

## BAB II KERANGKA TEORITIS

### A. Landasan Teori

#### 1. Konsep Data Mining

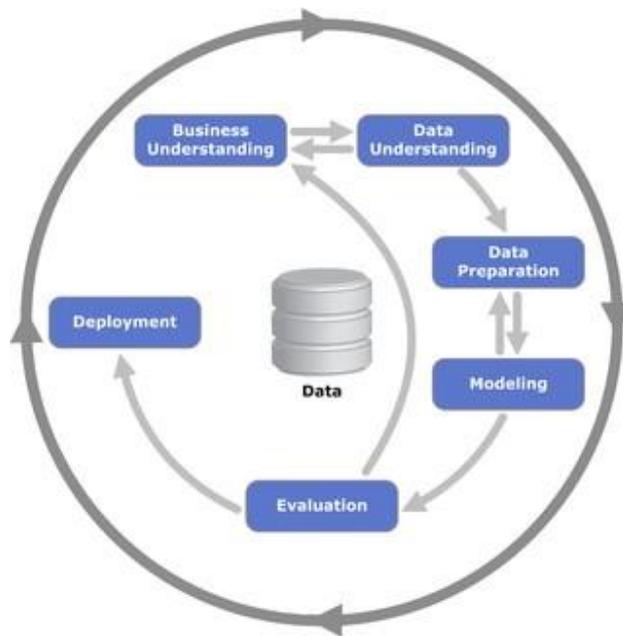
Menurut (Wanto, et al., 2020, p. 121) Data mining bertujuan untuk mengekstraksi informasi dari data dengan ukuran besar sehingga menghasilkan pola informasi yang menarik sebagai pengetahuan baru yang tersembunyi dari data tersebut. Data mining adalah proses mencari hubungan ringkasan antara entri-entri dalam matriks data yang entah sangat sering jarang terjadi data mining merupakan proses yang bertujuan untuk mencari hubungan hubungan antara data baik pola yang sangat sering atau jarang ( C. Aggarwal & K. Reddy, 2014, p. 14). Data mining merupakan tugas menentukan pola-pola yang menarik dan sejauh mana pola-pola tersebut digunakan dalam proses pengambilan keputusan pada perusahaan (Han, Kamber, & Pen, 2012, p. 601) . Tahapan proses data mining ada beberapa sesuai dengan proses KDD yaitu (Han, Kamber, & Pen, 2012, p. 601):

- (1) cleaning and integration;
- (2) selection and transformation;
- (3) proses mining;
- (4) evaluation and precentation.

CRISP-DM merupakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis ataupun unit penelitian (Kusrini, Emha Taufiq Luthfi, 2009, p.8). Di dalam CRISP-DM sebuah data mining memiliki siklus hidup yang terbagi menjadi enam fase pada gambar 2.1 yaitu( Kusrini, Emha Taufiq Luthfi, 2009, p.8-9):

- (1) fase pemahaman bisnis (*Business Understanding Phase*), yaitu berupa penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan;
- (2) fase pemahaman data (*Data Understanding*), yaitu berupa mengumpulkan data, mengenali lebih lanjut data, mengevaluasi kualitas data;
- (3) fase pengolahan data (*Data Preparation Phase*), memilih kasus dan variabel yang akan dianalisis yang sesuai dengan analisis yang akan dilakukan dan menyiapkan data awal untuk pemodelan ;
- (4) fase pemodelan (*Modeling Phase*), memilih teknik pemodelan dan mengaplikasikan pemodelan yang sesuai ;

- (5) fase evaluasi (*Evaluation Phase*), yaitu mengevaluasi satu atau lebih yang digunakan, menetapkan apakah model yang memenuhi tujuan;
- (6) fase penyebaran (*Deployment Phase*), menggunakan model yang dihasilkan, pembuatan laporan.



**Gambar 2.1 Alur CRISP-DM**

(Kusrini & Emha Taufiq Luthfi, 2009, p.9)

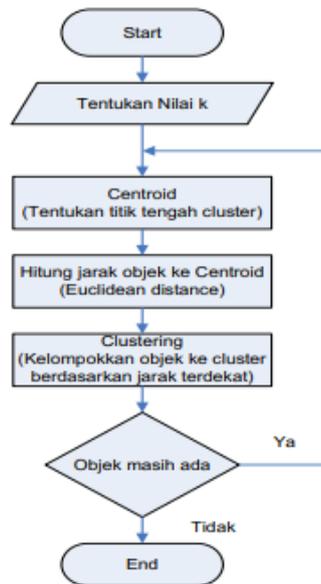
## 2. Pengertian Clustering

Clustering yaitu pengelompokan suatu set objek yang mirip kedalam kelas atau kelompok data dikelompokkan kedalam kelas atau kelas yang memiliki kemiripan tinggi satu sama lain tetapi memiliki perbedaan pada cluster lainnya (Parteek Bhatia, 2019 p. 155). Objek – objek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan intra kelas dan meminimalkan kesamaan di kelas lainnya ( C. Aggarwal & K. Reddy, 2014, p. 33). Analisis cluster dapat digunakan sebagai alat mandiri untuk mendapatkan wawasan tentang distribusi data, untuk mengamati karakteristik masing-masing cluster, dan untuk fokus pada serangkaian cluster tertentu untuk analisis lebih lanjut (Han, Kamber, & Pen, 2012, p. 443).

