informasi dengan cepat dan efisien.
b) Manajemen Kemanan Data yaitu mengontrol akses pengguna dan memberikan otoritas untuk mencegah akses tidak sah.
c) Manajemen Integritas Data yaitu memastikan bahwa data yang tersimpan dalam basis data tetap valid dan akurat melalui aturan integrita.
d) Manipulasi Data yaitu memungkinkan pengguna untuk menambahkan, menghapus, memperbarui, dan mengelola data menggunakan oprasi dasari SQL
e) Pemulihan dan Redundansi Data yaitu menyediakan mekanisme pencadangan dan pemulihan untuk mencegah kehilangan data akibar kegagalan sistem
f) Dukungan Multi User yaitu memungkinkan beberapa pengguna untuk mengakses dan memodifikasi data secara bersamaan tanpa konflik

Selain fungsi utama dari DBMS yang dibahas sebelumnya, S.K Singh membagi DBMS menjadi beberapa komponen antara lain yaitu, *database engine, query processor, storage manager, transaction manager, dan user interface.*

10. Python

Python merupakan bahasa pemrograman tingkat tinggi yang bersifat umum dan dirancang dengan filosofi kemudahan dalam penulisan serta keterbacaan kode. Menurut (Maesaroh dkk., 2024, p.1) “Python merupakan bahasa pemrograman tingkat tinggi yang mudah dipelajari karena sintaksnya sederhana dan mudah dibaca”. Python bersifat open source, multiplatform, dan mendukung paradigma pemrograman prosedural maupun berorientasi objek.

Python banyak digunakan dalam pengembangan perangkat lunak, analisis data, pembelajaran mesin, dan berbagai aplikasi komputasi lainnya. Keunggulan Python terletak pada sintaksis yang ringkas dan pustaka yang sangat lengkap seperti numpy, pandas, dan scikit-learn yang sangat berguna dalam pengolahan data dan penerapan algoritma klasterisasi seperti K-Means.

B. Tinjauan Studi

Penelitian ini merujuk pada beberapa penelitian betikut :

1. (Hendrastuty, 2024) **Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa**

   Berdasarkan analisis yang telah dilakukan pada Penelitian ini bertujuan untuk mengevaluasi hasil pembelajaran siswa dengan memanfaatkan metode data mining menggunakan algoritma K-Means Clustering. Evaluasi ini dilakukan untuk mengidentifikasi pola tersembunyi dalam data hasil pembelajaran siswa, serta untuk mengelompokkan siswa berdasarkan tingkat pencapaian atau karakteristik pembelajaran mereka. Pengelompokan ini bertujuan memberikan informasi yang bermanfaat bagi guru dan pihak sekolah dalam proses pengambilan keputusan pendidikan. Metode yang digunakan dalam penelitian ini adalah klasterisasi menggunakan algoritma K-Means, yang terbukti mampu membentuk dua kelompok siswa, yaitu kelompok C0 (Siswa Rajin) sebanyak 63 siswa dan kelompok C1 (Siswa Sangat Rajin) sebanyak 91 siswa. Hasil evaluasi menggunakan silhouette score menunjukkan nilai tertinggi sebesar 0,9168 pada jumlah klaster 2, yang mengindikasikan bahwa pengelompokan yang dilakukan memiliki kualitas yang baik. Penggunaan silhouette score sebagai metrik evaluasi juga membantu dalam menentukan jumlah klaster optimal serta memberikan gambaran yang lebih jelas dalam interpretasi hasil klasterisasi data pembelajaran siswa.
2. (Safnita *dkk.*, 2024) **PENERAPAN ALGORTIMA K-MEANS DALAM PENGKLASTERAN HASIL EVALUASI AKADEMIK MAHASISWA**

