BAB II
KERANGKA TEORITIS

A. Landasan Teori

1. *Data mining*

Syahdan dan Sindar mengatakan bahwa *Data mining* merupakan iteratif dan interaktif untuk menemukan pola-pola atau model baru yang shahih (sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang besar (*massive database), data mining* berisi pencarian trend atau pola yang diinginkan dalam basisdata besar untuk membantu pengambilan keputusan diwaktu yang akan datang (Wahyudi, Masitha, & rekan, 2020, p. 1). Proses data mining melibatkan penggunaan berbagai teknik, statistik, matematis, dan kecerdasan buatan untuk menganalisis data dengan cara yang sistematis dan otomatis, hasil dari data mining dapat digunakan untuk mendukung pengambilan keputusan, mengidentifikasi tren pasar, meningkatkan efisiensi operasional, atau merumuskan strategi bisnis (Rahayu, et al., 2024, p. 2).

Secara sistematis, langkah utama untuk melakukan *Data mining* terdiri dari tiga tahap, yaitu sebagai berikut menurut Gonunescu dalam (Muslim, Prasetiyo, & rekan, 2019, p. 2));

(1) Eksplorasi atau Pemrosesan awal data, terdiri dari pembersihan data, normalisasi data, transformasi data, penanganan missing value, reduksi dimensi, pemilihan subset fitur, dan sebagainya;
(2) Membangun model dan validasi, melakukan analisis dari berbagai model dan memilih model sehingga menghasilkan kinerja yang terbaik. Pembangunan model dilakukan menggunakan metode-metode seperti klasifikasi, regresi, analisis *cluster,* dan asosiasi;
(3) Penerapan, dilakukan dengan menerapkan model yang dipilih pada data yang baru untuk menghasilkan kinerja yang baik pada masalah yang diinvestigasi.

2. Clustering

Clustering merupakan teknik dengan cara mengelompokkan data secara otomatis tanpa diberitahukan label kelasnya, clustering dapat digunakan untuk memberikan label pada kelas data yang belum diketahui, karena clustering sering digolongkan sebagai metode unsupervised learning, metode clustering yang sering digunakan yaitu, K-Medoids, K-Means, Fuzzy C-Means, Self-Organizing Map (SOM), dll (Muslim, Prasetiyo, & rekan, 2019, p. 45). Clustering adalah salah satu metode untuk mengelompokkan data, yang sering dimasukkan sebagai salah satu metode dalam Data Mining, yang tujuannya adalah untuk mengelompokkan

data dengan karakteristik yang sama ke suatu ‘wilayah atau kelompok’ yang sama dan data dengan karakteristik yang berbeda ke ‘wilayah atau kelompok’ yang lain, beberapa pendekatan yang digunakan dalam mengembangkan metode clustering contohnya adalah pendekatan partisi dan hirarki. (Mustika, Ardilla, & rekan, 2021, p. 160). Salah satu contoh yaitu *clustering* K-Means yang merupakan suatu metode penganalisaan data atau metode *data mining* yang melakukan pemodelan tanpa supervisi, selain itu merupakan salah satu metode yang mengelompokkan data dengan sistem partisi (Mustika, Ardilla, & rekan, 2021, p. 140)

3. Algoritma K-Means

K-Means adalah metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik. Algoritma K-Means termasuk *partitioning clustering* yang memisahkan data ke daerah bagian yang terpisah, dalam algoritma K-Means, setiap data harus termasuk ke kluster tertentu dan bisa dimungkinkan bagi setiap data yang termasuk kluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke kluster lainnya, pada dasarnya penggunaan algoritma dalam melakukan proses *clustering* tergantung dari data yang ada dan kongklusi yang ingin dicapai (Wahyudi, Masitha, & rekan, 2020, p. 5). Wu dan Kumar mengatakan bahwa Algoritma K-Means merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K cluster yang sudah ditetapkan di awal, algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek (Prasetyo, 2014). Algoritma klastering K-Means dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetanpkan, algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama, jarak itu sendiri dihitung menggunakan fungsi euclidean, kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat (Wahyudi, Masitha, & rekan, 2020, pp. 6, 7).

Langkah-langkah pengelompokan data adalah :

(1) Pilih jumlah klaster;
(2) Inisialisasi awal dan pusat klaster dilakukan secara random;
(3) Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar obyek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan Metode Jarak Euclidean (Euclidean Distance) dengan rumus seperti dibawah ini (Witantom, Ratnawati dkk dalam (Wahyudi, Masitha, & rekan, 2020):

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \text{.....(1)}$$

dimana :

$d(x, y)$ = ukuran ketidakmiripan

$x_i = (x_1, x_2, \dots, x_i)$ yaitu variabel data

$y_i = (y_1, y_2, \dots, y_j)$ yaitu variabel pada titik pusat

(4) Hitung pusat klaster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata obyek pada klaster. Penghitungan bisa juga dengan menggunakan median;

(5) Hitung kembali jarak tiap objek dengan pusat klaster yang baru, hingga klaster tidak berubah, maka proses pengklasteran selesai.

Data persentasi wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen) berdasarkan provinsi yang diambil dari Badan Pusat Statistik (BPS) tingkat nasional yang dihitung dengan menggunakan 2 nilai kluster, yaitu : tinggi dan rendah. Data berikut selanjutnya akan diolah menggunakan perangkat lunak. Berikut uraian perhitungan manual proses clustering wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen) dengan menggunakan sebuah Algoritma K- Means. Berikut adalah langkah-langkah penyelesaian dalam mengelompokkan wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen) berdasarkan provinsi dengan menggunakan Algoritma K-Means (Wahyudi, Masitha, & rekan, 2020):

Tabel 2.1 Persentase wanita umur 15 s/d 49 tahun yang menggunakan alat kontrasepsi untuk KB

No	Provinsi	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1	Aceh	0	43,04	42,8	42,4	49,08	49,55	49,87	52,53	52,69	52,09	46,92
2	Sumatera Utara	42,51	45,08	45,53	41,91	49,71	48,67	50,25	52,86	51,79	51,87	49,06
3	Sumatera Barat	47,59	49,06	48,37	47,32	50,57	53,13	53,47	51,96	51,71	53,2	48,53
4	Riau	49,8	53,69	54,17	52,41	56,53	56,29	56,74	57,39	58,43	56,29	54,42
5	Jambi	62,94	61,63	64,66	62,16	66,72	65,8	67,32	68,05	68,5	67,5	64,16
6	Sumatera Selatan	59,42	62,44	61,97	62,92	64,63	65,78	66,37	67,23	67,98	66,47	68,06
7	Bengkulu	66,39	70,08	67,3	67,62	68,46	68,98	70,47	70,34	71,42	70,61	67,83
8	Lampung	65,97	64,49	64,03	64,58	67,81	69,28	66,41	67,74	69,55	69,36	67,35
9	Kep, Bangka Belitung	63,72	63,44	63,57	64,3	66,16	68,17	65,05	67,21	69,05	67,06	64,99
10	Kep, Riau	49,51	55,41	51,2	53,07	55,54	51,9	50,91	52,22	50,21	47,19	47,05

No	Provinsi	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
11	Dki Jakarta	54,13	55,25	54,69	52,68	56,62	57,42	55,19	57	57,55	55,14	54,75
12	Jawa Barat	62,88	62,84	62,28	60,51	63,67	64,57	64,09	65,53	65,12	65,36	64,67
13	Jawa Tengah	61,32	62,1	60,65	59,19	63,67	63,85	63,7	64,5	64,54	63,88	62,15
14	Di Yogyakarta	62,15	61,13	56,11	57,42	62,21	61,93	60,85	59,89	63,04	61,41	59,33
15	Jawa Timur	59,72	59,52	59,65	59,54	63,72	64,16	64,53	65,38	66,11	65,33	63,79
16	Banten	58,61	60,33	56,64	58	60,51	62,18	62,38	62,9	62,11	62,71	61,16
17	Bali	68,2	67,43	67,22	65,06	67,85	65,17	63,85	64,33	62,8	64,64	60,03
18	Nusa Tenggara Barat	55,71	54,82	52,44	53,07	57,88	57,75	59,68	58,67	60,34	58,79	59,07
19	Nusa Tenggara Timur	33,8	32,63	34,35	35,91	40,77	39,89	41,46	40,75	43,7	44,92	42,08
20	Kalimantan Barat	61,29	59,49	61,26	60,73	64,05	65,59	66,73	68,56	67,1	69,07	65,76
21	Kalimantan Tengah	67,08	66,64	67,46	68,4	70,34	68,16	70,85	72,49	72,88	72,07	68,5
22	Kalimantan Selatan	64,85	66,7	63,27	64,25	67,76	68,03	70,05	70,02	69,91	70,8	70,13
23	Kalimantan Timur	54,52	54,67	55,8	55,29	58,32	61,01	60,52	61,82	62,88	60,74	59,86
24	Sulawesi Utara	70,01	69,75	67,07	65,19	67,54	68,38	68,48	68,14	65,24	68,29	66,67
25	Sulawesi Tengah	54,97	54,68	56,83	55,91	61,5	61,08	58,25	60,8	59,7	60,38	57,55
26	Sulawesi Selatan	41,88	42,59	43,67	43,18	48,65	50,01	50,18	52,07	51,91	53,04	48,38
27	Sulawesi Tenggara	47,4	46,8	46,61	46,34	50,72	52,6	53,3	53	54,26	54,1	48,66
28	Gorontalo	59,91	61,24	64,22	59,54	62,83	64,22	61,6	65,08	65,13	66,83	64,78
29	Sulawesi Barat	0	38,82	38,47	45,23	49,78	48,83	47,84	50,92	47,93	49	47,69
30	Maluku	28,08	30,13	30,09	32,1	36,36	39,54	41,84	41	39,77	41,71	43,21
31	Maluku Utara	44,49	39,61	41,9	43,33	48,58	53,13	50,92	52,58	53,13	52,93	51,73
32	Papua Barat	0	31,73	28,29	26,69	36,47	38,48	37,84	41,25	42,91	42,12	43,96
33	Papua	32,8	31,22	31,92	27,71	33,71	26,97	23,91	24,77	23,87	27,88	23,37

1. Menentukan jumlah data yang akan dikluster, dimana sampel data Persentase Wanita Berumur 15-49 Tahun dan Berstatus Kawin yang Sedang Menggunakan/Memakai Alat KB (Persen) yang akan digunakan dalam proses clustering dengan jumlah data sebanyak 33 provinsi dan diambil dari tahun 2005 s/d 2015. Berikut yaitu cara mencari nilai rata-rata :

R1 = 0 + 43,04 + 42,8 + 42,2 + 49,08 + 49,55 + 49,87 + 52,53 + 52,69 + 52,09 + 46,92 / 11 = 43,72

R2 = 42,51 + 45,08 + 45,53 + 41,91 + 49,71 + 48,67 + 50,25 + 52,86 + 51,79 + 51,87 + 49,06 / 11 = 48,11

Lakukan pencarian rata-rata tersebut sampai R33, berikut nilai rata-rata yang telah didapat :

Tabel 2.2 Nilai Rata-rata

No	Provinsi	Rata-Rata
1	Aceh	43,72
2	Sumatera Utara	48,11
3	Sumatera Barat	50,45
4	Riau	55,11
5	Jambi	65,4
6	Sumatera Selatan	64,84
7	Bengkulu	69,05
8	Lampung	66,96
9	Kep, Bangka Belitung	65,7
10	Kep, Riau	51,29
11	Dki Jakarta	55,49
12	Jawa Barat	63,77
13	Jawa Tengah	62,69
14	Di Yogyakarta	60,5
15	Jawa Timur	62,86
16	Banten	60,68
17	Bali	65,14
18	Nusa Tenggara Barat	57,11
19	Nusa Tenggara Timur	39,11
20	Kalimantan Barat	64,51
21	Kalimantan Tengah	69,53
22	Kalimantan Selatan	67,8
23	Kalimantan Timur	58,68
24	Sulawesi Utara	67,71
25	Sulawesi Tengah	58,33
26	Sulawesi Selatan	47,78
27	Sulawesi Tenggara	50,34

No	Provinsi	Rata-Rata
28	Gorontalo	63,22
29	Sulawesi Barat	42,23
30	Maluku	36,71
31	Maluku Utara	48,39
32	Papua Barat	33,61
33	Papua	28,01

2. Menetapkan nilai K jumlah kluster sebanyak 2 kluster, kluster yang dibentuk yaitu: kluster tertinggi dan kluster rendah.
3. Menentukan nilai centroid awal yang telah ditentukan secara random. Kluster tertinggi (C1) diperoleh dari nilai maximum pada rata-rata data, kluster rendah (C2) diperoleh dari nilai minimum pada rata-rata data.