## 3. Algoritma K-means

Kmeans berusaha mengelompokkan data yang ada ke dalam beberapa kelompok setiap kelompok mempunyai karakteristik yang sama antara satu dengan yang lainnya, pemrosesan data mining diawali dengan pengelompokan centroid pertama yang dipilih secara acak sebagai titik awal

untuk setiap cluster kemudian dihitung ulang agar posisi centroid berulang (Wanto, et al., 2020, p. 4). Langkah-langkah algoritma K-Means dapat dijelaskan pada gambar 2.2 (Wanto, et al., 2020, p. 5):



**Gambar 2.2 Flowchart Algoritma K-Means**

(Wanto, et al., 2020, p. 5)

- (1) tentukan jumlah cluster (k) pada data set;
- (2) tentukan nilai pusat (Centroid); penentuan nilai Centroid pada tahap awal dilakukan secara random, sedangkan pada tahap iterasi digunakan rumus seperti pada persamaan berikut ini;

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

keterangan :

$V_{ij}$  = centroid rata-rata cluster ke-l untuk variabel ke-j;

$N_i$  = jumlah anggota cluster ke-i l;

$K$  = indeks dari cluster;

$j$  = indeks dari variable;

$X_{kj}$  = nilai data ke-k variabel ke-j untuk cluster tersebut;

- (3) pada masing-masing record, hitung jarak terdekat dengan Centroid jarak centroid yang digunakan adalah Euclidean Distance, dengan rumus pada persamaan dibawah ini:

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

keterangan :

De = euclidean distance;

i = banyaknya objek;

(x, y) = koordinat objek;

(s, t) = koordinat centroid;

- (4) kelompokkan objek berdasarkan jarak ke Centroid terdekat, ulangi langkah ke-3 hingga langkah ke-4, lakukan iterasi hingga centroid bernilai optimal. Berikut ini menurut (Wanto, et al., 2020, p. 5) merupakan contoh penerapan metode K-Means pada pengelompokan jumlah kunjungan wisatawan mancanegara. Berikut ini adalah proses perhitungan manual:

- (1) Menentukan data awal yang akan di cluster;

No	Kebangsaan	Tahun 2017	Tahun 2018
1	Afganistan	595	1034
2	Albania	592	706
3	Armenia	772	794
4	Austria	27208	26676
5	Azerbaijan	4405	1138
6	Bahrain	2457	2208
7	Bangladesh	56503	50423
8	Belarus	4576	4340
9	Belgium	48477	47693
10	Bhutan	610	594
11	...	...	...
91	Yunani (Greece)	9896	6985

- (2) Menentukan Jumlah Cluster ,menetapkan sebanyak 2 (dua) cluster.

- (3) Menentukan nilai Centroid;

Centroids	C1 MAX	2121888	2255115
	C2 MIN	140	294

- (4) Menghitung Jarak Centroid Untuk menghitung jarak antara titik Centroid dengan titik tiap objek menggunakan Euclidian Distance.

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Menghitung jarak Centroid 1 (C1) dan Centroid 2 (C2)

$$D_{A1C1} = \sqrt{(595 - 2121888)^2 + (1034 - 225115)^2} = 3095281,11$$

$$D_{A1C2} = \sqrt{(595 - 140)^2 + (1034 - 294)^2} = 868,69$$

$$D_{A2C1} = \sqrt{(592 - 2121888)^2 + (706 - 225115)^2} = 3095522,03$$

$$D_{A2C2} = \sqrt{(592 - 140)^2 + (706 - 294)^2} = 611,59$$

$$D_{A3C1} = \sqrt{(772 - 2121888)^2 + (794 - 225115)^2} = 3095334,60$$

$$D_{A3C2} = \sqrt{(772 - 140)^2 + (794 - 294)^2} = 805,87$$

$$D_{A4C1} = \sqrt{(27208 - 2121888)^2 + (26676 - 225115)^2} = 3058369,61$$

$$D_{A4C2} = \sqrt{(27208 - 140)^2 + (26676 - 294)^2} = 37797,97$$

$$D_{A5C1} = \sqrt{(4405 - 2121888)^2 + (1138 - 225115)^2} = 3092595,44$$

$$D_{A5C2} = \sqrt{(4405 - 140)^2 + (1138 - 294)^2} = 4347,71$$

$$D_{A6C1} = \sqrt{(2457 - 2121888)^2 + (2208 - 225115)^2} = 3093150,13$$

$$D_{A6C2} = \sqrt{(2457 - 140)^2 + (2208 - 294)^2} = 3005,31$$

$$D_{A7C1} = \sqrt{(56503 - 2121888)^2 + (50423 - 225115)^2} = 3021006,79$$

$$D_{A7C2} = \sqrt{(56503 - 140)^2 + (1034 - 294)^2} = 75430,13$$

$$D_{A8C1} = \sqrt{(4576 - 2121888)^2 + (4340 - 225115)^2} = 3090145,34$$

$$D_{A8C2} = \sqrt{(4576 - 140)^2 + (1034 - 294)^2} = 6004,02$$

$$D_{A9C1} = \sqrt{(4847 - 2121888)^2 + (47693 - 225115)^2} = 3028488,91$$

$$D_{A9C2} = \sqrt{(4847 - 140)^2 + (47693 - 294)^2} = 67698,82$$

$$D_{A10C1} = \sqrt{(610 - 2121888)^2 + (594 - 225115)^2} = 3095591,27$$

$$D_{A10C2} = \sqrt{(610 - 140)^2 + (594 - 294)^2} = 557,58$$

Berikut hasil dari perhitungan Centroid 1 dan Centroid 2 serta jarak terpendeknya.

No	Kebangsaan	C1	C2	Jarak Terpendek
1	Afganistan	3095281,11	868,69	868,69
2	Albania	3095522,03	611,59	611,59
3	Armenia	3095334,60	805,87	805,87
4	Austria	3058369,61	37797,97	37797,97
5	Azerbaijan	3092595,44	4347,71	4347,71
6	Bahrain	3093150,13	3005,31	3005,31
7	Bangladesh	3021006,79	75430,13	75430,13
8	Belarus	3090145,34	6004,02	6004,02
9	Belgium	3028488,91	67698,82	67698,82

No	Kebangsaan	C1	C2	Jarak Terpendek
10	Bhutan	3095591,27	557,58	557,58
11	...	...	...	
91	Yunani (Greece)	3084574,32	11830,00	11830,00

(5) Menentukan Posisi Cluster;

Kebangsaan	C1	C2
Afganistan		1
Albania		1
Armenia		1
Austria		1
Azerbaijan		1
Bahrain		1
Bangladesh		1
Belarus		1
Belgium		1
Bhutan		1
Bosnia		1
Brunei		1
Bulgaria		1
Canada		1
China	1	
Croasia		1
Cyprus		1
Czech		1
Denmark		1
Egypt		1
Estoria		1
Finland		1
France		1
Georgia		1
Germany		1