   Berdasarkan analisi yang telah dilakukan pada Penelitian ini bertujuan untuk mengelompokkan mahasiswa berdasarkan hasil evaluasi akademik dengan menerapkan algoritma K-Means Clustering sebagai bagian dari teknik data mining. Penelitian ini dilatarbelakangi oleh banyaknya data akademik yang tersimpan dalam sistem informasi berbasis komputer, yang jika diolah dengan tepat dapat memberikan pengetahuan berharga bagi pihak institusi. K-Means merupakan algoritma pengelompokan iteratif yang mempartisi himpunan data ke dalam sejumlah klaster yang telah ditentukan sebelumnya (nilai K). Metode ini dikenal mudah

diimplementasikan, cepat, serta banyak digunakan dalam berbagai praktik analisis data. Dataset yang digunakan berasal dari Fakultas Teknik, Program Studi Teknik Informatika, Universitas Islam Riau, yang mencakup 180 data mahasiswa dari semester 1 hingga semester 4. Hasil penelitian menghasilkan empat kelompok mahasiswa, yaitu: klaster mahasiswa berprestasi sebanyak 104 orang (57,72%), mahasiswa berpotensi berprestasi sebanyak 62 orang (34,41%), mahasiswa berpotensi bermasalah sebanyak 10 orang (5,55%), dan mahasiswa bermasalah sebanyak 4 orang (2,22%). Hasil ini diharapkan dapat memberikan pengetahuan baru bagi pihak akademik, serta menjadi model acuan dalam memantau dan memprediksi perkembangan performa akademik mahasiswa.

3. (Kurniawan dkk, 2023) **PENERAPAN ALGORITMA K-MEANS DAN FUZZY C-MEANS DALAM MENGELOMPOKAN PRESTASI SISWA BERDASARKAN NILAI AKADEMIK**

   Berdasarkan analisis yang telah dilakukan pada penelitian ini bertujuan untuk mengelompokkan prestasi siswa berdasarkan nilai akademik dengan menerapkan algoritma K-Means dan Fuzzy C-Means, serta membandingkan hasil dari kedua algoritma untuk menentukan metode terbaik. Latar belakang penelitian ini adalah adanya kesulitan pihak sekolah dalam menentukan siswa berprestasi sejak ditiadakannya Ujian Nasional pada tahun 2020 akibat pandemi COVID-19. Padahal, SMK PGRI 2 Karawang secara rutin memberikan beasiswa pendidikan dan merekomendasikan siswa-siswinya kepada perusahaan mitra berdasarkan capaian akademik. Data yang digunakan dalam penelitian ini merupakan nilai rapor siswa dari semester 1 hingga semester 6. Hasil klasterisasi dengan algoritma K-Means menunjukkan bahwa terdapat 52 siswa berprestasi, 25 siswa dengan prestasi sedang, dan 28 siswa tidak berprestasi. Sementara itu, hasil dari algoritma Fuzzy C-Means menghasilkan 48 siswa berprestasi, 29 siswa dengan prestasi sedang, dan 28 siswa tidak berprestasi. Uji coba menggunakan perangkat lunak RapidMiner Studio hanya memungkinkan penggunaan algoritma K-Means (karena keterbatasan fitur), dan menghasilkan 49 siswa berprestasi, 27 siswa dengan prestasi sedang, serta 29 siswa tidak berprestasi. Berdasarkan evaluasi menggunakan metode Davies-Bouldin Index (DBI), algoritma K-Means menunjukkan performa terbaik dengan nilai DBI yang mendekati nol dan tidak bernilai negatif. Hal ini mengindikasikan bahwa K-Means lebih optimal dalam proses pengelompokan data prestasi siswa dibandingkan Fuzzy C-Means pada studi ini.

4. (Silvia Ningsih, 2022) **PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN KELAS KELOMPOK BIMBINGAN BELAJAR TAMBAHAN**

   Berdasarkan anailisi yang dilakukan padaPenelitian ini dilakukan untuk mendukung pelaksanaan program belajar tambahan yang dilakukan di luar kegiatan intrakurikuler sekolah, atau yang biasa dikenal dengan sebutan program belajar tambahan sore. Program ini dirancang secara khusus untuk meningkatkan pemahaman siswa terhadap materi pelajaran, terutama dalam rangka persiapan menghadapi ujian tengah semester, ujian akhir semester, hingga ujian akhir nasional. Pembelajaran dilakukan setelah jam sekolah berakhir dan dibimbing langsung oleh guru mata pelajaran dari sekolah yang bersangkutan. Dalam proses pengelompokan kelas untuk program belajar tambahan ini, digunakan algoritma *K-Means Clustering* untuk mengklasifikasikan siswa ke dalam kelompok belajar yang sesuai. Sampel data yang digunakan dalam penelitian ini terdiri dari 26 siswa jurusan IPA. Tujuan penerapan metode klasterisasi ini adalah untuk memudahkan penentuan kelompok belajar yang homogen berdasarkan karakteristik nilai atau tingkat pemahaman siswa, sehingga kegiatan tambahan belajar menjadi lebih efektif dan terarah sesuai dengan kebutuhan masing-masing siswa.

