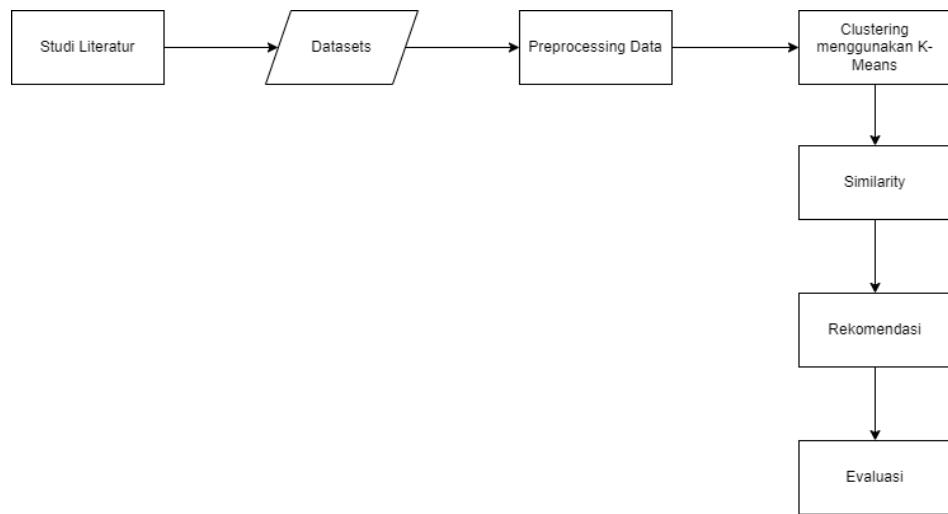


BAB III

METODE PENELITIAN

3.1 Diagram Penelitian

Penelitian ini memerlukan tahapan-tahapan atau alur proses yang dapat dilihat pada gambar 3.1.



Gambar 3.1 Alur Penelitian

3.1.1 Studi Literatur

Pada tahap studi literatur dilakukan pembelajaran dimulai dari teori-teori yang tersedia dari berbagai sumber berupa Jurnal, Buku dan penelitian lain yang berkaitan dengan sistem, serta sumber-sumber pendukung lain yang memiliki hubungan langsung dengan objek penelitian yang dipilih.

3.1.2 Preprocessing Data

Pada tahap *preprocessing* data, dilakukan pembersihan atau cleaning data dan transformasi data. Berikut *datasets* sebelum *preprocessing* dan setelah *preprocessing* yang dapat dilihat pada tahap di bawah ini.

Pada tahap sebelum *preprocessing* data, diambil sampel 50 *user* dari 943 *user* dan 30 *movie* dari 1682 *movie*. *User* tersebut akan mengalami

proses *cleaning* dan akan di ganti dengan data *user* selanjutnya sampai terbentuk 50 data yang dapat dilihat pada gambar 3.2.

user id	age	gender	occupation	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20
1	24	M	technician	5	3	4	3	3	0	4	1	5	0	2	0	5	0	5	5	0	4	5	0
2	53	F	other	4	0	0	0	0	0	0	0	0	2	0	0	0	4	0	0	0	0	0	0
3	23	M	writer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	24	M	technician	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
5	33	F	other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	42	M	executive	4	0	0	0	0	0	2	4	4	0	0	4	2	0	0	0	0	0	0	0
7	57	M	administrator	0	0	0	5	0	0	0	0	5	0	3	5	0	0	0	0	0	0	0	0
8	36	M	administrator	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
9	29	M	student	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
10	53	M	lawyer	0	0	0	0	0	0	0	0	4	0	4	5	0	0	0	0	0	0	0	0
11	39	F	other	0	0	0	0	0	0	0	4	0	0	2	0	0	0	5	0	0	0	0	0
12	28	F	other	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	47	M	educator	3	0	0	5	1	0	2	4	3	0	0	5	5	0	0	0	1	0	0	0
14	45	M	scientist	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5
15	49	F	educator	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	4	0	0	0	0
16	21	M	entertainment	5	0	0	5	0	0	0	0	0	0	0	5	5	0	0	5	0	0	0	0
17	30	M	programmer	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
18	35	F	other	0	0	0	3	0	5	0	5	9	0	0	0	0	0	4	0	0	0	0	0
19	40	M	librarian	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	42	F	homemaker	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0
21	26	M	writer	5	0	0	0	2	0	5	0	5	0	0	0	0	0	4	0	4	0	0	0
22	25	M	writer	0	2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
23	30	F	artist	0	0	0	0	0	0	4	4	0	0	0	0	0	4	0	0	0	0	4	0
24	21	F	artist	0	0	0	0	0	0	4	5	0	0	5	5	0	0	0	0	0	0	0	0