Kebangsaan	C1	C2
Iran		1
Irlandia		1
Israel		1
Japan		1
Jordan		1
Kamboja		1
Kazakhstan		1
Kirgistan		1
Korea Utara		1
Kuwait		1
Laos		1
Latvia		1
Libanon		1
Lithuania		1
Luxemburg		1
Macao		1
Maccodinia		1
Maladewa		1
Malaysia	1	
Mala		1
Moldova		1
Mongolia		1
Myanmar		1
Nepal		1
Netherlands		1

Kebangsaan	C1	C2
Hongkong		1
Hungaria		1
Italy		1
Islandia		1
India		1
Irak		1
Czech		1
Portugal		1
Qatar		1
Rusia		1
Romania		1
Saudi Arabia		1
Serbia		1
Singapore	1	
Slovakia		1
Slovenia		1
South Korea		1
Spain		1
Srilanka		1
Switzerland		1
Syria		1
Taiwan		1
Tajikistan		1

Kebangsaan	C1	C2
Norway		1
Orman		1
Pakistan		1
Palestina		1
Philipines		1
Polandia		1
Maladewa		1
Timor leste	1	
Turki		1
Turkmenistan		1
Ukraine		1
Uni Emirat Arab		1
United Kingdom		1
USA		1
Uzbekistan		1
Vietnam		1
Yaman		1
Yunani		1

Berikut perhitungannya :

$$C1_x = (2093171 + 2121888 + 1554119 + 960026) / 4$$

$$= 1682301$$

$$C1_y = (1994159 + 2255115 + 1526918 + 1610578) / 4$$

$$= 1846692$$

$$C2_x = (595 + 592 + 772 + 27208 + 4405 + 2457 + 56503 +$$

$$4576 + 48477 + 610 + 1784 + 23455 + 8695 +$$

$$96139 + 6620 + 1909 + 19904 + 43721 + 20345 +$$

$$\begin{aligned}
& 7569 + 24447 + 274117 + 1265 + 267823 + 98272 + \\
& 12600 + 90022 + 4776 + 536902 + 2167 + 16301 + \\
& 29400 + 688 + 573310 + 6773 + 6506 + 7219 + \\
& 1254 + 140 + 5760 + 4036 + 3932 + 6115 + 8550 + \\
& 1720 + 618 + 1381 + 2668 + 2173 + 1030 + 2414 + \\
& 48133 + 12821 + 210426 + 22838 + 18615 + 11424 + \\
& 2035 + 308977 + 32704 + 33223 + 1859 + 117532 + \\
& 18787 + 182086 + 7054 + 9264 + 5264 + 423191 + \\
& 81690 + 35669 + 51417 + 61191 + 2328 + 264278 + \\
& 350 + 138235 + 34433 + 233 + 32964 + 8387 + \\
& 378131 + 344766 + 4057 + 77466 + 8453 + 9896) / 87 \\
& = 61734,39
\end{aligned}$$

$$\begin{aligned}
C2_y &= (1034 + 706 + 794 + 26676 + 1138 + 2208 + 50423 \\
& + 4340 + 47693 + 594 + 982 + 14394 + 6908 \\
& + 89533 + 3723 + 1815 + 20250 + 44140 + 16297 \\
& + 7039 + 24230 + 275020 + 587 + 257233 + 84825 \\
& + 12488 + 89411 + 2232 + 535550 + 2189 + 11029 \\
& + 25972 + 460 + 485888 + 5726 + 7950 + 6671 \\
& + 1147 + 608 + 5199 + 3577 + 3873 + 4868 + 7790 \\
& + 1976 + 294 + 944 + 2204 + 1181 + 738 + 2992 \\
& + 25320 + 13630 + 196169 + 23155 + 24366 + 12107 \\
& + 2339 + 198182 + 29431 + 33700 + 1970 + 110028 \\
& + 12807 + 152569 + 3658 + 8494 + 4149 + 328471 \\
& + 81022 + 29709 + 44041 + 56384 + 1792 + \\
& \quad 191986 + 488 + 113565 + 19491 + 408 + 23468 \\
& \quad \quad \quad + 6540 + 362080 + 353030 + 3184 \\
& \quad \quad \quad + 70963 + 9036 + 6985) / 87 \\
& = 54715,59
\end{aligned}$$

Dari perhitungan di atas maka diperoleh hasil Centroid baru sebagai berikut:

**Tabel 2. 1 Nilai Centroid Iterasi ke 2**

Centroid	C1 (Max)	1682301	1846692,5
	C2(Min)	61734,39	54715,59

Untuk menghitung Iterasi ke-2, dapat dilakukan dengan cara yang sama menggunakan persamaan:

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Sehingga menghasilkan perhitungan seperti berikut:

No	Kebangsaan	C1	C2	Jarak Terpendek
1	Afganistan	24966916,17	81361,77	81361,77
2	Albania	2497160,65	81580,80	81580,80
3	Armenia	2496974,38	81387,66	81387,66
4	Austria	2460039,21	44477,97	44477,97
5	Azerbaijan	2494274,72	78467,93	78467,93
6	Bahrain	2494794,37	79188,73	79188,73
7	Bangladesh	2422767,68	6767,11	6767,11
8	Belarus	2491791,31	76189,12	76189,12
9	Belgium	2430181,08	15002,50	15002,50
10	Bhutan	2497231,33	81641,52	81641,52
11	.....	.....	.....	.....
91	Yunani	2486254,65	70465,79	70465,79

Berikut posisi cluster pada iterasi ke-2.