5. (Hutagalung, 2022) **PEMETAAN SISWA KELAS UNGGULAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING**

   Berdasarkan anailisi yang dilakukan pada penelitian ini bertujuan untuk mengelompokkan data siswa kelas unggulan agar proses pembelajaran di sekolah dapat difasilitasi sesuai dengan kemampuan masing-masing siswa. Mengingat volume data yang besar dan terus bertambah setiap tahunnya, pengolahan data secara manual menjadi tidak efektif dan efisien. Oleh karena itu, diperlukan sistem berbasis komputer dengan pendekatan data mining untuk membantu pengambilan keputusan, khususnya dalam mengelompokkan siswa yang memiliki potensi atau prestasi unggul di bidang tertentu. Dalam penelitian ini, digunakan algoritma K-Means Clustering untuk mempercepat proses pengelompokan siswa berdasarkan nilai rapor. Sampel yang digunakan berjumlah 120 siswa dari SMK Raksana 2 Medan. Dengan menerapkan nilai centroid dan mencari jarak terdekat dalam perhitungan klaster, diperoleh tiga kelompok utama, yaitu: klaster 1 sebanyak 10 siswa dengan penguasaan pemrograman Android, klaster 2 sebanyak 62 siswa dengan penguasaan pemrograman Web, dan klaster 3 sebanyak 48 siswa dengan penguasaan pemrograman Desktop. Implementasi metode K-Means berbasis web dalam pengolahan data ini dirancang untuk mempermudah pendidik dalam mengambil

keputusan secara tepat dan cepat, serta memanfaatkan data secara optimal guna mendukung efektivitas proses pembelajaran di kelas unggulan.

6. (Saintikom *dkk.*, 2023) **Penerapan Algoritma K-Means dan Analisisnya untuk Menentukan Kebijakan Strategis Penyelesaian Studi Mahasiswa**

   Berdasarkan analisis yang dilakukan pada penelitian ini Peningkatan jumlah mahasiswa yang menyelesaikan studi setiap tahun di Program Studi PTIK UIN Bukittinggi menghasilkan akumulasi data yang besar dan menimbulkan pertanyaan penting: “Pengetahuan apa yang dapat diperoleh dari tumpukan data tersebut?”. Untuk menjawab pertanyaan ini, penelitian dilakukan dengan menerapkan teknik *data mining* sebagai bagian dari proses *Knowledge Discovery in Database* (KDD). Permasalahan yang dihadapi program studi mencakup penurunan kualitas tugas akhir mahasiswa dibandingkan dengan tahun-tahun sebelumnya, kurangnya sosialisasi mengenai judul dan arah penelitian (roadmap) kepada dosen dan mahasiswa, serta ketidaksesuaian judul tugas akhir dengan empat domain keterampilan utama di prodi PTIK. Untuk mengatasi kendala tersebut, digunakan metode *Clustering* dengan algoritma *K-Means* dalam kerangka metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Hasil penerapan algoritma *K-Means* menghasilkan tiga klaster mahasiswa berdasarkan data tugas akhir yang dianalisis. Setiap klaster dianalisis untuk mengidentifikasi pola dan permasalahan yang ada, sehingga program studi dapat menyusun rekomendasi dan kebijakan strategis yang tepat dalam mendukung penyelesaian studi mahasiswa. Temuan ini memberikan dasar pengambilan keputusan yang berbasis data untuk meningkatkan mutu akademik dan kesesuaian tema penelitian dengan kompetensi program studi.

7. (Isnanto dan Widodo, 2021) **PENERAPAN DATA MINING PADA PENERIMAAN MAHASISWA BARU DENGAN ALGORITMA K-MEANS CLUSTERING**

   Penelitian ini bertujuan untuk mengelompokkan data akademik mahasiswa berdasarkan nilai IPK semester 1 dan 2 menggunakan metode *Clustering* dengan algoritma *K-Means*. Data berasal dari Politeknik STMI Jakarta periode 2017–2020 dan diproses menggunakan RapidMiner. Hasil klasterisasi menunjukkan bahwa pada program studi Administrasi Bisnis Otomotif terbentuk 3 klaster, dengan klaster terbaik didominasi lulusan SMA jurusan IPA dan IPS. Program Sistem Informasi Industri Otomotif dan Teknik Industri Otomotif masing-masing menghasilkan 2 klaster, dengan klaster terbaik didominasi lulusan SMA IPA dan SMK Teknik Mesin. Sementara itu, pada program studi Teknik Kimia Polimer terbentuk 6 klaster, dan klaster terbaik seluruhnya berasal dari SMA jurusan IPA..