Berikut adalah nilai centroid data awal pada iterasi 1:

Tabel 2.3 Centroid awal iterasi 1

C1	Maximum	69,53
C2	Minimum	28,01

4. Setelah data nilai pusat kluster ditentukan, selanjutnya menghituhng jarak setiap data terhadap pusat kluster dengan menggunakan rumus sebagai berikut yang dilakukan dengan titik pusat (centroid) pada kluster pertama. Berikut adalah perhitungannya:

   $D\ (1.1) = \sqrt{(69{,}53 - 43{,}72)^2} = 25{,}81$

   $D\ (1.2) = \sqrt{(69{,}53 - 48{,}11)^2} = 21{,}42$

   Lakukan sampai dengan D (1.33)

   Perhitungan rata-rata data pada titik pusat centroid pada kluster 2:

   $D\ (2.1) = \sqrt{(28{,}01 - 43{,}72)^2} = 15{,}71$

   $D\ (2.2) = \sqrt{(28{,}01 - 48{,}11)^2} = 20{,}10$

   Lakukan sampai dengan D (2.33)

   Berikut adalah hasil dari perhitungan rata-rata data pada titik pusat centroid pada setiap kluster:

Tabel 2.4 Hasil Perhitungan Centroid pada setiap Kluster (Iterasi 1)

No	Provinsi	Rata-Rata	C1	C2
1	Aceh	43,72	25,81	15,71
2	Sumatera Utara	48,11	21,42	20,1
3	Sumatera Barat	50,45	19,09	22,43
4	Riau	55,11	14,43	27,09
5	Jambi	65,4	4,13	37,39

No	Provinsi	Rata-Rata	C1	C2
6	Sumatera Selatan	64,84	4,69	36,83
7	Bengkulu	69,05	0,49	41,03
8	Lampung	66,96	2,57	38,95
9	Kep, Bangka Belitung	65,7	3,83	37,69
10	Kep, Riau	51,29	18,24	23,28
11	Dki Jakarta	55,49	14,04	27,48
12	Jawa Barat	63,77	5,76	35,76
13	Jawa Tengah	62,69	6,85	34,67
14	Di Yogyakarta	60,5	9,04	32,49
15	Jawa Timur	62,86	6,67	34,85
16	Banten	60,68	8,85	32,67
17	Bali	65,14	4,39	37,13
18	Nusa Tenggara Barat	57,11	12,42	29,1
19	Nusa Tenggara Timur	39,11	30,42	11,1
20	Kalimantan Barat	64,51	5,02	36,5
21	Kalimantan Tengah	69,53	0	41,52
22	Kalimantan Selatan	67,8	1,74	39,79
23	Kalimantan Timur	58,68	10,86	30,66
24	Sulawesi Utara	67,71	1,83	39,69
25	Sulawesi Tengah	58,33	11,2	30,32
26	Sulawesi Selatan	47,78	21,76	19,77
27	Sulawesi Tenggara	50,34	19,19	22,33
28	Gorontalo	63,22	6,32	35,2
29	Sulawesi Barat	42,23	27,31	14,22
30	Maluku	36,71	32,82	8,7
31	Maluku Utara	48,39	21,14	20,38
32	Papua Barat	33,61	35,92	5,6
33	Papua	28,01	41,52	0

Lalu hitung jarak terdekat dengan menggunakan Euclidean Distance.

Tabel 2.5 Hasil Perhitungan Jarak Data pada Iterasi 1

No	Jarak Terpendek	Hasil	C1	C2
1	15,71	C2		1
2	20,1	C2		1
3	19,09	C1	1	
4	14,43	C1	1	
5	4,13	C1	1	
6	4,69	C1	1	
7	0,49	C1	1	
8	2,57	C1	1	
9	3,83	C1	1	
10	18,24	C1	1	
11	14,04	C1	1	

No	Jarak Terpendek	Hasil	C1	C2
12	5,76	C1	1	
13	6,85	C1	1	
14	9,04	C1	1	
15	6,67	C1	1	
16	8,85	C1	1	
17	4,39	C1	1	
18	12,42	C1	1	
19	11,1	C2		1
20	5,02	C1	1	
21	0	C1	1	
22	1,74	C1	1	
23	10,86	C1	1	
24	1,83	C1	1	
25	11,2	C1	1	
26	19,77	C2		1
27	19,19	C1	1	
28	6,32	C1	1	
29	14,22	C2		1
30	8,7	C2		1
31	20,38	C2		1
32	5,6	C2		1
33	0	C2		1

Berikut ini adalah hasil dari perhitungan jarak data ke titik pusat kluster pada iterasi 1:

Tabel 2.6 Hasil Kluster pada Iterasi 1

Kluster	Hasil
C1	24
C2	9

5. Selanjutnya lakukan kembali langkah 4 dan 5 jika nilai centroid hasil dari iterasi 1 dengan nilai centroid berikutnya bernilai sama ataupun nilai centroid sudah optimal serta posisi kluster tidak mengalami perubahan lagi maka proses iterasi berhenti. Namun jika posisi kluster masih berubah-ubah, maka proses iterasi terus berlanjut pada iterasi berikutnya sampai hasil cluster bernilai sama.
6. Menghitung titik pusat baru menggunakan hasil dari setiap anggota pada masing-masing cluster. Berikut contoh perhitungan titk pusat kluster baru pada iterasi ke - 2 (x):

$$C1x = \frac{\begin{matrix}50{,}54 + 55{,}11 + 65{,}40 + 64{,}84 + 69{,}05 + 66{,}96 + 65{,}70 + \\ 51{,}29 + 55{,}49 + 63{,}77 + 62{,}69 + 60{,}50 + 62{,}86 + 60{,}68 + 65{,}14 + \\ 57{,}11 + 64{,}51 + 69{,}53 + 67{,}80 + 58{,}68 + 67{,}71 + 58{,}33 + 50{,}34 + 63{,}22\end{matrix}}{24}$$

$$= 61{,}55$$

$$C2x = \frac{\begin{matrix}43{,}72 + 48{,}11 + 39{,}11 + 47{,}78 + 42{,}23 \\ 36{,}71 + 48{,}39 + 33{,}61 + 28{,}01\end{matrix}}{9}$$

$$= 40{,}85$$

Tabel 2.7 Nilai Centroid Baru

C1	Maximum	61,55
C2	Minimum	40,85

Selanjutnya lakukan sebuah perhitungan nilai centroid baru tersebut dengan nilai rata-rata data sebagai berikut:

Perhitungan rata-rata data pada titik centroid pada cluster 1:

$E\ (1.1) = \sqrt{(61{,}55 - 43{,}72)^2} = 17{,}82$

$E\ (1.2) = \sqrt{(61{,}55 - 48{,}11)^2} = 13{,}44$

Lakukan sampai dengan E (1.33)

Perhitungan rata-rata data pada titik centroid pada cluster 2:

$E\ (2.1) = \sqrt{(40{,}85 - 43{,}72)^2} = 2{,}78$

$E\ (2.2) = \sqrt{(40{,}85 - 48{,}11)^2} = 7{,}26$

Lakukan sampai dengan E (2.33)

Berikut merupakan hasil perhitungan rata-rata pada titik centroid setiap kluster:

Tabel 2.8 Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 2)

No	Provinsi	Rata-Rata	C1	C2
1	Aceh	43,72	17,82	2,87
2	Sumatera Utara	48,11	13,44	7,26
3	Sumatera Barat	50,45	11,1	9,59
4	Riau	55,11	6,44	14,25
5	Jambi	65,4	3,86	24,55
6	Sumatera Selatan	64,84	3,29	23,99
7	Bengkulu	69,05	7,5	28,19
8	Lampung	66,96	5,41	26,11
9	Kep, Bangka Belitung	65,7	4,15	24,85
10	Kep, Riau	51,29	10,26	10,44
11	Dki Jakarta	55,49	6,06	14,64
12	Jawa Barat	63,77	2,23	22,92
13	Jawa Tengah	62,69	1,14	21,83
14	Di Yogyakarta	60,5	1,05	19,64
15	Jawa Timur	62,86	1,31	22
16	Banten	60,68	0,86	19,83
17	Bali	65,14	3,6	24,29
18	Nusa Tenggara Barat	57,11	4,44	16,26
19	Nusa Tenggara Timur	39,11	22,43	1,74
20	Kalimantan Barat	64,51	2,96	23,66
21	Kalimantan Tengah	69,53	7,99	28,68

No	Provinsi	Rata-Rata	C1	C2
22	Kalimantan Selatan	67,8	6,25	26,94
23	Kalimantan Timur	58,68	2,87	17,82
24	Sulawesi Utara	67,71	6,16	26,85
25	Sulawesi Tengah	58,33	3,22	17,48
26	Sulawesi Selatan	47,78	13,77	6,92
27	Sulawesi Tenggara	50,34	11,2	9,49
28	Gorontalo	63,22	1,67	22,36
29	Sulawesi Barat	42,23	19,32	1,37
30	Maluku	36,71	24,84	4,14
31	Maluku Utara	48,39	13,15	7,54
32	Papua Barat	33,61	27,94	7,24
33	Papua	28,01	33,54	12,84

Lalu hitung jarak terdekat centroid menggunakan *Euclidean Distance*

Tabel 2.9 Hasil Perhitungan Jarak data pada Iterasi 2

No	Jarak Terpendek	Hasil	C1	C2
1	2,87	C2		1
2	7,26	C2		1
3	9,59	C2		1
4	6,44	C1	1	
5	3,86	C1	1	
6	3,29	C1	1	
7	7,5	C1	1	
8	5,41	C1	1	
9	4,15	C1	1	
10	10,26	C1	1	
11	6,06	C1	1	
12	2,23	C1	1	
13	1,14	C1	1	
14	1,05	C1	1	
15	1,31	C1	1	
16	0,86	C1	1	
17	3,6	C1	1	
18	4,44	C1	1	
19	1,74	C2		1
20	2,96	C1	1	
21	7,99	C1	1	
22	6,25	C1	1	
23	2,87	C1	1	
24	6,16	C1	1	
25	3,22	C1	1	
26	6,92	C2		1
27	9,49	C2		1

No	Jarak Terpendek	Hasil	C1	C2
28	1,67	C1	1	
29	1,37	C2		1
30	4,14	C2		1
31	7,54	C2		1
32	7,24	C2		1
33	12,84	C2		1

Berikut adalah hasil dari perhitungan jarak data ke titik pusat kluster pada iterasi 2:

Tabel 2.10 Hasil Kluster Iterasi 2

Kluster	Hasil
C1	22
C2	11

Selanjutnya menghitung titik pusat cluster baru menggunakan hasil dari setiap anggota data pada masing-masing cluster. Berikut perhitungan titik pusat cluster baru iterasi ke tiga (x):

$$C1x = \frac{\begin{array}{l}55{,}11 + 65{,}40 + 64{,}84 + 69{,}05 + 66{,}96 + 65{,}70 + 51{,}26 + \\ 55{,}49 + 63{,}77 + 62{,}69 + 60{,}50 + 62{,}86 + 60{,}68 + 65{,}14 + \\ 57{,}11 + 64{,}51 + 69{,}53 + 67{,}80 + 58{,}68 + 67{,}71 + 58{,}33 + 63{,}22\end{array}}{22}$$

$$= 62{,}56$$

$$C2x = \frac{\begin{array}{l}43{,}72 + 48{,}11 + 50{,}45 + 39{,}11 + 47{,}78 + \\ 50{,}34 + 42{,}23 + 33{,}71 + 48{,}39 + 33{,}61 + 28{,}01\end{array}}{11}$$

$$= 42{,}59$$

Tabel 2 11 Nilai Centroid Baru (Iterasi 3)

C1	Maximum	62,56
C2	Minimum	42,59

Perhitungan rata-rata data pada titik centroid pada kluster 1:

$F\ (1.1) = \sqrt{(62{,}56 - 43{,}72)^2} = 18{,}84$

$F\ (1.2) = \sqrt{(43{,}59 - 48{,}11)^2} = 14{,}45$

Lakukan sampai dengan F (1.33)

Perhitungan rata-rata data pada titik centroid pada cluster 2:

$F\ (1.1) = \sqrt{(62{,}56 - 43{,}72)^2} = 1{,}14$

$F\ (1.2) = \sqrt{(43{,}59 - 48{,}11)^2} = 5{,}52$

Lakukan sampai dengan F (2.33)

Berikut adalah hasil dari perhitungan rata-rata data pada titik pusat centroid setiap kluster:

Tabel 2 12 Hasil Perhitungan Cenroid pada setiap Kluster (Iterasi 3)

No	Provinsi	Rata-Rata	C1	C2
1	Aceh	43,72	18,84	1,14
2	Sumatera Utara	48,11	14,45	5,52
3	Sumatera Barat	50,45	12,12	7,86
4	Riau	55,11	7,46	12,52
5	Jambi	65,4	2,84	22,81
6	Sumatera Selatan	64,84	2,28	22,25
7	Bengkulu	69,05	6,48	26,46
8	Lampung	66,96	4,4	24,37
9	Kep, Bangka Belitung	65,7	3,14	23,11
10	Kep, Riau	51,29	11,27	8,7
11	Dki Jakarta	55,49	7,07	12,9
12	Jawa Barat	63,77	1,21	21,19
13	Jawa Tengah	62,69	0,12	20,1
14	Di Yogyakarta	60,5	2,07	17,91
15	Jawa Timur	62,86	0,3	20,27
16	Banten	60,68	1,88	18,1
17	Bali	65,14	2,58	22,55
18	Nusa Tenggara Barat	57,11	5,45	14,52
19	Nusa Tenggara Timur	39,11	23,45	3,47
20	Kalimantan Barat	64,51	1,95	21,92
21	Kalimantan Tengah	69,53	6,97	26,94
22	Kalimantan Selatan	67,8	5,23	25,21
23	Kalimantan Timur	58,68	3,89	16,09
24	Sulawesi Utara	67,71	5,14	25,12
25	Sulawesi Tengah	58,33	4,23	15,74
26	Sulawesi Selatan	47,78	14,78	5,19
27	Sulawesi Tenggara	50,34	12,22	7,76
28	Gorontalo	63,22	0,65	20,63
29	Sulawesi Barat	42,23	20,33	0,36
30	Maluku	36,71	25,85	5,88
31	Maluku Utara	48,39	14,17	5,8
32	Papua Barat	33,61	28,95	8,98
33	Papua	28,01	34,55	14,58

Lalu hitung jarak terdekat centroid menggunakan *Euclidean Distance.*

Tabel 2 13 Hasil Perhitungan jarak data pada Iterasi 3

No	Jarak Terpendek	Hasil	C1	C2
1	1,14	C2		1
2	5,52	C2		1
3	7,86	C2		1

No	Jarak Terpendek	Hasil	C1	C2
4	7,46	C1	1	
5	2,84	C1	1	
6	2,28	C1	1	
7	6,48	C1	1	
8	4,4	C1	1	
9	3,14	C1	1	
10	8,7	C2		1
11	7,07	C1	1	
12	1,21	C1	1	
13	0,12	C1	1	
14	2,07	C1	1	
15	0,3	C1	1	
16	1,88	C1	1	
17	2,58	C1	1	
18	5,45	C1	1	
19	3,47	C2		1
20	1,95	C1	1	
21	6,97	C1	1	
22	5,23	C1	1	
23	3,89	C1	1	
24	5,14	C1	1	
25	4,23	C1	1	
26	5,19	C2		1
27	7,76	C2		1
28	0,65	C1	1	
29	0,36	C2		1
30	5,88	C2		1
31	5,8	C2		1
32	8,98	C2		1
33	14,58	C2		1

Berikut adalah hasil perhitungan jarak data ke titik pusat kluster pada iterasi 2:

Tabel 2.14 Hasil Kluster pada Iterasi 3

Kluster	Hasil
C1	21
C2	12

Selanjutnya menghitung titik pusat cluster baru menggunakan hasil dari setiap anggota data pada masing-masing cluster. Berikut perhitungan titik pusat cluster baru iterasi ke tiga (x):

$$C1x = \frac{\begin{array}{c}55{,}11 + 65{,}40 + 64{,}84 + 69{,}05 + 66{,}96 + 65{,}70 + \\ 55{,}49 + 63{,}77 + 62{,}69 + 60{,}50 + 62{,}86 + 60{,}68 + 65{,}14 + \\ 57{,}11 + 64{,}51 + 69{,}53 + 67{,}80 + 58{,}68 + 67{,}71 + 58{,}33 + 63{,}22\end{array}}{21}$$

$$= 63{,}10$$

$$C2x = \frac{\begin{array}{c}43{,}72 + 48{,}11 + 50{,}45 + 51{,}29 + 39{,}11 + 47{,}78 + \\ 50{,}34 + 42{,}23 + 36{,}71 + 48{,}39 + 33{,}61 + 28{,}01\end{array}}{12}$$

$$= 43{,}31$$

Tabel 2 15 Nilai Centroid Baru (Iterasi 4)

C1	Maximum	63,10
C2	Minimum	43,31

Perhitungan rata-rata data pada titik centroid pada kluster 1:

$F\ (1.1) = \sqrt{(63{,}10 - 43{,}72^2} = 19{,}37$

$F\ (1.2) = \sqrt{(43{,}31 - 48{,}11)^2} = 14{,}99$

Lakukan sampai dengan F (1.33)

Perhitungan rata-rata data pada titik centroid pada cluster 2:

$F\ (1.1) = \sqrt{(63{,}10 - 43{,}72^2} = 0{,}41$

$F\ (1.2) = \sqrt{(43{,}31 - 48{,}11)^2} = 4{,}80$

Lakukan sampai dengan F (2.33)

Berikut adalah hasil dari perhitungan rata-rata data pada titik pusat centroid setiap kluster:

Tabel 2.16 Hasil Perhitungan centroid pada setiap Kluster (Iterasi 4)

No	Provinsi	Rata-Rata	C1	C2
1	Aceh	43,72	19,37	0,41
2	Sumatera Utara	48,11	14,99	4,80
3	Sumatera Barat	50,45	12,65	7,13
4	Riau	55,11	7,99	11,79
5	Jambi	65,4	2,3	22,09
6	Sumatera Selatan	64,84	1,74	21,53
7	Bengkulu	69,05	5,95	25,73
8	Lampung	66,96	3,86	23,65
9	Kep, Bangka Belitung	65,7	2,6	22,39
10	Kep, Riau	51,29	11,81	7,98
11	Dki Jakarta	55,49	7,61	12,18
12	Jawa Barat	63,77	0,68	20,46
13	Jawa Tengah	62,69	0,41	19,37
14	Di Yogyakarta	60,5	2,6	17,18
15	Jawa Timur	62,86	0,24	19,54
16	Banten	60,68	2,41	17,37
17	Bali	65,14	2,04	21,83
18	Nusa Tenggara Barat	57,11	5,99	13,80
19	Nusa Tenggara Timur	39,11	23,98	4,20
20	Kalimantan Barat	64,51	1,41	21,20
21	Kalimantan Tengah	69,53	6,43	26,22
22	Kalimantan Selatan	67,8	4,7	24,48
23	Kalimantan Timur	58,68	4,42	15,36
24	Sulawesi Utara	67,71	4,61	24,39

No	Provinsi	Rata-Rata	C1	C2
25	Sulawesi Tengah	58,33	4,77	15,02
26	Sulawesi Selatan	47,78	15,32	4,46
27	Sulawesi Tenggara	50,34	12,75	7,03
28	Gorontalo	63,22	0,12	19,90
29	Sulawesi Barat	42,23	20,87	1,09
30	Maluku	36,71	26,39	6,60
31	Maluku Utara	48,39	14,71	5,08
32	Papua Barat	33,61	29,49	9,70
33	Papua	28,01	35,09	15,30

Lalu hitung jarak terdekat centroid dengan menggunakan *Euclidean Distance*:

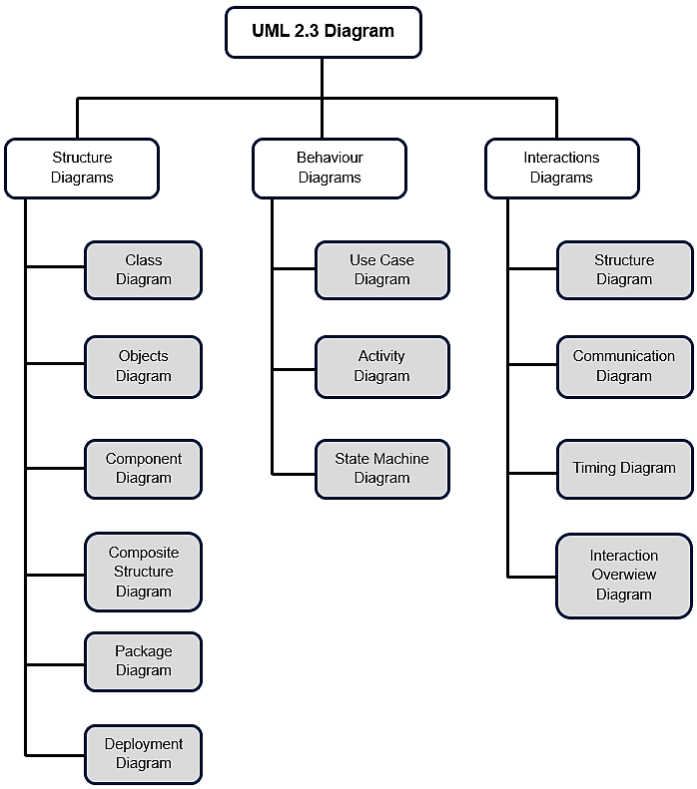
Tabel 2 17 Hasil Perhitungan jarak Data pada Iterasi 4

No	Jarak Terpendek	Hasil	C1	C2
1	0,41	C2		1
2	4,8	C2		1
3	7,13	C2		1
4	7,99	C1	1	
5	2,3	C1	1	
6	1,74	C1	1	
7	5,95	C1	1	
8	3,86	C1	1	
9	2,6	C1	1	
10	7,98	C2		1
11	7,61	C1	1	
12	0,68	C1	1	
13	0,41	C1	1	
14	2,6	C1	1	
15	0,24	C1	1	
16	2,41	C1	1	
17	2,04	C1	1	
18	5,99	C1	1	
19	4,2	C2		1
20	1,41	C1	1	
21	6,43	C1	1	
22	4,7	C1	1	
23	4,42	C1	1	
24	4,61	C1	1	
25	4,77	C1	1	
26	4,46	C2		1
27	7,03	C2		1
28	0,12	C1	1	
29	1,09	C2		1
30	6,6	C2		1

31	5,08	C2		1
32	9,7	C2		1
33	15,3	C2		1

4. *Unified Modeling Language* (UML)

UML adalah keluarga notasi grafis yang didukung oleh meta-model tunggal, yang membantu pendeskripsian dan desain sistem perangkat lunak, khususnya sistem yang dibangun menggunakan pemrograman berorientasi objek (OO), hal ini dikarenakan oleh sejarahnya sendiri dan oleh perbedaan persepsi tentang apa yang membuat sebuah proses rancang-bangun perangkat lunak efektif (Fowler, 2005). UML muncul karena adanya kebutuhan pemodelan visual untuk menspesifikasikan, menggambarkan, membangun, dan mendokumentasikan sistem perangkat lunak, UML versi 2.3 terdiri dari 13 macam diagram yang dikelompokkan dalam 3 kategori, pembagian kategori dan macam-macam diagram dapat dilihat pada gambar 2.1



Gambar 2.1 Bagan UML

Berdasarkan Gambar 2.1 berikut penjelasan singkat dari pembagian kategori tersebut (Hasanah & Untari, 2020):

(a). Structure diagram, merupakan kumpulan diagram yang digunakan untuk menggambarkan strutur statis dari sistem yang dimodelkan.