Kebangsaan	C1	C2	Kebangsaan	C1	C2
Afganistan		1	Iran		1
Albania		1	Irlandia		1
Armenia		1	Israel		1
Austria		1	Japan		1
Azerbaijan		1	Jordan		1
Bahrain		1	Kamboja		1
Bangladesh		1	Kazakhstan		1
Belarus		1	Kirgistan		1
Belgium		1	Korea Utara		1
Bhutan		1	Kuwait		1
Bosnia		1	Laos		1
Brunei		1	Latvia		1
Bulgaria		1	Libanon		1

Canada		1
China	1	
Croatia		1
Cyprus		1
Czech		1
Denmark		1
Egypt		1
Estonia		1
Finland		1
France		1
Georgia		1
Germany		1
Hongkong		1
Hungaria		1
Italy		1
Islandia		1
India		1
Irak		1
Portugal		1
Qatar		1
Rusia		1
Romania		1
Saudi Arabia		1
Serbia		1
Singapore	1	
Slovakia		1
Slovenia		1
South Korea		1
Spain		1
Srilanka		1
Sweden		1

Lithuania		1
Luxemburg		1
Macao		1
Macedonia		1
Maladewa		1
Malaysia	1	
Mala		1
Moldova		1
Mongolia		1
Myanmar		1
Nepal		1
Netherlands		1
Norway		1
Orman		1
Pakistan		1
Palestina		1
Philippines		1
Polandia		1
Timor leste	1	
Turki		1
Turkmenistan		1
Ukraine		1
Uni Emirat Arab		1
United Kingdom		1
USA		1
Uzbekistan		1
Vietnam		1
Yaman		1
Yunani		1

Switzerland		1
Syria		1
Taiwan		1
Tajikistan		1
Thailand		1

Diperoleh hasil dari posisi Cluster 2 adalah C1 berjumlah 4 data dan C2 berjumlah 87, maka proses perhitungan dihentikan.

#### 4. Pengertian dan Konsep Sistem Pendukung Keputusan

Sistem Pendukung Keputusan dirancang untuk menunjang seluruh tahapan pembuatan keputusan yang dimulai dari tahap mengidentifikasi masalah, memilih data yang relevan, menentukan pendekatan yang digunakan pada proses pembuatan keputusan, sampai pada kegiatan mengevaluasi pemilihan alternatif Sistem Pendukung Keputusan menyajikan informasi yang nantinya bisa dijadikan sebagai bahan alternatif pengambilan keputusan yang baik untuk membantu pengambil keputusan untuk memecahkan masalah semi dan atau tidak terstruktur (Yuswardi, 2022, p.3).

Tahapan dari sistem pendukung keputusan (Jeperson hutahaeen. Dkk, 2023, p.3):

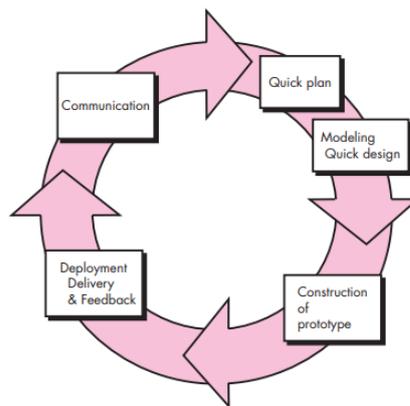
- (1) mendefinisikan masalah;
- (2) mengumpulkan data atau informasi yang relevan dan saling berkaitan;
- (3) pengolahan data dapat menjadi informasi dalam bentuk laporan tulisan atau grafik;
- (4) menentukan alternatif berupa solusi yang dapat berbentuk persentase.

Komponen Sistem Pendukung Keputusan (Jeperson hutahaeen. Dkk, 2023, p.3):

- (a) Subsistem Manajemen Data yaitu Subsistem manajemen data berupa basis data yang terdiri dari data yang relevan dengan keadaan.
- (b) Subsistem Manajemen Model yaitu Subsistem manajemen model berupa paket – paket software berisi yang berisi model – model finansial, statistik, ilmu manajemen, atau model kuantitatif yang menyediakan kemampuan analisa dan manajemen..
- (c) Subsistem User Interface yaitu Subsistem user interface merupakan subsistem yang dapat digunakan oleh user untuk berkomunikasi dengan sistem dan juga memberi perintah SPK.

#### 5. Prototyping pada SDLC

Menurut (Abdul Khadir, 2014, p. 344) bahwa SDLC (*System Development Life Cycle*) atau daur hidup pengembangan sistem merupakan metode klasik yang digunakan untuk mengembangkan, memelihara dan menggunakan sistem informasi. Prototyping merupakan suatu metode dalam pengembangan sistem yang menggunakan pendekatan untuk program dengan cepat dan bertahap sehingga dapat di evaluasi oleh pengguna, prototyping merupakan bagian dari SDLC (Abdul Khadir, 2014, p. 357). Pendekatan prototyping dapat dilihat pada gambar 2.3;



**Gambar 2.3 Metode Prototyping**

(Roger S. Pressman, 2010, p. 43)

Menurut Roger S. Pressman (2010, p.43-44) bahwa prototyping dimulai dengan *communication* yaitu komunikasi yaitu bertemu dengan pihak-pihak terkait untuk mendefinisikan tujuan keseluruhan dari perangkat lunak. Ini melibatkan diskusi dan pembicaraan untuk memahami apa yang ingin dicapai dengan pengembangan perangkat lunak tersebut; *Quick Plan* yaitu perencanaan secara cepat setelah komunikasi; selanjutnya *Modeling quick design* yaitu membuat model dan desain cepat ini fokus pada representasi aspek-aspek perangkat lunak yang akan terlihat atau dirasakan oleh pengguna akhir contohnya adalah tata letak antarmuka manusia (*human interface layout*) atau tampilan *output*; *Construction of prototype* hasil dari *quick design* adalah pembuatan prototipe; prototipe adalah versi awal dari perangkat lunak yang dirancang untuk memberikan gambaran nyata tentang bagaimana perangkat lunak tersebut akan berfungsi; *Deployment Delivery & Feedback* pada tahap ini prototipe di terapkan oleh pengguna akhir dan evaluasi oleh pengguna; feedback dari pengguna akan di gunakan untuk memperbaiki dan menyempurnakan prototipe.