8. (Marpaung dkk, 2024) **Penerapan Data Mining Dalam Menentukan Tingkat Kedisiplinan Karyawan Perhotelan Menggunakan Algoritma K-Means Clustering**
   Penelitian ini bertujuan untuk menentukan tingkat kedisiplinan karyawan Hotel Halay menggunakan algoritma K-Means Clustering. Sebelumnya, penilaian disiplin dilakukan secara manual oleh manajer melalui observasi, yang dinilai kurang efektif dan profesional. Dengan menerapkan metode clustering, karyawan dikelompokkan ke dalam tiga klaster: Sangat Baik (C1), Baik (C2), dan Cukup (C3). Dari 10 data karyawan yang dianalisis, diperoleh 6 karyawan dalam klaster C1, 2 karyawan dalam klaster C2, dan 2 karyawan dalam klaster C3. Hasil ini digunakan sebagai dasar dalam memberikan penghargaan maupun pembinaan terhadap karyawan, serta mendukung peningkatan kualitas pelayanan hotel secara keseluruhan.
9. (Handayani dkk, 2024) **Implementasi Metode K-Medoids Dalam Pengelompokkan Kepuasan Masyarakat Terhadap Pelayanan Rumah Sakit**
   Berdasarkan analisis pada penelitian ini dilakukan untuk mengelompokkan tingkat kepuasan masyarakat terhadap pelayanan rumah sakit menggunakan metode K-Medoids. Penelitian ini memanfaatkan sembilan variabel terkait mutu pelayanan, seperti persyaratan, prosedur, waktu, biaya, kompetensi pelaksana, dan penanganan pengaduan. Hasil pengelompokan menghasilkan empat kategori kepuasan, yaitu sangat baik, baik, kurang baik, dan tidak baik. Evaluasi menggunakan Silhouette Coefficient menunjukkan struktur klaster yang kuat dengan nilai antara 0,9 hingga 1,0. Dari implementasi sistem berbasis website dan analisis terhadap 400 data, diketahui bahwa mayoritas masyarakat termasuk dalam klaster 4 dengan kategori sangat baik, khususnya berdasarkan akumulasi hasil pada 26 data percobaan awal di RSUD Dr. Soedarso Pontianak..
10. (Ningrum dkk, 2024) **Penerapan Metode K-Means dan Euclidean Distance Untuk Seleksi Metode Judul Tugas Akhir**
    Berdasarkan analisis pada penelitian ini dilakukan untuk mengetahui sebaran bidang kajian dan metode yang digunakan dalam judul skripsi mahasiswa pada salah satu perguruan tinggi swasta di Bogor, dengan tujuan membantu bagian akademik dalam mengevaluasi dan menyetujui pengajuan judul skripsi secara lebih selektif. Penelitian ini menggunakan metode K-Means dengan pengukuran jarak Euclidean Distance terhadap data judul skripsi mahasiswa tahun 2019 hingga 2022. Proses pengolahan data dilakukan melalui tahap transformasi judul menjadi atribut metode dan bidang kajian, dilanjutkan dengan normalisasi agar data dapat digunakan dalam proses klasterisasi. Hasil penelitian menghasilkan dua klaster utama, yaitu klaster dengan

metode Data Mining dan klaster dengan metode Decision Support System (DSS), dengan nilai Silhouette Coefficient sebesar 0,62 yang menunjukkan struktur klaster berada dalam kategori sedang (medium structure).