(b). Behavior diagram, merupakan kumpulan diagram yang digunakan untuk menggambarkan kelakuan sistem atau rangkaian perubahan yang terjadi pada suatu sistem.

(c). Interaction diagram, merupakan kumpulan diagram yang digunakan untuk menggambarkan interaksi sistem dengan sistem lain maupun interaksi antar subsistem pada suatu sistem.

Pada penelitian & pengembangan ini sendiri hanya menggunakan beberapa diagram dari 13 macam diagram yang telah diperlihatkan pada gambar 2.1. Diagram yang dipakai hanya dua yaitu *Structure Diagrams* dan *Behaviour Diagrams*, pada dua diagram tersebut pun masih ada beberapa diagram yang digunakan antara lain pada *Structure Diagrams,* diagram yang digunakan yaitu *Class Diagram*, *Deployment Diagram* dan *Component Diagram* sedangkan untuk *Behaviour Diagrams*, diagram yang digunakan hanya *Use Case Diagram.*

5. Basis Data

Basis adalah Gudang / markas / tempat berkumpul / tempat bersarang, data adalah Representasi fakta dunia nyata yang mewakili suatu obyek (manusia, benda, kejadian, dll) yang disimpan dalam bentuk teks, angka, gambar, bunyi, simbol, atau kombinasinya, basis data adalah kumpulan dari item data yang saling berhubungan satu dengan lainnya yang diorganisasikan berdasar sebuah skema atau struktur tertentu, tersimpan di hardware komputer dan dengan software digunakan untuk melakukan manipulasi data (diperbaharui, dicari, diolah dengan perhitunganperhitungan tertentu, dan dihapus) dengan tujuan tertentu (Fikry, 2019, pp. 1-2). Perusahaan menyimpan setiap data -data penting yang dimiliki ke dalam basis data, kekuatan basis data berasal dari sebuah pengetahuan dan teknologi yang telah berkembang pesat dan diwujudkan dalam perangkat lunak khusus, adapun perangkat lunak khusus tersebut dikenal sebagai sistem manajemen basis data, atau *Database Management System* (DBMS), atau disebut juga Sistem Basis Data (Putri, 2022, p. 4). Menurut (Fikry, 2019, p. 2) dalam sistem basis data memiliki beberapa komponen yaitu:

(1) Perangkat Keras (*Hardware*), yang biasanya terdapat dalam sistem basis data adalah memori sekunder *hardisk.*

(2) Sistem Operasi (*Operating System*), yang mengaktifkan atau mengfungsikan sistem komputer, mengendalikan seluruh sumber daya (*resource*) dan melakukan operasi-operasi dalam komputer.

(3) Basis Data (*Database*), sebuah basis data dapat memiliki sekumpulan data yang terorganisisr dengan baik sehingga data tersebut mudah disimpan, diakses, dan juga dapat diubah.

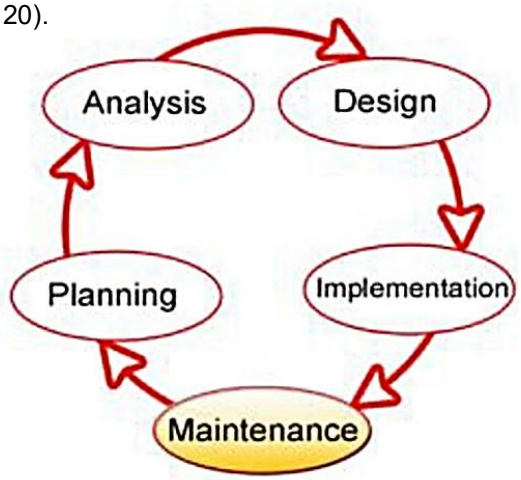
(4) *Management System* (DBMS), pengolahan basis data secara fisik tidak dilakukan oleh pemakai secara langsung, tetapi ditangani oleh sebuah perangkat lunak yang disebut DBMS.

(5) Pemakai (*User*), dapat berinteraksi dengan basis data dan memanipulasi data dalam program yang ditulis dalam bahasa pemograman

(6) Aplikasi atau Perangkat Lain, tergantung kebutuhan, pemakai basis data bisa dibuatkan program khusus untuk melakukan pengisian, pengubahan atau pengambilan data yang mudah dalam pemakainya.

6. *Software Development Life Cycle* (SDLC)

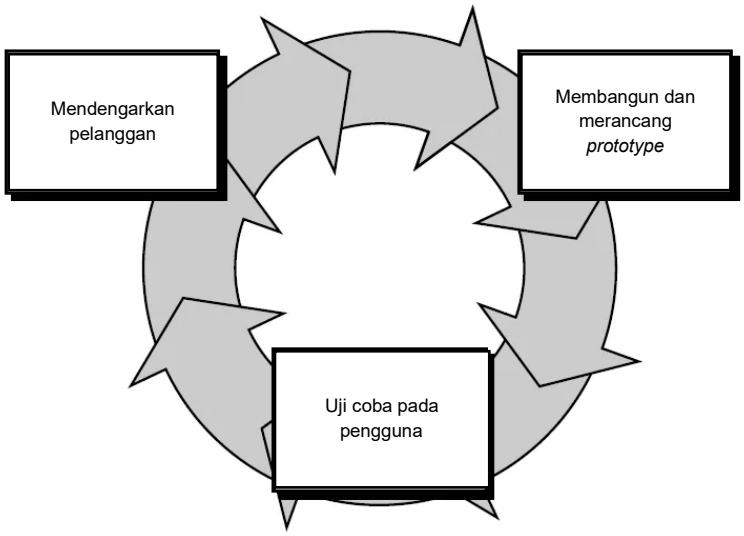
SDLC atau Software Development Life Cycle merupakan proses pengembangan atau mengubah suatu system perangkat lunak dengan menggunakan model-model dan metodologi yang digunakan orang untuk mengembangkan system perangkat lunak, SDLC juga merupakan pola yang diambil untuk mengembangkan sistem perangkat lunak, yang terdiri dari tahap-tahap: rencana(planning), analisis (analysis), desain (design), implementasi (implementation), uji coba (testing) dan pengelolaan (maintenance) (Hasanah & Untari, 2020, p. 20).



Gambar 2.2 Model Siklus Pengembangan Sistem
(Hasanah & Untari, 2020)

Dalam rekayasa perangkat lunak, konsep SDLC mendasari berbagai jenis metodologi pengembangan perangkat lunak, metodologi-metodologi ini membentuk suatu kerangka kerja untuk perencanaan dan pengendalian pembuatan sistem informasi, yaitu 21 proses pengembangan perangkat lunak, hal yang paling penting yaitu mengenali type pelanggan/ customer dan memilih menggunakan model SDLC yang sesuai dengan karakter pelanggan dan sesuai dengan karakter pengembang (Hasanah & Untari, 2020, p. 21). *Prototyping* merupakan proses merancang sebuah *prototype* dimana sebuah model produk yang mungkin belum memiliki semua fitur produk sesungguhnya namun sudah

memiliki fitur – fitur utama dari produkl sesungguhnyaa dan biasa digunakan untuk keperluan testing/uji coba untuk bahan uji coba sebeluym berlanjut ke fase pembuatan produk sesungguhnya (Hasanah & Untari, 2020, p. 23), model protoyping dapat dilihat pada gambar 2.4 berikut:



Gambar 2.3 Ilustrasi Model Prototype menurut Roger S. Pressman dalam (Hasanah & Untari, 2020)

tahapan pengembangan model dari gambar 2.3 dapat dijelaskan sebagai berikut:

(1) Mendengarkan pelanggan, pada tahap ini dilakukan pengumpulan kebutuhandari system dengan cara mendengar keluhan dari pelanggan. Untuk membuat suatu system yang sesuai kebutuhan, maka harus diketahui terlebih dahulu bagaimana system yang sedang berjalan untuk kemudian mengetahui masalah yang terjadi.

(2) Merancang dan Membuat Prototype, pada tahap ini, dilakukan perancangan dan pembuatan prototype system. Prototype yang dibuat disesuaikan dengan kebutuhan system yang telah didefinisikan sebelumnya dari keluhan pelanggan atau pengguna.

(3) Uji coba Pada tahap ini, Prototype dari system di uji coba oleh pelanggan atau pengguna. Kemudian dilakukan evaluasi kekurangan-kekurangan dari kebutuhan pelanggan. Pengembangan kemudian kembali mendengarkan keluhan dari pelanggan untuk memperbaiki Prototype yang ada.

7. Streamlit

Streamlit adalah framework Python yang digunakan untuk membangun aplikasi web dengan antarmuka pengguna interaktif untuk proyek-proyek data science dan machine learning, streamlit juga menyediakan API sederhana untuk input dan output data, sehingga mudah untuk terhubung ke berbagai sumber data dan API. (Tholib, 2023, p. 12). Streamlit dikembangkan untuk berfokus pada model *prototype* aplikasi berbasis data dan pembelajaran mesin, adapun kelebihan lain dari *framework* streamlit yaitu tidak membutuhkan sumberdaya yang banyak dalam menjalankan programnya, selain itu, streamlit banyak digunakan karena dapat meningkatkan efektivitas dalam pengembangan *web* baik dari segi waktu ataupun biaya (Raghavendra, 2023, p. 2). *Streamlit library* yang digunakan merupakan hasil penggabungan antara *front end* dengan *back end* sebagai aplikasi, pengguna tidak perlu lagi menggunakan pengembangan web manual untuk mengembangkan aplikasi berbasis data karena *streamlit* secara otomatis akan membuatnya (Raghavendra, 2023, p. 4)

8. *Python*

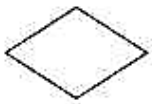
Python adalah bahasa pemrograman dinamis, tingkat tinggi, dimana merupakan bahasa pemrograman interpreter yaitu bahasa yang mengkonversi source code menjadi machine code secara langsung ketika program dijalankan, sintaks dan pengetikan python sangat dinamis dengan sifat interpretasinya menjadikannya bahasa yang ideal untuk skrip dan pengembangan aplikasi yang cepat, python mendukung banyak pola pemrograman, termasuk gaya pemrograman berorientasi objek, imperatif, dan fungsional serta prosedural (Suharto, 2023, p. 1). Menurut (Suharto, 2023, p. 10) python menyediakan banyak fitur berguna yang membuatnya populer dan berharga dari bahasa pemrograman lain, python mendukung pemrograman berorientasi objek, pendekatan pemrograman prosedural dan menyediakan alokasi memori dinamis. Python dapat digunakan di server untuk membuat aplikasi web, python juga dapat digunakan bersama perangkat lunak untuk menyusun alur kerja dan Python juga dapat terhubung ke sistem basis data, selain itu bahasa pemrograman Python dapat digunakan untuk menangani data besar, dan melakukan perhitungan yang kompleks, serta dapat digunakan untuk pembuatan *prototype*, atau untuk pengembangan perangkat lunak siap produksi (MA'ARIF, 2020, p. 7).