Gambar 3.2 Datasets sebelum *preprocessing* data

Pada tahap *cleansing* data, data yang dibersihkan adalah data yang berusia di bawah 17 tahun dan 60 tahun ke atas. Selain itu, data yang perlu dibersihkan adalah data user yang sama sekali tidak merating movie 1-30 yang dapat dilihat pada gambar 3.3.

93	48	M	executive	5	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
94	26	M	student	0	0	0	4	0	0	4	0	0	0	5	0	0	0	0	0	0	0	0	0
95	31	M	administrator	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
96	25	F	artist	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
97	43	M	artist	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98	49	F	executive	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
99	20	M	student	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100	36	M	executive	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101	15	M	student	3	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
102	38	M	programmer	3	2	0	2	0	0	2	0	0	0	3	0	3	0	0	0	0	0	0	0
103	26	M	student	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
104	27	M	student	0	0	3	0	0	0	0	0	2	2	0	0	3	0	5	0	0	0	0	0
105	24	M	engineer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
106	61	M	retired	0	0	0	0	0	0	0	4	0	0	4	0	4	0	4	0	0	0	0	0
107	39	M	scientist	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
108	44	M	educator	4	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
109	29	M	other	0	0	0	0	0	0	0	0	0	0	4	4	0	0	4	0	4	0	0	0
110	19	M	student	0	3	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0

Gambar 3.3 Tahap *Cleansing* Data

Pada tahap transformasi data, data yang memiliki rentang nilai yang berbeda seperti gender dan *occupation* akan di normalisasi ke dalam bentuk yang sama (numerik) yang dapat dilihat pada gambar 3.4 dan 3.5.

user id	age	gender	occupation	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	..	m30
1	24	1	20	5	3	4	3	3	0	4	1	5	0	2	0	5	...	3
2	53	2	14	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0
4	24	1	20	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
5	33	2	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	42	1	7	4	0	0	0	0	0	0	2	4	4	0	0	4	2	0
7	57	1	1	0	0	0	5	0	0	0	0	5	0	3	5	0	0	0
8	36	1	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
9	29	1	19	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
10	53	1	10	0	0	0	0	0	0	0	0	4	0	4	5	0	0	0
11	39	2	14	0	0	0	0	0	0	0	4	0	0	2	0	0	0	0
12	28	2	14	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
13	47	1	4	3	0	0	5	1	0	2	4	3	0	0	5	5	0	0
14	45	1	18	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0
15	49	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
16	21	1	6	5	0	0	5	0	0	0	0	0	0	0	5	5	0	0
17	30	1	15	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
18	35	2	14	0	0	0	3	0	5	0	5	9	0	0	0	0	0	0
20	42	2	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	26	1	21	5	0	0	0	2	0	5	0	5	0	0	0	0	0	0
22	25	1	21	0	2	0	5	0	0	0	0	0	0	0	0	0	0	0
23	30	2	2	0	0	0	0	0	0	4	4	0	0	0	0	0	0	0
24	21	2	2	0	0	0	0	0	0	4	5	0	0	0	5	5	0	0
25	39	1	5	5	0	0	0	0	0	4	0	0	0	0	0	0	0	0