Penelitian rujukan di atas kemudian dapat disajikan secara ringkas sebagaimana Tabel 2.12 berikut :

Tabel 2.12 Tinjauan Penelitian

No	Peneliti/Tahun	Judul	Jurnal Sumber	Kontribusi
1	Nirwana Hendrastuty, tahun 2021	Penerapan Data mining menggunakan algoritma k-means clustering dalam evaluasi hasil pembelajaran siswa	https://e.publication.diskoplampung.com/index.php/jima-ilkom/article/view/26	Memberikan pemahaman bagaimana cara perhitungan K-Means berdasarkan nilai yang dihasilkan dari proses penilaian sebelumnya
2	Fitri Safnita dkk, tahun 2024	Penerapan Algoritma K-Means dalam pengklasteran hasil evaluasi akademik mahasiswa	https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/360	Memeberikan pemahaman dalam penggunaan algoritma clustering
3	Salim Kurniawan dkk, tahun 2023	Penerapan Algoritma K-Means dan Fuzzy C Means Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik	http://repository.ubpkarawang.ac.id/id/eprint/2905/	Menunjukan cara penggunaan algortima K-Means dapat diperhitungkan menggunakan rapid minder studio senganka

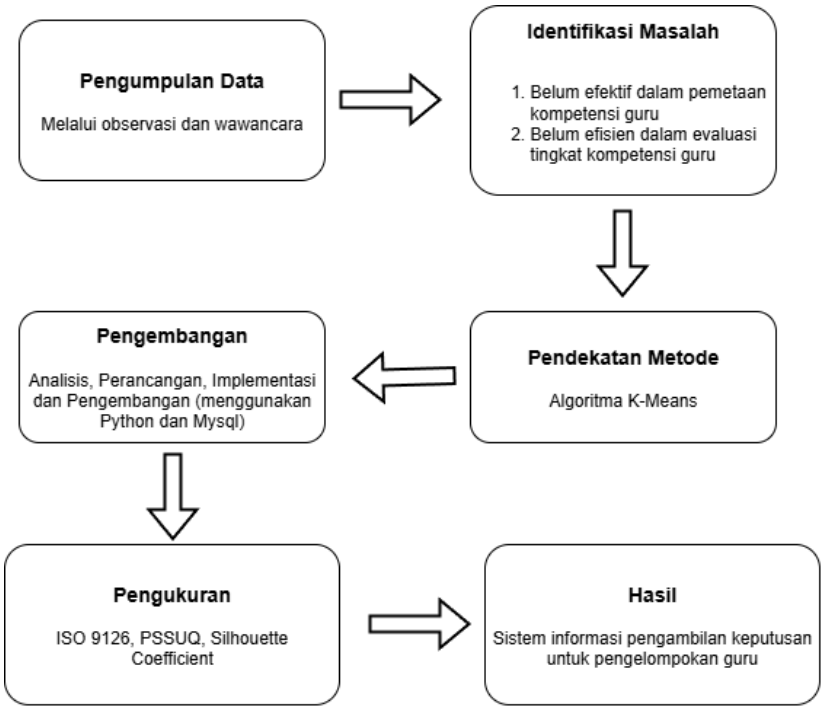
No	Peneliti/Tahun	Judul	Jurnal Sumber	Kontribusi
4	Silvia Ningsing, tahun 2022	Penerapan Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan	https://pdfs.semanticscholar.org/07cf/53542e4fef175a9892092d55ff3076f46ccc.pdf	Kontribusi yang dapat diambil dari penelitian ini adalah bagaimana cara mengklaster data berdasarkan nilai
5	Hutagalung, tahun 2022	Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering	https://scholar.google.co.id/scholar_url?url=https://jurnal.mdp.ac.id/index.php/jatisi/article/download/1516/671&hl=en&sa=X&ei=Mm6QaN7HNYrO6rQPh4OHiAw&scisig=AAZF9b9knlZIHBISHDTX6P3midFu&oi=scholarr	Membantu saya dalam melakukan clusterisasi dengan tiga cluster
6	Saintikom dkk, tahun 2023	Penerapan Algoritma K-Means dan Analisanya untuk Menentukan Kebijakan Strategis Penyelesaian Studi Mahasiswa	https://ojs.trigunadharma.ac.id/index.php/jis/article/view/8461	Kontribusi yang dapat diambil dari penelitian ini adalah bagaimana cara penerapan metodologi CRIPS-DM
7	Isnanto dkk, tahun 2024	Penerapan Data Mining Pada Penerimaan Mahasiswa Baru	https://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/367	Menunjukan bahwa motode yang digunakan efektif dalam