9. *Business Process Modelling Notation* (BPMN)

BPMN merupakan tahapan awal dalam rangkaian aktifitas pemodelan proses bisnis yang dikeluarkan oleh BPMI (Business Process Management initiative), BPMN menggambarkan suatu Diagram Proses Bisnis (BPD) yang

didasarkan pada suatu flowcharting teknik yang dikhususkan untuk menciptakan model grafis dari operasi proses bisnis, sehingga dapat dihasilkan bentuk diagram yang sederhana, mudah dipahami dan cukup menggambarkan proses secara keseluruhan (Wasilah & Karnila, 2017, p. 43). Dalam pengunaannya BPMN menyediakan sejumlah notasi dasar dan notasi-notasi tambahan. Notasi tambahan dapat digunakan untuk menggambarkan proses yang lebih komplek. Hal ini dapat dilakukan tanpa mengubah bentuk diagram awal. Terdapat empat (4) kategori elemen, yaitu (Wasilah & Karnila, 2017, p. 44):

(1) Flow Objects;
(2) Connecting objects;
(3) Swimlanes;
(4) Artifacts.

	*Event Start*, menyatakan sesuatu yang terjadi selama proses berlangsung dan mempengaruhi aliran proses, dimana dalam hal ini notasi menyatakan event yang merupakan titik dimulainya suatu proses.
	*Event Stop*, menyatakan event yang merupakan titik akhir suatu proses.
	*Activity*, menyatakan pekerjaan yang dilakukan oleh suatu pelaku
	*Gateway*, menyatakan percabangan atau pertemuan aliran dalam proses
	*Event Start with Message Trigger*, menyatakan titik awal proses yang dimulai akibat adanya penyebab berupa informasi atau pesan
	Sequential Flow, menyatakan urutan / sekuens suatu aktifitas dilakukan dalam proses

Gambar 2.4 Notasi BPMN

BPMN membentuk Business Process Diagram (BPD), yang didasarkan dan disesuaikan untuk model grafis dari operasi proses bisnis. Flow Objects Suatu diagram proses bisnis dimungkinkan terdiri dari tiga elemen utama, yaitu event, activity, dan gateway yang masing-masing diuraikan sebagai berikut :

(a). Event : adalah sesuatu yang “terjadi” selama perjalanan proses bisnis. Events mempengaruhi alur proses dan biasanya memiliki Pemodelan

Proses Bisnis 45 penyebab (cause/trigger) dan akibat (impact / result). Ada 3 macam event yang mungkin yaitu start, intermediate dan end.

Gambar 2.5 Lambang Event

(b). Activity : diwakili oleh kotak dengan sudut bulat dan menunjukkan pekerjaan yang dilakukan dalam proses bisnis. Tipe dari activity adalah Task dan Sub Process. Sub Process dibedakan dengan menambahkan tanda plus di bagian bawah tengah dari kotak.

Gambar 2.6 Lambang Activity

(c). Gateway : diwakili oleh bentuk mutiara dan dipakai untuk menentukan pertemuan atau pemisahan dari aliran proses. Digunakan untuk keputusan apakah pemisahan, penyatuan atau penggabungan alur.

Gambar 2.7 Lambang Gateaway

Tiga elemen diatas merupakan elemen dasar pembentuk struktur suatu proses bisnis. Terdapat tiga jenis konektor yang berfungsi sebagai penghubung yaitu: sequence flow, massage flow dan association yang masing-masing diuraikan sebagai berikut :

(a). Sequence Flow : digambarkan dengan garis tak putus dengan kepala panah penuh, dan digunakan untuk menunjukkan urutan aktifitas (sequence) di dalam proses.

Gambar 2.8 Lambang Sequence Flow

(b). Message Flow : digambarkan dengan garis putus dan kepala panah terbuka, dan digunakan untuk menunjukkan aliran pesan antara dua Process Participants. Dalam BPMN dua pool yang berbeda dianggap dua participant yang berbeda.

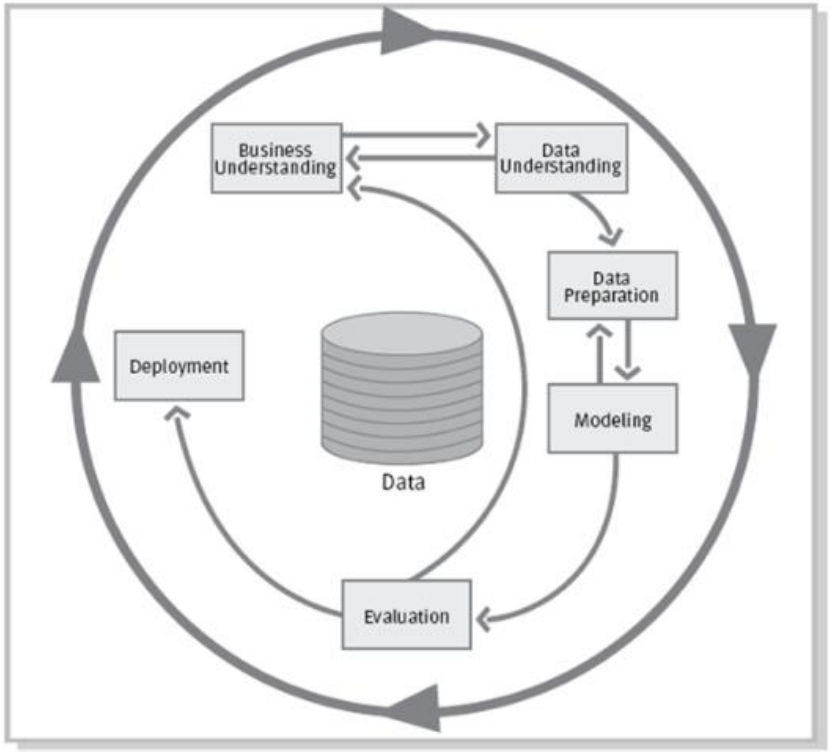
Gambar 2.9 Lambang Message Flow

(c). Association : digambarkan dengan garis titik-titik dengan kepala panah berupa garis, dan digunakan untuk mengasosiasikan data, teks dan Artifact lain dengan flow objects. Association digunakan untuk menunjukkan input dan output dari Activity.

Gambar 2.10 Lambang Association

10. *Cross – Industry Standard Process for Data Mining* (CRISP-DM)

Menurut (Ginantra, Arifah, & Wijaya, 2021, p. 15), dalam melaksanakan berbagai proyek dalam data mining secara sistematis, suatu proses yang dijadikan standar umum biasanya diterapkan. Salah satu prosesnya paling populer adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Dalam CRISP-DM, suatu proyek data mining mempunyai siklus hidup yang dibagi ke dalam enam tahap atau fase yang berurutan dan bersifat adaptif. Keluaran dari fase sebelumnya akan sangat berpengaruh bagi fase berikutnya. Dapat dilihat pada gambar 2.11 terkait proses CRISP-DM menurut Chapman dalam (Ginantra, Arifah, & Wijaya, 2021, p. 16) :



Gambar 2.11 Proses Data Mining CRISP-DM
(Ginantra, Arifah, & Wijaya, 2021)

(1) *Business Understanding*

Fase pertama dalam CRISP-DM ini merupakan tahapan yang cukup penting, karena di dalam fase ini memerlukan pengetahuan dari objek bisnis, proses membangun dan mendapatkan data serta mencocokkan tujuan permodelan dengan tujuan bisnis. Adapun aktivitas yang dilakukan antara lain:

(a). Menetapkan tujuan proyek serta kebutuhan dalam lingkup bisnis dengan detail atau dalam penelitian secara keseluruhan dengan jelas.

(b). Menerjemahkan tujuan dan menentukan batasan dan tujuan dalam perumusan permasalahan data mining

(c). Menyiapkan awalan strategi guna mencapai sasaran tersebut

(2) *Data Understanding*

Tahap ini dilaksanakan untuk memberikan fondasi analitik pada suatu penelitian dengan menyusun ringkasan atau summary dan mengidentifikasi potensi berbagai masalah yang ada di dalam data. Tahap data understanding ini juga harus dilakukan dengan teliti dan hati-hati, karena dikhawatirkan jika terdapat kendala pada fase ini yang belum terjawab, maka dapat mengganggu tahap modeling. Adapun kegiatan yang dilakukan dalam tahapan ini adalah sebagai berikut:

(a). Pengumpulan data

(b). Menggunakan hasil analisis penyelidikan data untuk mengenali data lebih lanjut serta proses pencarian pengetahuan awal.

(c). Mengevaluasi kualitas data yang sudah dihasilkan sebelumnya

(d). Jika dibutuhkan, kita dapat memilih sebagian kecil grup data yang memiliki kemungkinan dan mengandung pola masalah yang ada.

(3) *Data Preparation*

Fase data preparation ini memerlukan pemikiran matang dan upaya tinggi untuk memperbaiki masalah dalam data dan dibuat variable derived serta memastikan apakah data sudah tepat untuk algoritma yang digunakan. Tahap ini sering mengalami peninjauan ulang ketika menemukan kendala pada pembangunan model, sehingga dilaksanakan iterasi hingga menemukan hal yang sesuai dengan data yang dimaksud. Adapun aktivitas yang dilakukan antara lain:

(a). Menyiapkan data dari awal, termasuk himpunan data yang akan digunakan untuk keseluruhan fase yang berikutnya.

(b). Memilih kasus serta variabel yang sesuai dengan analisis yang akan dilaksanakan.

(c). Merubah beberapa variabel jika dibutuhkan

(d). Menyiapkan data awal hingga siap digunakan untuk perangkat permodelan.

(4) *Modelling*

Pada tahap ini, dilakukan metode statistika dan *machine learning* untuk menentukan teknik, alat bantu serta algoritma data mining yang akan diterapkan. Yang perlu digaris bawahi di sini, beberapa teknik memungkinkan untuk digunakan pada data mining yang memiliki permasalahan yang sama. Jika diperlukan penyesuaian data terhadap metode data mining, kita dapat kembali ke tahapan *data preparation*.

Dalam fase *modelling* ini, terdapat beberapa aktivitas yang dilakukan seperti:

(a). Memilih serta mengaplikasikan metode pemodelan yang sesuai.

(b). Mengalibrasi aturan model guna mendapatkan *output* yang optimal.

(5) *Evaluation*

Tahap *evaluation* ini merupakan tahap evaluasi dengan melaksanakan interpretasi terhadap *output* dari data mining yang dihasilkan dalam tahapan sebelumnya. Evaluasi di sini bertujuan agar model yang sudah ditentukan dapat sesuai dengan tujuan yang ingin dipenuhi pada fase pertama. Adapun aktivitas yang dilakukan dalam tahap *evaluation* ini antara lain:

(a). Melakukan evaluasi satu model atau lebih yang digunakan dalam tahap pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan pada tahap berikutnya

(b). Menetapkan model yang sudah memenuhi tujuan pada awal tahapan.