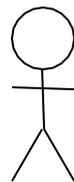
## 6. Pemrograman

Menurut Bunafit Nugroho (2004, p.201) Bahasa pemrograman merupakan sebuah paket bahasa yang digunakan untuk membentuk sebuah bahasa turunan, bahasa turunan ini dapat berupa bahasa pemrograman, atau dapat juga berupa hasil akhir yang sering disebut aplikasi pemrograman. Python adalah bahasa pemrograman yang dinamis yang berbasis objek, python berorientasi objek (*Object-Oriented Programming*), python memiliki modul-modul yang siap pakai (Faisal Dharma Ardhinata & Gita Fadila Fitriana, 2022, p.15).

### 7. **Unified Modeling Language (UML)**

Dalam perancangan sistem, dibutuhkan pemodelan yang dapat menggambarkan sistem (Henderi, 2007, p. 4) menyatakan bahwa Unified Modeling Language (UML) adalah bahasa pemodelan yang telah jadi standar industri perangkat lunak untuk memvisualisasikan, merancang, dan mendokumentasikan sistem perangkat lunak; sehingga UML biasa digunakan untuk menggambarkan batasan sistem, proses bisnis, struktur statistik suatu sistem, sifat suatu sistem, arsitektur implementasi dan memperluas functionality. Menurut (Rosa, 2018, p. 155) UML memiliki beberapa diagram yaitu:

(a) Use Case Diagram yaitu pemodelan untuk kelakuan (behavior) sistem informasi yang akan dibuat. Berikut simbol usecase diagram :



*ACTOR*

Orang yang berinteraksi dengan sistem informasi;



*USECASE*

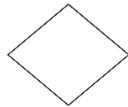
Usecase menggambarkan fungsionalitas yang disediakan sistem, biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama usecase.

_____	<b>ASOSIASI/ ASSOCIATION</b>	Asosiasi digunakan untuk menghubungkan antara actor dan usecase yang berpartisipasi pada usecase atau usecase memiliki interaksi dengan actor.
<<extend>> ----->	<b>EKSTENSI/ EXTEND</b>	<i>Extend</i> yaitu relasi usecase tambahan ke sebuah usecase dimana usecase yang ditambahkan dapat berdiri sendiri walau tanpa usecase tambahan.
—————>	<i>GENERALISASI /</i> <b>GENERALIZATION</b>	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah usecase dimana fungsi yang satu adalah fungsi yang lebih <b>umum dari lainnya.</b>
<<include>> ----->	<i>MENGGUNAKAN</i> <b>INCLUDE</b>	relasi usecase tambahan ke sebuah usecase untuk menjalankan fungsional yang prosesnya diharuskan atau yang menjadi saray terpenuhi;

(b) Class diagram merupakan alur database dalam sebuah progam. Dimana pada setiap laporan sistem yang hendak dibuat. Berikut simbol class diagram:

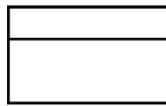
_____	Generalization	Hubungan dimana objek anak (descendent)
-------	----------------	---

berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (ancestor)



Navy Association

Upaya untuk menghindari asosiasi dengan lebih dari 2 objek



Class

Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama



Collaboration

Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor.



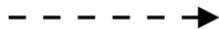
Realization

Operasi yang benar-benar dilakukan oleh suatu objek



Association

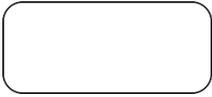
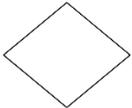
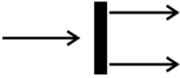
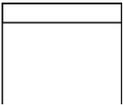
Yang menghubungkan antara objek satu dengan objek lainnya.



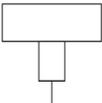
Dependency

Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempengaruhi elemen yang bergantung pada elemen yang tidak mandiri.

(c) Gambaran umum serta fungsi activity diagram adalah suatu dasar-dasar dalam pemrograman atau diagram yang menggambarkan workflow aktivitas dari sebuah sistem atau proses bisnis. Berikut simbol-simbol activity diagram:

	Status Awal/Initial	Status awal aktivitas sistem
	Activity	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	Decision	Asosiasi percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu.
	Join	Asosiasi penggabungan dimana lebih dari satu aktivitas lebih dari satu.
	Status Akhir/Final	Status akhir yang dilakukan sistem,.
	Swimlane	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi.

(d) Sequence diagram, yaitu diagram interaksi yang merinci bagaimana sebuah operasi dilakukan. Berikut simbol-simbol pada sequence diagram:

	Actor	Mempresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.
	Lifeline	Menghubungkan objek selama sequence (message dikirim atau diterima).
	General	Merepresentasikan entitas tunggal dalam sequence

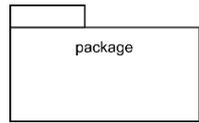
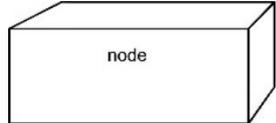
	Boundary	Berupa tepi dari sistem, seperti user interface dan alat yang berinteraksi dengan yang lain.
	Control	Elemen mengatur aliran dari informasi untuk sebuah scenario
	Entity	Elemen yang bertanggung jawab menyimpan informasi
	Message Entry	Berfungsi untuk menggambarkan pesan/ hubungan antar objek.
	Message to Self	Simbol ini menggambarkan pesan/hubungan objek itu sendiri, yang menunjukkan urutan kejadian yang terjadi.
	Message Return	Menggambarkan hasil dari pengiriman message yang digambarkan dengan arah dari kanan ke kiri.

(e) Component diagram yaitu diagram yang menggambarkan kumpulan komponen dan hubungan antara komponen. Berikut ini merupakan simbol-simbol yang dimiliki oleh component diagram:

	Package	<i>Package</i> merupakan sebuah bungkus dari satu atau lebih komponen
	Component	Komponen sistem

	Dependency	Kebergantungan antar komponen, arah panah mengarah pada komponen yang dipakai.
	Link	Relasi antar komponen
	Interface	Sebagai antarmuka komponen agar tidak mengakses langsung komponen.