No	Peneliti/Tahun	Judul	Jurnal Sumber	Kontribusi
		Dengan Algoritma K-Means Clustering		menyesuaikan individu dengan kebutuhan yang relevan.
8	Marpaung dkk, tahun 2024	Penerapan Data Mining Dalam Menentukan Tingkat Kedisiplinan Karyawan Perhotelan Menggunakan Algoritma K-Means Clustering	https://ejournal.sisfokomtek.org/index.php/jikom/article/view/2905	Penelitian ini menunjukan bagaimana cara menggunakan algoritma K-Means Clustering
9	Hamdayani dkk., tahun 2024	Implementasi Metode K-Medoids Dalam Pengelompokkan Kepuasan Masyarakat Terhadap Pelayanan Rumah Sakit	http://ejurnal.seminar-id.com/index.php/josh/article/view/5331	Menunjukan bagaimana cara melakukan memetakan data dengan jumlah yang cukup besar
10	Ningrum dkk, tahun 2023	Penerapan Metode K-Means dan Euclidean Distance Untuk Seleksi Metode Judul Tugas Akhir	https://journal.global.ac.id/index.php/AJCSR/article/view/10766	Kontribusi yang bisa diambil dari penelitian ini adalah bagaimana cara penerapan perhitungan silhouette coefficient

Berdasarkan tinjauan penelitian yang telah dipaparkan di atas, diperoleh dasar pengetahuan yang menjadi rujukan dalam pelaksanaan penelitian ini. Rujukan tersebut berkontribusi dalam memberikan pemahaman mengenai permasalahan penilaian kompetensi guru dan proses pengelompokannya. Pada penelitian ini digunakan algoritma K-Means Clustering, dengan kriteria yang mengacu pada blanko penilaian kompetensi guru tahun 2023–2024, mencakup muatan materi dan cara penyajian materi saat kegiatan belajar-mengajar di kelas, kemudahan pemahaman (delivery) bagi siswa, manajemen kelas, serta proses penilaian yang dilaksanakan..

C. Kerangka Berfikir

Berdasarkan dukungan landasan teoritis yang diperoleh dari eksplorasi teori yeng dijadikan rujukan penelitian, maka dapat disusun kerangka berfikir sebagai berikut :



Gambar 2.6 Kerangka Penelitian

Kerangka berfikir pada gambar 2.6 dapat dijelaskan sebagai berikut :

1. Penelitian ini diawali dengan munculnya permasalah melalui pengumpulan data dengan cara observasi dan wawancara yang dilakukan pada objek permasalahan

2. Pendekatan metode yang digunakan adalah metode K-Means klastering yang berfungsi untuk mengklasterisasi dari data sebelumnya menjadi kelompok berdasarkan keanggotaan
3. Pengembangan melalui analisis dan desain serta implementasi dan pengembangan
4. Pengukuran menggunakan ISO 9126, PSSUQ, dan Silhouette Coefficient
5. Hasil yang dikeluarkan yaitu sistem informasi pengambilan keputusan untuk pengelompokan guru.

D. Hipotesis Penelitian

Pada penelitian yang membahas tentang pemilihan jurusan dengan menerapkan algoritma K-Means dalam penelitiannya membahas tentang bagaimana cara melakukan klasterisasi data siswa untuk menentukan kelas unggulan. Di mana pada penelitian tersebut algoritma K-Means dianggap mampu mengatasi permasalahan dalam proses pemetaan siswa untuk kelas unggulan. Hal ini dapat memperkuat penelitian ini dalam penerapan K-Means Clustering untuk melakukan pemetaan kompetensi guru (Hutagalung, 2022).

Berdasarkan saran dari penelitian tersebut dan permasalahan yang dihadapi, yaitu belum akuratnya pengelompokan guru berdasarkan tingkat kompetensinya, maka diperlukan suatu pendekatan untuk mengatasi hal tersebut. Dalam teori data mining, terdapat metode untuk melakukan klasterisasi data berdasarkan variabel-variabel tertentu. Salah satu metode klasterisasi tersebut adalah K-Means Clustering, yang akan menghasilkan pengelompokan data berdasarkan kedekatan nilai antar data. Metode ini digunakan untuk mengelompokkan guru berdasarkan kategori kompetensi yang dinilai, yaitu penyajian materi, delivery, manajemen kelas, dan penilaian.

Dengan demikian, hipotesis pada penelitian ini adalah bahwa penerapan algoritma K-Means diduga dapat mengelompokkan guru berdasarkan nilai kompetensi guna mendukung proses evaluasi dan peningkatan kompetensi guru di Sekolah Menengah Kejuruan.