(6) *Deployment*

Tahap *deployment* atau rencana penggunaan model merupakan fase yang penting dalam proses CRISP-DM. Perencanaan untuk tahap *deployment* sudah dimulai sejak proses *Business Understanding* (fase pertama) dilakukan. Fase *deployment* ini tidak hanya menghasilkan suatu model, tapi juga mengonversi skor putusan serta menggabungkan keputusan dalam sistem operasional

Namun terbentuknya model tidak menandakan proyek selesai begitu saja. Model dibangun dari data yang mewakili pada jangka waktu tertentu, sehingga ada kemungkinan perubahan karakteristik data seiring perubahan waktu. (Ginantra, Arifah, & Wijaya, 2021, p. 18).

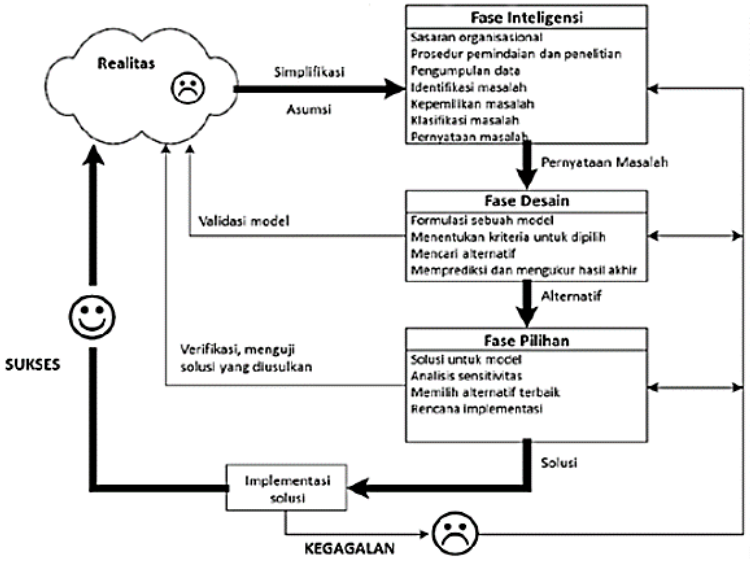
11. PSSUQ (Post-Study System Usability Questionnaire)

PSSUQ adalah kuesioner yang dirancang dengan tujuan untuk menilai kepuasan yang dirasakan pengguna terhadap sistem komputer, jenis perbandingan kelompok dengan menggunakan PSSUQ dinilai valid karena

seluruh *effect of respons* harus dihilangkan terlebih dahulu dari kondisi eksperimental, adapun potensi kritikal terhadap penggunaan PSSUQ yaitu, bahwa item tidak dapat mengikuti konvensi umum untuk kemudian memvariasikan keselarasan item, oleh karena itu, seringkali separuh hasil perhitungan PSSUQ akan bernilai positif dan separuh lainnya akan bernilai negatif (Wiley & Sons, 2012). PSSUQ merupakan kuesioner yang dibuat dengan tujuan untuk membantu pengguna mendapatkan kepuasan dalam penggunaan suatu sistem komputer, PSSUQ pada versi 3 yang digunakan sampai sekarang pun memiliki 16 poin penilaian yang berisikan pertanyaan-pertanyaan untuk pengguna dimana pengguna akan menjawab sesuai dengan kepuasan yang dirasakan (Salvendy, 2012, p. 1301).

12. Sistem Pendukung Keputusan

Hermawan mengatakan bahwa Sistem Pendukung Keputusan (SPK), dapat diartikan atau didefinisikan sebagai sebuah sistem yang dapat dan mampu memberikan solusi atau kemampuan baik kemampuan pemberian solusi atau pemecahan masalah maupun kemampuan mengkomunikasikan terhadap masalah masalah semi-terstruktur, SPK dideskripsikan atau dijelaskan sebagai sebuah sistem yang dapat mensupport kerja seorang pengambil keputusan dalam memecahkan / memberikan solusi terhadap masalah yang bersidat semiterstruktur melalui cara memberikan informasi ataupun saran menuju pada keputusan tertentu (Mashuri & Mujianto, 2021, pp. 5-6). Menurut Turban, proses pengambilan keputusan meliputi tiga fase utama yaitu inteligensi, desain, dan kriteria. Ia Kemudian menambahkan fase keempat yakini implementasi. Gambaran konseptual pengambilan keputusan menurut Turban dalam (Mashuri & Mujianto, 2021, p. 9) dapat dilihat pada gambar 2.12.

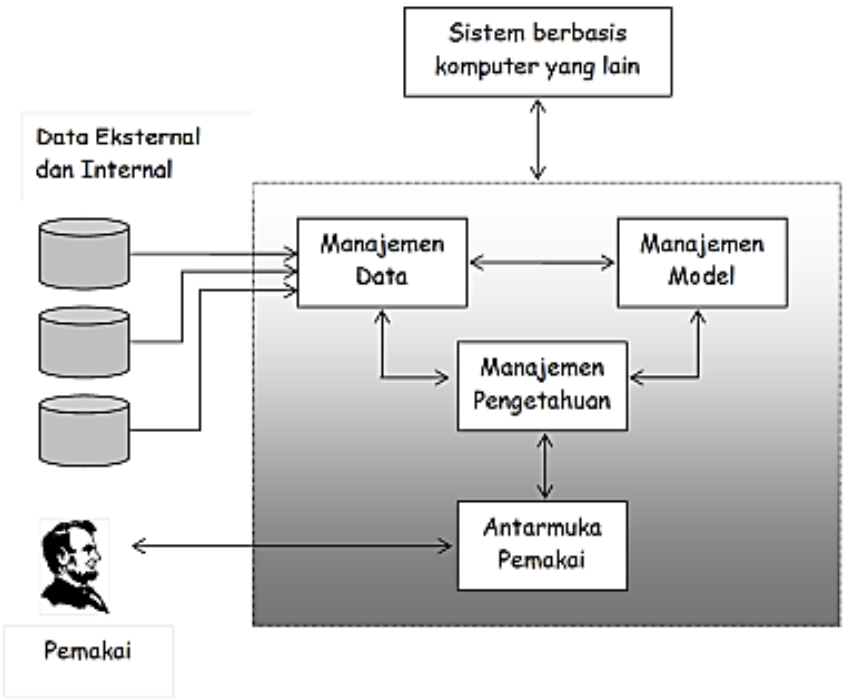


Gambar 2.12 Fase-Fase Pengambilan Keputusan/Proses Pemodelan SPK

Proses pengambilan keputusan dimulai dari fase intelegensi. Realitas diuji, dan masalah diidentifikasi dan ditentukan, selanjutnya pada fase desain akan dikonstruksi sebuah model yang merepresentasikan sistem, selanjutnya adalah fase pilihan yang meliputi pilihan terhadap solusi yang diusulkan untuk model (tidak memerlukan masalah yang disajikan), solusi ni diuji untuk menentukan viabilitasnya, begitu solusi yang diusulkan tampak masuk akal, maka kita siap untuk masuk kepada fase terakhir yakni fase implementasi keputusan (Mashuri & Mujianto, 2021, pp. 9-10).

Adapun Komponen-komponen SPK adalah sebagai berikut menurut (Mashuri & Mujianto, 2021, p. 17):

(1) Data Management Termasuk database, yang mengandung data yang relevan untuk berbagai situasi dan diatur oleh software yang disebut Database Management System (DBMS);
(2) Model Management Melibatkan model finansial, statistikal, management science, atau berbagai model kualitatif lainnya, sehingga dapat memberikan ke sistem suatu kemampuan analitis, dan manajemen software yang dibutuhkan;
(3) Communication User dapat berkomunikasi dan memberikan perintah pada DSS melalui subsistem ini. Ini berarti menyediakan antarmuka;
(4) Knowledge Management Subsistem optional ini dapat mendukung subsistem lain atau bertindak atau bertindak sebagai komponen yang berdiri sendiri. Untuk dapat lebih jelas memahami model konseptual SPK, dapat dilihat pada gambar 2.13.



Gambar 2.13 Model Konseptual SPK

B. Tinjauan Pustaka

Pada pelaksanaan penelitian ini, dilakukan proses pencarian informasi berdasarkan pada penelitian yang sudah dilakukan sebelumnya. Penelitian terkait dengan penentuan pola pelayanan dalam gereja sudah pernah dilakukan sebelumnya, dengan menggunakan algoritma K-Means. Penelitian tersebut berkontribusi sebagai bahan perbandingan dan juga sumber pengetahuan penelitian yang dilaksanakan. Berikut ini adalah beberapa penelitian yang menjadi pendukung pada penulisan penelitian ini:

(1) Implementasi Algoritma K-Means Untuk Mengenali Pola Jemaat Dalam Kegiatan Pelayanan Gereja (Pandie, Widiastuti, Mola, & Djahi, 2019)**;** pada penelitian didapatkan bahwa implementasi algoritma k-means clustering ke dalam sistem informasi klasterisasi memberikan hasil klasifikasi pengelompokan data yang efektif; proses setiap iterasi perputaran jarak centroid, penetuan titik cluster dibentuk; berdasarkan hasil clustering maka dapat menjadi catatan bagi pihak gereja yaitu jemaat yang belum dibaptis berjenis kelamin laki dan perempuan dengan kisaran usia menyebar di antara 1 tahun sampai dengan 80 tahun dan jumlah jemaat yang belum dibaptis mencapai 1031 orang; untuk kedepannya variabel pendukung bisa ditambahkan untuk dapat menganalisa faktor-faktor yang menyebabkan jemaat belum dibaptis, misalnya pendidikan, pekerjaan, jumlah anggota keluarga dan faktor sosial lainnya;

(2) Penerapan Metode Clustering Pada Sistem Informasi Pelayanan Di Gereja Pekabaran Injil Jalan Suci Yogyakarta (Sutopo & Duwit, 2018)**;** semakin berkembang dan majunya dunia teknologi, dapat mendukung untuk terciptanya manajemen yang baik dalam suatu organisasi gereja; teknologi Informasi berguna agar manajemen dalam organisasi dapat berjalan dan bekerja dengan baik; masalah yang dihadapi oleh Jalan Suci Yogyakarta pada saat ini adalah belum mempunyai Sistem informasi berbasis website yang mampu memberikan informasi pelayanan bagi jemaat; oleh karena itu perlu merancang sistem informasi pelayanan jemaat di Jalan Suci Yogyakarta dengan Metode clustering sebagai pendukung Sistem berbasis web; dengan adanya sistem informasi berbasis web site ini akan menghasilkan informasi yang akurat, tepat waktu dan relevan; hasil penelitian dan perancangan ini sangat bermanfaat bagi jemaat di Gereja Pekabaran Injil Jalan Suci Yogyakarta;

(3) Penerapan Algoritma K-Medoids Dalam Klasterisasi Penyebaran Tempat Ibadah Di Sumatera Utara (Maulana & Sundari, 2022)**;** tempat ibadah adalah fasilitas umum yang dibangun untuk memenuhi kebutuhan umat beragama dalam menjalankan kewajiban beribadah kepada Tuhan Yang Maha Esa; Tempat peribadatan di Provinsi Sumatera Utara antara lain mesjid, gereja, vihara, candi, dan pura; jumlah jamaah yang semakin banyak dapat mengakibatkan kapasitas tempat ibadah yang tidak

memadai sehingga masyarakat harus mencari tempat ibadah lainnya; hingga saat ini, pemerintah dan masyarakat sekitar berusaha menentukan lokasi pembangunan tempat ibadah yang strategis, yang dapat digunakan wisatawan dalam hal beribadah kepada Tuhan; mengingat pemerintah belum melakukan pemetaan untuk mengetahui daerah mana saja yang sudah dibangun atau belum menjadi tempat ibadah; agar proses menjadi lebih objektif tentunya diperlukan alat yaitu sistem informasi yang dapat mengolah data yang ada menjadi informasi; teknik yang digunakan dalam pengolahan data adalah data mining; metode yang digunakan adalah K-Medoids; implementasi algoritma k-medoids yang dilakukan dengan menggunakan Microsoft Visual Basic 2010; hasil yang diperoleh bahwa data persebaran tempat ibadah di Sumatera Utara tahun 2011 sampai 2020 yang terbagi menjadi 5 cluster dimana pada cluster 1 terdapat 44 anggota, klaster 2 berjumlah 23 anggota, klaster 3 berjumlah 49 anggota, klaster 4 berjumlah 64, dan klaster 5 berjumlah 150;