(f) Deployment diagram yaitu menunjukkan konfigurasi komponen dalam proses eksekusi aplikasi. Berikut simbool-simbol yang digunakan dalam deployment diagram:

	Package	Package merupakan sebuah bungkus dari satu atau lebih node
	Dependency	Kebergantungan antar <i>node</i> , arah panah mengarah pada <i>node</i> yang dipakai.
	Link	Relasi antar <i>node</i> .
	Node	Node biasanya mengacu pada perangkat keras ( <i>hardware</i> ), perangkat lunak yang tidak dibuat sendiri ( <i>software</i> ),

## B. Tinjauan Studi

Dalam pelaksanaan penelitian, dilakukan proses penggalian informasi berdasarkan penelitian yang telah dilakukan sebelumnya. Penelitian mengenai klusterisasi objek wisata untuk rekomendasi prioritas pengembangan pariwisata yang pernah dilakukan sebelumnya menggunakan teori Metode K-Means. Penelitian tersebut berkontribusi sebagai bahan perbandingan dan sumber

pengetahuan penelitian yang dilaksanakan. Berikut adalah penelitian rujukan pada penelitian ini::

- (1) **Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Diprov. Dki Jakarta Dengan K-Means** (Maulida, 2018)

Penelitian ini menghitung data jumlah kunjungan asing di Prov. DKI Jakarta dengan metode K-Means terdapat 8 objek wisata unggulan diperoleh 3 klaster dengan klaster C1 terdapat 1 objek wisata , Klaster C2 terdapat 2 objek wisata dan klaster C3 terdapat 5 objek wisata. Dimana Hasil hasil klaster C3 yang menjadi catatan bagi pemerintah Prov. DKI. Jakarta.

- (2) **Pengelompokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering** (Naning et al., 2022)

Hasil analisis menunjukkan terdapat tiga cluster objek wisata yaitu C1: jumlah pengunjung rendah sebanyak 20 objek wisata, C2: jumlah pengunjung sedang sebanyak 15 objek wisata, dan C3: jumlah pengunjung tinggi sebanyak 2 objek wisata.

- (3) **Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran** (Al-Fahmi et al., 2023)

menghasilkan 3 cluster dengan rincian Cluster 0 terdiri dari Taman Pinggir Gawan, Negeri Atas Angin, Mumbul Kalianyar, Agrowisata Salak Wedi, dan Wisata Edukasi Pejambon. Cluster 1 terdiri dari Agrowisata Belimbing, Agrowisata Jambu Padang, Wali Kidangan, Teksas Wonocolo, Rumah Singgah, Waduk Bendo, Penangkaran Rusa Malo, Growgoland, Rumah Tua Padangan, Puthuk Kreweng, Banyu Kuning, Masyarakat Samin, Petilasan Angling Dharma, dan Sumber Maha Dewi. Cluster 2 terdiri dari Dander Waterpark, Kayangan Api, Waduk Pacal, Wisata Edukasi Gerabah, dan Outbond Toyoaji. Maka dari itu, dengan adanya pengelompokan tempat wisata sesuai dengan kondisi data yang ada dapat menentukan strategi pemasaran perlu dilaksanakan lebih efisien dan efektif.

- (4) **Implementasi Algoritma K-Means Dalam Pengelompokan Jumlah Wisatawan Akomodasi Di Jawa Barat** (Lusianah et al., 2023)

dikelompokkan menjadi 2 cluster, kemudian setiap cluster dibuat berdasarkan daerah dan tahun dengan jumlah wisatawan yang ramai berkunjung, hasil dari pengelompokan jumlah tertinggi yaitu Kota Bandung dengan jumlah wisatawan daerah sebanyak 8411064 pengunjung di tahun 2019. Sedangkan untuk Analisis Jenis Wisatawan Yang Berkunjung. Dari hasil analisis diperoleh

dengan jumlah tertinggi yaitu wisatawan Nusantara atau (dalam negeri) sebanyak 20445162 pengunjung di tahun 2016.

(5) **Penggunaan Algoritma K-Means Pada Aplikasi Pemetaan Kluster Daerah Pariwisata** (Marini & Suhendra, 2023)

Dari hasil analisis kluster diperoleh kluster 1 dengan jumlah kunjungan sedikit yang terdiri dari distrik Kuri Wamesa, Raisey, Roswar, Wasior, Wondiboy sedangkan kluster 2 dengan jumlah kunjungan terbanyak terdiri dari distrik Teluk Duairi, Roon, Windesi. Untuk penelitian selanjutnya adalah dengan menambah jumlah distrik yang tersebar di wilayah Kawasan Teluk Cendrawasih yaitu distrik- distrik yang masuk pada Kabupaten Nabire Provinsi Papua Tengah, selain itu penelitian selanjutnya juga dapat dilakukan dengan menghitung peringkat prioritas utama distrik yang akan dioptimalkan daerah pariwisatanya menggunakan sistem pengambilan keputusan. melalui penelitian ini juga diperoleh banyak daerah dari Kabupaten Teluk Wondama yang tidak memiliki kunjungan yang disebabkan banyak wilayah yang potensi wisatanya belum terdeteksi dan diperhatikan dengan maksimal oleh pemerintah daerah.

(6) **Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi** (Sari & Hartama, 2018)

Diperoleh pengelompokan Jumlah Tamu Asing berdasarkan provinsi dengan 1 provinsi dalam cluster C1, 2 provinsi cluster C2, dan 31 provinsi cluster C3. Hasil yang didapat dari penelitian bahwa hampir 90% provinsi di Indonesia masih memiliki potensi wisata yang rendah. Itu dibuktikan dari jumlah tamu asing yang menginap di hotel berdasarkan laporan dari badan pusat statistik. Hal ini menjadi masukan bagi pemerintah, untuk meningkatkan pertimbangan pemerintah khususnya Dinas Pariwisata dan Ekonomi Kreatif di Provinsi-provinsi untuk mengembangkan potensi objek daya tarik wisata di provinsi-provinsi serta sebagai bahan pertimbangan pengembangan hotel-hotel berbintang yang ada di Indonesia dan bertujuan untuk meningkatkan devisa negara.