(4) Klasterisasi Jemaat Pemenerima Bantuan Sosial Dengan K-Means Clustering (Sitanggang, Rumapea, Sarkis, & Rumahorbo, 2022); proses klasterisasi jemaat yang layak menerima bantuan di Gereja HKBP Resort Pardomuan Nauli sedang mengalami masalah; bervariasinya data jemaat membuat pihak gereja kurang objektif dalam memberikan bantuan sosial kepada jemaatnya; untuk itu diperlukan sebuah metode pengklasifikasian jemaat yang penerima bantuan sosial untuk mengetahui dan mengelompokkan jemaat bedasarkan status ekonomi, sehingga mempermudah pihak gereja dalam menentukan jemaat mana saja yang layak menerima bantuan; dalam pendekatan pengklasteran K-Means, pembagian kelompok jemaat dapat dilakukan berdasarkan status, usia, pekerjaan, tanggungan, pendidikan, dan status kepemilikan rumah (SKR); dari penelitian yang dilakukan, terdapat 6 kelompok cluster yaitu C1= cluster kaya, C2 = cluster menengah atas, C3 = cluster menengah, C4 = cluster menengah bawah, C5 = cluster miskin, dan C6 = clustet sangat miskin, dengan tingkat akurasi 65,58% benar;

(5) Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means *Clustering* (Fakhri, Defit, & Sumijan, 2021)**;** knowledge Discovery in Database (KDD) merupakan proses analisa yang terstruktur bertujuan mendapatkan informasi yang baru dan benar, menemukan pola dari data yang komplek, dan bermanfaat; data mining merupakan inti dari proses KDD; clustering merupakan metode data mining yang cocok untuk pengoptimalisasikan pelayanan perpustakaan dikarenakan dapat mengklasterisasikan buku dengan dengan efektif dan efesien, dengan algoritma K-Means data dapat di clustering dan informasi dari setiap nilai centroid dari setiap cluster; pelayanan perpustakaan dapat mengoptimalisasikan penempatan buku sehingga santri bisa dengan cepat mencari buku sesuai dengan

minat bacanya dengan lebih efektif dan bisa tertarik dengan buku yang lain karena berada dalam satu pengelompokan; sedangkan untuk pihak perpustakaan bisa memprioritaskan untuk pengadaan buku selanjutnya; optimalisasi pelayanan perpustakaan di cluster dengan menggunakan metode K-Means.Clustering minat baca memiliki kreteria jumlah kesediaan buku, buku yang di pinjam, dan lama buku di pinjam; data buku di clustering menjadi 3 yaitu sangat diminiti, diminati, dan kurang diminati; setelah melakukan proses perhitungan dari 40 sampel jenis buku maka menghasilkan 6 kali iterasi, dan di dapatkan hasil akhir 3 clustering yaitu cluster 1 sebanyak 4 buku yang sangat diminati, cluster 2 sebanyak 20 buku yang diminati, dan cluster 3 sebanyak 16 buku yang kurang diminati; penelitiaan ini dapat dijadikan sebagai acuan rekomendasi untuk pengoptimalisasikan pelayanan perpustakaan baik untuk tata letak maupun pengadaan buku dengan memprioritaskan jenis buku yang sangat diminati;

(6) Pengelompokkan Daerah Berdasarkan Ketersediaan Masjid Muhammadiyah Dengan Algoritma K-Means (Purwayoga & Susanto, 2020)**;** pentingnya peranan masjid dalam kehidupan sehari-sehari yang sebagaimana diketahui adalah untuk sarana ibadah; namun ketersediaan masjid pada suatu daerah belum sepenuhnya merata, bahkan masih terdapat daerah yang belum memiliki masjid; kelebihan finansial pada organisasi Muhammadiyah belum dimanfaatkan dengan baik; salah satu solusinya yaitu dengan memetakan dan mengelompokkan suatu daerah berdasarkan jumlah masjid pada daerah tersebut; pengelompokkan daerah dalam penelitian ini menggunakan algoritma K-Means; karakteristik untuk proses pengelompokkan daerah adalah nama kabupaten, nama kecamatan dan jumlah masjid pada suatu kecamatan atau cabang; area studi dalam penelitian adalah provinsi jawa barat; penentuan K atau jumlah cluster dalam penelitian ini adalah 3; jumlah cluster ditentukan berdasarkan 3 kategori yaitu sedikit, sedang dan banyak; terdapat 4 daerah yang masuk kategori dengan jumlah masjid yang banyak, 21 daerah dengan jumlah sedang dan 78 untuk daerah yang berkategori sedikit; evaluasi hasil pengelompokkan menunjukkan hasil yang baik dengan nilai SSE sebesar 90.05 %;

(7) *Mapping Religious Harmony in the Special Capital Region Jakarta using K-Means Algorithm* (Istiawan, Sulistijanti, Santoso, & Ustyannie, 2023)**;** DKI Jakarta, yang sering disebut sebagai jendela Indonesia, merupakan kota yang dilanda berbagai masalah sosial yang kompleks karena statusnya sebagai salah satu kota terbesar di Indonesia; tujuan dari penelitian ini adalah untuk mengidentifikasi dan mengkaji kondisi kerukunan umat beragama di DKI Jakarta; penelitian sebelumnya tentang kerukunan beragama terutama menggunakan pendekatan berbasis indeks, yang hanya memberikan gambaran umum tentang kondisi kerukunan beragama tanpa

menunjukkan faktor-faktor spesifik yang mempengaruhi pengukurannya; dalam penelitian ini, pendekatan pengelompokan digunakan untuk menganalisis kerukunan umat beragama di DKI Jakarta; penelitian ini menunjukkan bahwa klaster 0 menunjukkan tantangan yang signifikan yang sangat mempengaruhi kerukunan umat beragama dibandingkan dengan klaster-klaster lainnya; dengan demikian, kebijakan pemerintah daerah yang bertujuan untuk membina kerukunan umat beragama harus memusatkan upaya mereka pada klaster 0, terutama berfokus pada variabel-variabel yang telah diidentifikasi memiliki tingkat yang rendah, seperti empati, anti-kekerasan, komitmen kebangsaan, dan kemampuan beradaptasi dengan budaya lokal;

(8) ***K-Means Method For Clustering Public Service Assessment of Government Organization In Kediri City*** (Arkham & Swanjaya, 2020)**;** Unit Pelayanan Publik(UPP) adalah merupakan unit kerja non struktural yang melakukan kegiatan penyelengaraan pelayanan publik langsung ke masyarakat; setiap tahun, Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi (PANRB), melalui Bagian Organisasi Pemerintah Kota Kediri, membentuk sebuah tim yang bertugas membuat asesmen tentang pelayanan publik di UPP wilayah kota Kediri; dalam pelaksanaan Asesmen tersebut masih menggunakan media kertas, kinerja kunjungan assesor ke tempat yang akan dilakukan monev tidak termonitoring dengan baik dan dan untuk menentukan sebuah lembaga mendapatkan predikat hasil yang baik masih belum punya standar yang jelas; maka, langkah yang tepat untuk mengatasi permasalahan tersebut adalah membuat aplikasi yang dapat menghemat penggunaan kertas, memantau kinerja, dan melakukan clustering untuk memberikan saran dan hasil yang lebih terukur; K-Means Clustering merupakan metode untuk mengklasterisasi input non hirarki kedalam beberapa bentuk cluster; penelitian ini berusaha membuktikan metode K-Means Clustering dapat diandalkan untuk membantu Assesor dalam mempertimbangkan kualitas dari UPP yang dilakukan asesmen; setiap masukan input akan diperhitungan menjadi 3 cluster predikat; hasilnya adalah persentasi predikat untuk dipertimbangkan menjadi rekomendasi untuk perbaikan kualitas UPP tersebut dimasa mendatang;

(9) ***Service Quality Dealer Identification: The Optimization of K-Means Clustering*** (Wella, Okfalisa, Insani, Saeed, & Hussin, 2023)**;** kualitas layanan dan kepuasan pelanggan secara langsung memengaruhi branding perusahaan, reputasi, dan loyalitas pelanggan; sebagai penghubung antara produsen dan konsumen, dealer harus menjaga hubungan konsumen yang berharga untuk meningkatkan kepuasan dan kepatuhan pelanggan; rendahnya pengukuran dan standarisasi yang komprehensif mengenai kualitas layanan muncul sebagai masalah pertimbangan terhadap keunggulan layanan perusahaan; dengan demikian, mengidentifikasi kinerja kualitas layanan dan pengelompokannya berkembang menjadi kontribusi yang

berharga dalam pengambilan keputusan untuk mengendalikan dan meningkatkan niat perusahaan; penelitian ini menerapkan Algoritma K-Means dengan mengoptimalkan jumlah cluster dalam mengidentifikasi kinerja kualitas layanan dealer; sehingga akan didapatkan formasi kualitas layanan terbaik; berdasarkan hasil analisis, didapatkan tiga kategori identifikasi dealer, yaitu cluster satu, dengan 125 dealer yang dikelompokkan sebagai dealer dengan performa baik; cluster dua, dengan 30 dealer yang dikelompokkan sebagai dealer dengan performa sangat baik; dan cluster tiga, dengan 38 dealer yang dikelompokkan sebagai dealer dengan performa kurang baik; guna mengevaluasi keampuhan nilai k yang optimum, daftar pendekatan pengujian dilakukan dan dibandingkan, di mana calinski-harabasz, elbow, silhouette score, dan davies-bouldin index (dbi) memberikan kontribusi pada nilai k=3; hasilnya, klaster optimum ditentukan melalui kinerja tertinggi dari nilai k yang berjumlah tiga;

(10) ***The Analysis of Online Worship Services Acceptance using the UTAUT 2 Method and Clustering K-Means*** (Phita & Nataliani, 2022)**;** karakteristik dan latar belakang dari sebuah jemaat gereja dan penerimaan teknologi dalam pelayanan ibadah online menjadi salah satu tolak ukur untuk mengetahui sikap atau perilaku pengguna dalam menerima sebuah teknologi; tujuan dari penelitian ini yaitu untuk melihat penerimaan teknologi jemaat terhadap pelayanan ibadah online menggunakan unified theory of acceptance and use of the technology 2 (utaut 2), dan jawaban responden dari berbagai kalangan dan usia dikelompokkan menggunakan clustering k-means; pengumpulan data dilakukan melalui kuesioner dengan jumlah responden sebanyak 220 responden; dari uji regresi linear berganda terhadap variabel utaut 2 didapatkan bahwa variabel habit pada penggunaan ibadah online memiliki pengaruh yang tingg terhadap variabel behavioral intention; selain itu variabel habit pada penggunaan ibadah online memiliki pengaruh yang tinggi terhadap variabel use behavior; selanjutnya, clustering dengan kmeans dilakukan untuk melihat kelompok usia yang puas terhadap ibadah online, berdasarkan empat cluster yang didapat menggunakan metode elbow dan enam kelompok usia; hasil dari clustering k-means dari seluruh variabel utaut 2, kelompok usia 12-16 tahun dan 26-35 tahun merupakan kelompok yang paling puas dan menerima ibadah online.

Berikut ini tabel tinjauan pustaka dimana penjelasan secara menyeluruh dengan sumber dan juga kontribusi terhadap penelitian & pengembangan ini dapat dilihat pada tabel 2.1;

Tabel 2.18 Tinjauan Pustaka

No	Nama Peneliti	Judul	Sumber	Kontribusi
1	Emersensye S. Y. Pandi; Tiwuk Widiastuti; Sebastianus A. S. Mola; Bertha S. Djahi	Implementasi Algoritma K-Means Untuk Mengenali Pola Jemaat Dalam Kegiatan Pelayanan Gereja	J-ICON, Vol. 7 No. 2, Oktober 2019, pp. 110~115 DOI 10.35508/jicon .v7i2.1679 http://ejurnal.undana. ac.id/jicon/article /view/1679	Hasil penelitian ini menunjukkan *clustering* k-means dapat diimplementasikan untuk mengenali pola jemaat seperti apa yang ada pada kegiatan pelayanan gereja menggunakan metode K-Means
2	Joko Sutopo; Obaja Duwit	Penerapan Metode Clustering Pada Sistem Informasi Pelayanan Di Gereja Pekabaran Injil Jalan Suci Yogyakarta	Januari 2018, pp 3~10 http://eprints .uty.ac.id /id/eprint/691	Hasil penelitian ini menunjukkan bahwa Perhitungan menggunakan Algoritma K-means bisa menentukan berapa jumlah data clustering yang akan di random oleh data mining
3	Didik Maulana; Siti Sundari	Penerapan Algoritma K-Medoids Dalam Klasterisasi Penyebaran Tempat Ibadah Di Sumatera Utara	SNASIKOM Vol. 2 No. 1 2022, pp. 51~59 https://proceeding.unive d.ac.id/index.php/ snasikom/ article/view/77	Penelitian Ini membantu dalam perhitungan untuk skala provinsi guna mengetahui klaster seperti apa yang dapat di gunakan untuk mengetahui penyebaran tempat ibadah
4	Mitha Febriana Sitanggang; Humuntal Rumapea; Indra M Sarkis; Benget Rumahorbo	Klasterisasi Jemaat Pemenerima Bantuan Sosial Dengan K-Means Clustering	METHOSISFO: Jurnal Ilmiah Sistem Informasi Vol. 2 No. 2, Oktober 2022, pp 19~25 http://ojs.fikom- methodist.net/ index.php/methosisfo	Hasil dari penelitian ini menunjukkan bagaimana pengelompokkan data jemaat pemenerima bantuan sosial dengan metode K-Means
5	Dwiki Aulia Fakhri; Sarjon Defit; Sumijan	Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means *Clustering*	JIDT : Jurnal Informasi Dan Teknologi 2021 Vol. 3 No. 3, Hal: 160~166 https://jidt.org/index .php/jidt/ article/view/137	Dari hasil penelitian ini dapat diketahui seperti apa penggunaan metode K-Means terhadap pengoptimalisasian pelayanan
6	Vega Purwayoga; Budi Susanto	Pengelompokkan Daerah Berdasarkan Ketersediaan	Jurnal Teknologi Volume 13 No. 1, Januari 2021, pp 75~80 https://jurnal.umj	Hasil penelitian ini menunjukkan bahwa pengelompokkan suatu daerah dapat di

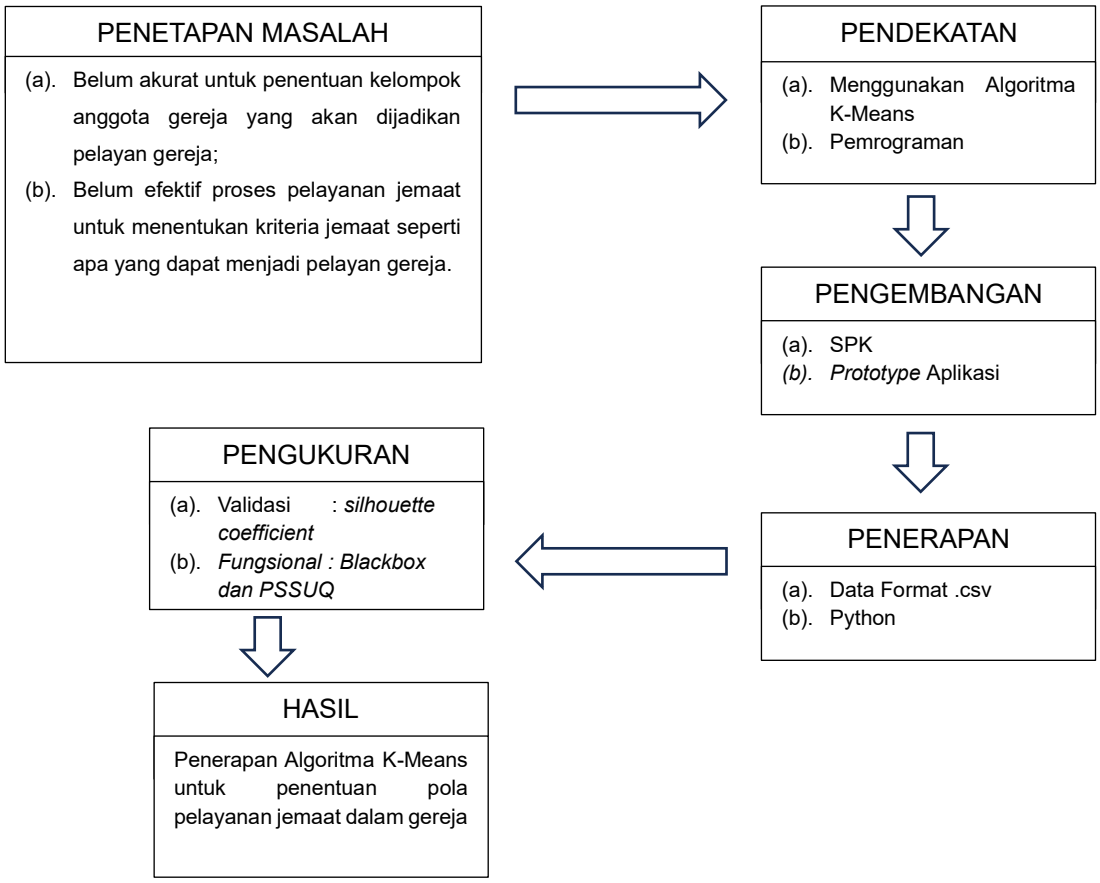
No	Nama Peneliti	Judul	Sumber	Kontribusi
		Masjid Muhammadiyah Dengan Algoritma K-Means	.ac.id/index.php/jurtek/article/view/6131	dapatkan dengan algoritma K-Means
7	Deden Istiawan; Wellie Sulistijanti; Arif Gunawan Santoso; Windyaning Ustyannie	*Mapping Religious Harmony in the Special Capital Region Jakarta using K-Means Algorithm*	Journal of Intelligent Computing and Health Informatics (JICHI) Vol. 4 No. 1, 2023, pp 27~34 https://jurnal.unimus.ac.id/index.php/ICHI/article/view/11715	Penelitian ini menunjukkan bahwa Dalam penelitian ini, pendekatan pengelompokan digunakan untuk menganalisis kerukunan umat beragama di DKI Jakarta.
8	Dany Arkham; Daniel Swanjaya	*K-Means Method For Clustering Public Service Assessment of Government Organization In Kediri City*	Seminar Nasional Inovasi Teknologi UN PGRI Kediri, Juli 2020, pp 155~160 https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/79	Penelitian ini berusaha membuktikan metode K-Means Clustering dapat diandalkan untuk membantu Assesor dalam mempertimbangkan kualitas dari UPP yang dilakukan asesmen.
9	Yolanda Enza Wella; Okfalisa Okfalisa; Fitri Insani; Faisal Saeed; Ab Razak Che Hussin	*Service Quality Dealer Identification: The Optimization of K-Means Clustering*	SINERGI Vol. 27 No. 3, October 2023, pp 433~442 http://publikasi.mercubuana.ac.id/index.php/sinergi http://doi.org/10.22441/sinergi.2023.3.014	Penelitian ini menerapkan Algoritma K-Means dengan mengoptimalkan jumlah cluster dalam mengidentifikasi kinerja kualitas layanan dealer. Sehingga akan didapatkan formasi kualitas layanan terbaik.
10	Gilang Jonathan Pitha; Yessica Nataliani	*The Analysis of Online Worship Services Acceptance using the UTAUT 2 Method and Clustering K-Means*	SISTEMASI: Jurnal Sistem Informasi Volume 11 Nomor 3, September 2022, pp 664~680 https://repository.uksw.edu//handle/123456789/26091	Hasil dari penelitian ini yaitu untuk melihat penerimaan teknologi jemaat terhadap pelayanan ibadah online menggunakan UTAUT 2, dan jawaban responden dari berbagai kalangan dan usia dikelompokkan menggunakan clustering k-Means.

Penelitian & pengembangan saat ini memastikan bahwa penggunaan teknik komputasi pemodelan Algoritma K-Means dalam penentuan pola pelayanan jemaat dalam gereja merupakan sebuah hal yang baru dalam kajian literatur. Melalui pemilihan secara

objektif, dipastikan bahwa sebelumnya belum ada penelitian & pengembangan yang membahas secara khusus dan rinci terkait penerapan metode k-means untuk penentuan pola pelayanan jemaat dalam gereja. Dalam referensi yang didapat yang berkaitan dengan metode dan juga permasalahan penelitian & pengembangan saat ini, tidak ditemukan kesamaan khusus dalam penelitian & pengembangan terdahulu. Dengan demikian penelitian & pengembangan ini dapat memberikan kontribusi baru terkait dengan penerapan metode k-means untuk penentuan pola pelayanan jemaat dalam gereja

C. Kerangka Pemikiran

Berdasarkan dengan landasan teoritis yang diperoleh dari penelitian-penelitian sebelumnya yang dijadikan sebagai rujukan penelitian, maka dapat disusun kerangka pemikiran sebagai berikut:



Gambar 2 14 Kerangka Pemikiran

Keterangan kerangka pemikiran pada gambar sebagai berikut :

(1) Pada penelitian ini diawali dengan penetapan permasalahan mengenai objek dengan melakukan identifikasi masalah diantaranya yaitu :

(a). Bagaimana penerapan Algoritma K-Means untuk pengelompokan pelayan jemaat dalam gereja?;

(b). Seberapa akurat dan efektif penerapan algoritma K-Means untuk pengelompokan pelayan jemaat dalam gereja.

(2) Selanjutnya dilakukan pengembangan terhadap analisis data, dan aplikasi dengan metode *Prototyping;*

(3) Penerapan dengan data format .csv dan bahasa pemrograman Python;

(4) Penelitian diukur menggunakan pengujian fungsional dan pengujian validasi dengan *Silhoueette Coefisien*

(5) Hasil dari penelitian ini adalah sistem informasi penentuan pola pelayanan jemaat dalam gereja

D. Hipotesis

Metode K-Means merupakan salah satu metode yang ada pada metode clustering yang merupakan metode data clustering non hirarki yang berusha mempartisi data ke dalam cluster dan data yang mempunyai karakteristik cluster yang berbeda ke dalam cluster lain. Merujuk pada penelitian menggunakan K-Means, telah digunakan dalam berbagai kasus untuk menganalisa data seperti yang digunakan dalam “Implementasi Algoritma K-Means Untuk Mengenali Pola Jemaat Dalam Kegiatan Pelayanan Gereja” (Pandie, Widiastuti, Mola, & Djahi, 2019), dapat diketahui bahwa masih belum akurat dan efektif. Permasalahan dalam penelitian ini adalah belum akurat dan efektif penentuan pola pelayanan jemaat berdasarkan wilayah sehingga algoritma K-Means sesuai dalam menentukan pola pelayanan jemaat. Oleh karena itu metode K-Means yang digunakan pada penelitian & pengembangan ini diharapkan bahwa penerapannya dapat digunakan dalam menentukan anggota jemaat yang akan menjadi pelayan gereja dengan pengelompokkan pola pelayanan jemaat.