(7) **Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means** (Setio & Prasetyaningrum, 2021)

Hasil dari penelitian ini yaitu melalui metode algoritma K-Means bahwa telah melakukan iterasi sebanyak 5 kali dengan memilih cluster 1, cluster 2, cluster 3 secara acak (random) dengan cluster 1 memiliki 24 data dengan persentase

sebesar (50%), cluster 2 memiliki 11 data dengan persentase sebesar (23%), dan cluster 3 memiliki 13 data dengan persentase sebesar (27%).

(8) **Penerapan Metode Clustering Untuk Pengelompokan Potensi Wisata Di Kabupaten Sumedang** (No, 2018)

Menghasilkan 3 kluster dengan C1 terdapat 12 objek wisata, C2 terdapat 18 objek wisata, terdapat 7 objek wisata.

(9) **Data Mining Using K-Means Clustering Algorithm for Grouping Countries of Origin of Foreign Tourist** (Hasanah et al., 2021)

Menghasilkan 2 kluster, yaitu kluster 1 untuk kategori pengunjung rendah dengan 206 anggota dan kluster 2 untuk kategori pengunjung tinggi dengan 6 anggota, yaitu Malaysia, Singapura, China, Asia Lainnya, Timor Leste, dan Australia.

(10) **Clustering Tourist Destinations Based on Number of Visitors Using the K-Mean Method** (Rochman & Rachmad, 2020)

Menghasilkan 3 kluster untuk wisatawan Nusantara, yaitu C1 terdiri dari 2 objek wisata unggulan, C2 terdiri dari 16 objek wisata unggulan, dan C3 terdiri dari 2 objek wisata yang menonjol. Sementara itu, berdasarkan Wisatawan Asing, C1 terdiri dari 16 objek wisata unggulan, C2 terdiri dari 2 objek wisata unggulan, dan C3 terdiri dari 1 objek wisata unggulan. Berikut Tabel 2.2 merupakan tabel tinjauan studi:

**Tabel 2. 2 Tinjauan Studi**

No	Peneliti	Judul	Jurnal	Kontribusi
1	Linda Maulida (2018)	<b>Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Diprov. Dki Jakarta Dengan K-Means</b>	JISKa (Jurnal Informatika Sunan Kalijaga) Vol.2 No.3 ISSN 2527-5836 Diakses pada 2/11/2023. <a href="https://ejournal.uin-suka.ac.id/saintek/JI-SKA/article/view/23-06/1079">https://ejournal.uin-suka.ac.id/saintek/JI-SKA/article/view/23-06/1079</a>	Kontribusi dalam penelitian ini adalah:  Terdapat kesamaan atribut yang digunakan.
2	Nanning Savitri, Renhat Pranata, Aprilius Nadzario Mulya Clara (2021)	<b>Pengelompokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering</b>	Jurnal UMJ Diakses pada 2/11/2023. <a href="https://jurnal.umj.ac.id/index.php/just-it/index">https://jurnal.umj.ac.id/index.php/just-it/index</a>	Kontribusi dalam Penelitian ini adalah:  Terdapat kesamaan pada langkah – langkah dalam pengolahan data.

No	Peneliti	Judul	Jurnal	Kontribusi
3	Barron Mahardika Al-Fahmi, Endra Rachmawati, Tri Sagirani (2023)	<b>Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran</b>	Jurnal Nasional Teknologi dan Sistem Informasi Diakses pada 27/09/2023 <a href="http://teknosi.fti.unand.ac.id">http://teknosi.fti.unand.ac.id</a>	Kontribusi dalam penelitian ini adalah: Pengambilan metode data mining yang digunakan yaitu menggunakan metode <i>CRISP-DM</i> .
4	Neni Lusianah, Ade Irma Purnamasari, Bani Nurhakim (2023)	<b>Implementasi Algoritma K-Means Dalam Pengelompokan Jumlah Wisatawan Akomodasi Di Jawa Barat</b>	EBISMEN (Jurnal Ekonomi Bisnis dan Manajemen) Vol.2, No.1 p-ISSN:2962-763X Diakses pada 28/12/2023 <a href="https://journal.unima.ac.id/index.php/EBISMEN/article/view/682">https://journal.unima.ac.id/index.php/EBISMEN/article/view/682</a>	Kontribusi dalam penelitian ini adalah: Pengambilan proses tahapan dalam data mining yaitu menggunakan proses <i>KDD (Knowledge Discovery in Database)</i>
5	Lion Ferdinand Marini, Christian Dwi Suhendra (2023)	<b>Penggunaan Algoritma K-Means Pada Aplikasi Pemetaan Klaster Daerah Pariwisata</b>	Jurnal Media Informatika Budidarma Vol. 7 No.2 ISSN 2614-5278. Diakses pada 14/11/2023 <a href="https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/download/5558/3456">https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/download/5558/3456</a>	Kontribusi dari penelitian ini adalah: Pengambilan teknik pengujian yang digunakan yaitu menggunakan metode silhouette coefficient.
6	Riyani Wulan Sari, Dedy Hartama (2018)	<b>Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi</b>	SENSASI (Seminar Nasional Sains & Teknologi Informasi) <a href="http://prosiding.seminda.com/index.php/sensasi/article/view/47">http://prosiding.seminda.com/index.php/sensasi/article/view/47</a>	Kontribusi dalam penelitian ini adalah: Pengambilan teknik perhitungan jarak yang digunakan.

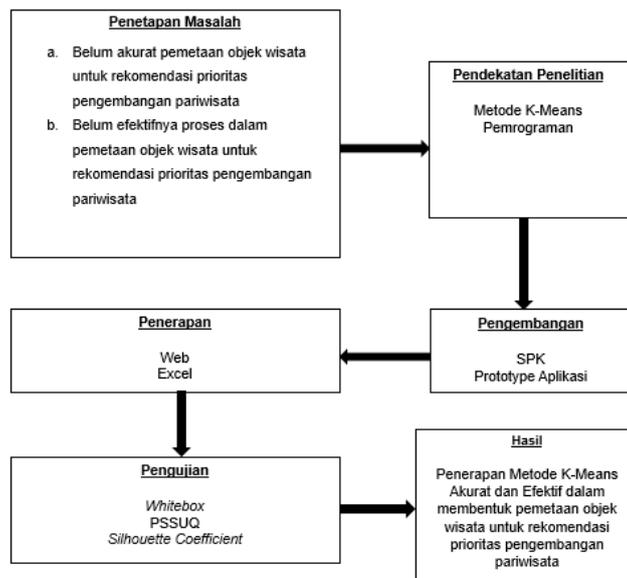
No	Peneliti	Judul	Jurnal	Kontribusi
7	Bambang Setio Purnomo, Putri Taqwa Prasetyaningrum (2021)	<b>Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means</b>	Journal Of Computer Science and Technology, Vol.1 No.1 Diakses pada 5/11/2023 <a href="https://journal.unwidha.ac.id/index.php/jcstec/article/view/38">https://journal.unwidha.ac.id/index.php/jcstec/article/view/38</a>	Kontribusi dari penelitian ini adalah:  Pengambilan teknik pengumpulan data yang digunakan.
8	M Egypt Pratama, Alif Finandhita (2019)	<b>Penerapan Metode Clustering Untuk Pengelompokan Potensi Wisata Di Kabupaten Sumedang</b>	Diakses pada 2/11/2023 <a href="https://onesearch.id/Record/IOS7445.1491">https://onesearch.id/Record/IOS7445.1491</a>	Kontribusi dalam penelitian ini adalah:  Pengambilan metode data mining yang digunakan yaitu menggunakan metode <i>CRISP-DM</i> .
9	Herliyani Hasanah, Nugroho Arif Sudibyo, Rhezka Mahendra Galih (2021)	<b>Data Mining Using K-Means Clustering Algorithm for Grouping Countries of Origin of Foreign Tourist</b>	Basic and Applied Science Conference (BASC) Diakses pada 1/01/2024 <a href="http://dx.doi.org/10.11594/nstp.2021.1112">http://dx.doi.org/10.11594/nstp.2021.1112</a>	Kontribusi dalam penelitian ini adalah:  Pengambilan teknik perhitungan jarak yang digunakan yaitu menggunakan Euclidean Distance).
10	Eka Mala Sari Rochman, Aeri Rachmad (2019)	<b>Clustering Tourist Destinations Based on Number of Visitors Using the K-Mean Method</b>	1 <sup>st</sup> International Multidisciplinary Conference on Education, Technology, and Engineering (IMCETE 2019) Diakses pada 1/01/2024 <a href="https://www.atlantispress.com/proceedings/imcete-19/125935521">https://www.atlantispress.com/proceedings/imcete-19/125935521</a>	Kontribusi dalam penelitian ini adalah:  Ada kesamaan atribut yang digunakan.

Berdasarkan tinjauan pustaka yang sudah dipaparkan pada tabel 2.8 didapatkan pengetahuan yang dijadikan rujukan dalam pelaksanaan penelitian ini rujukan penelitian berkontribusi dalam memberi pengetahuan mengenai permasalahan terkait pemetaan objek wisata untuk rekomendasi prioritas pengembangan pariwisata. Penelitian yang dilakukan merupakan pengembangan dari penelitian yang ada sebelumnya. Pada penelitian ini, dilakukan pengembangan penelitian dengan menggunakan variabel nama objek wisata, jumlah kunjungan, fasilitas, Aksesibilitas kendaraan umum, harga tiket masuk (HTM) dan luas.

### C. Kerangka Berpikir

Kerangka berpikir adalah skema untuk merangkai konsep pada penelitian.

Gambar 2. 4 merupakan kerangka penelitian pada penelitian ini:



**Gambar 2. 4 Kerangka Berpikir**

Deskripsi kerangka berpikir pada gambar 2.4:

- (1) Permasalahan pada penelitian yaitu belum akurat pemetaan objek wisata untuk rekomendasi prioritas pengembangan pariwisata dan belum efektifnya proses pada pemetaan objek wisata untuk rekomendasi prioritas pengembangan pariwisata;
- (2) Untuk menemukan solusi dari permasalahan maka dilakukan pendekatan penelitian dengan metode K-Means dan pemrograman dengan bahasa python;

- (3) Pengembangan penelitian dilakukan dalam pembangunan sistem ini menggunakan SPK dan prototype aplikasi;
- (4) Penerapan penelitian dilakukan dengan menggunakan web dan data Excel;
- (5) Untuk menguji penelitian dan pengembangan yang dilakukan, maka dilakukan pengukuran dengan menggunakan *whitebox*, PSSUQ, dan *silhouette coefficient*;
- (6) Hasil dari penelitian dan pengembangan adalah terbukti bahwa metode K-Means akurat dan efektif dalam membentuk pemetaan objek wisata untuk rekomendasi prioritas pengembangan pariwisata.

#### **D. Hipotesis**

Algoritma K-Means merupakan metode data mining yang digunakan untuk klusterisasi dengan cara mengelompokkan data yang memiliki karakteristik yang sama ke dalam satu cluster atau lebih. Merujuk pada ketercapaian penelitian sebelumnya, metode K-Means mampu membentuk kluster yang akurat. Berdasarkan permasalahan yang dihadapi yaitu belum akurat dan efektif proses pemetaan objek wisata rekomendasi prioritas pengembangan pariwisata, maka perlu adanya klusterisasi objek wisata rekomendasi prioritas pengembangan pariwisata dan sifat K-Means untuk melakukan pengelompokan maka dikatakan dengan permasalahan dalam penelitian ini menunjukkan pemecahan masalah untuk mengetahui objek wisata prioritas pengembangan pariwisata yang akurat dan efektif dengan metode K-Means. Dengan pemahaman diatas maka hipotesis penelitian dan pengembangan ini dapat diungkapkan bahwa “penerapan metode K-Means diduga akurat dan efektif untuk menentukan objek wisata rekomendasi prioritas pengembangan pariwisata”.