Gambar 3.4 Tahap Transformasi Data

user id	age	gender	occupation	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	..	m30
1	24	1	20	5	3	4	3	3	0	4	1	5	0	2	0	5	...	3
2	53	2	14	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0
4	24	1	20	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
5	33	2	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	42	1	7	4	0	0	0	0	0	0	2	4	4	0	0	4	2	0
7	57	1	1	0	0	0	5	0	0	0	0	5	0	3	5	0	0	0
8	36	1	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
9	29	1	19	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
10	53	1	10	0	0	0	0	0	0	0	0	4	0	4	5	0	0	0
11	39	2	14	0	0	0	0	0	0	0	4	0	0	2	0	0	0	0
12	28	2	14	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
13	47	1	4	3	0	0	5	1	0	2	4	3	0	0	5	5	0	0
14	45	1	18	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0
15	49	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
16	21	1	6	5	0	0	5	0	0	0	0	0	0	0	5	5	0	0
17	30	1	15	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
18	35	2	14	0	0	0	3	0	5	0	5	9	0	0	0	0	0	0
20	42	2	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	26	1	21	5	0	0	0	2	0	5	0	5	0	0	0	0	0	0
22	25	1	21	0	2	0	5	0	0	0	0	0	0	0	0	0	0	0
23	30	2	2	0	0	0	0	0	0	4	4	0	0	0	0	0	0	0
24	21	2	2	0	0	0	0	0	0	4	5	0	0	0	5	5	0	0
25	39	1	5	5	0	0	0	0	0	4	0	0	0	0	0	0	0	0
26	49	1	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	40	2	11	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0
28	32	1	21	0	0	0	0	3	0	0	0	0	0	4	0	0	0	0
29	41	1	15	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0
32	28	2	19	0	0	0	0	0	0	4	0	3	0	0	0	0	0	0
37	23	1	19	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
42	30	1	1	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0
43	29	2	11	0	0	2	4	0	0	0	4	4	4	0	0	5	0	0
44	26	1	20	4	0	0	0	0	0	5	0	0	0	0	0	0	0	0
45	29	1	15	0	0	0	0	0	0	3	0	0	0	0	0	5	0	0

Gambar 3.5 Setelah Preprocessing Data

3.1.3 Clustering Menggunakan K-Means Clustering

Pada tahap ini, peneliti akan melakukan clustering data berdasarkan usia user saja dengan menggunakan *K-Means Clustering* yang dapat dilihat pada tabel 3.1.

Tabel 3.1 Sampel Data

User id	Age	gender	occupation	m1	m2	m3	m4	m5	...	m30
1	24	1	20	5	3	4	3	3	...	3
2	53	2	14	4	0	0	0	0	...	0
3	24	1	20	0	0	0	0	0	...	0
4	33	2	14	0	0	0	0	0	...	0
5	42	1	7	4	0	0	0	0	...	0
6	57	1	1	0	0	0	5	0	...	0
7	36	1	1	0	0	0	0	0	...	0
8	29	1	19	0	0	0	0	0	...	0
9	53	1	10	0	0	0	0	0	...	0
10	39	2	14	0	0	0	0	0	...	0
11	28	2	14	0	0	0	5	0	...	0
12	47	1	4	3	0	0	5	1	...	0
13	45	1	18	0	0	0	0	0	...	0
14	49	2	4	0	0	0	0	0	...	0
15	21	1	6	5	0	0	5	0	...	0
16	30	1	15	0	0	0	0	0	...	0
17	35	2	14	0	0	0	3	0	...	0

18	42	2	9	0	0	0	0	0	...	0
19	26	1	21	5	0	0	0	2	...	0
20	25	1	21	0	2	0	5	0	...	0
...
50	19	1	19	0	0	0	0	0	...	0

Contoh perhitungan *K-Means Clustering* :

1. Peneliti membagi data menjadi 3 *cluster* (K), yaitu C0, C1, dan C2.
2. Titik awal dari setiap *cluster* (*centroid*) adalah C0 = 21, C1 = 29, C2 = 47.
3. Menghitung jarak setiap data usia *user* ke setiap *centroid* menggunakan rumus persamaan (1) adalah sebagai berikut :

Rentang usia	17-59
C0	17-25
C1	26-39
C2	40-59

Kelompok C0 :

- a) Jarak *user* id 1 dengan *centroid* C0

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

$$D_{(1,C0)} = \sqrt{(24 - 21)^2}$$

$$D_{(1,C0)} = \sqrt{(3)^2}$$

$$D_{(1,C0)} = \sqrt{9}$$

$$D_{(1,C0)} = 3$$

- b) Jarak *user* 2 dengan *centroid* C0

$$D_{(2,C0)} = \sqrt{(53 - 21)^2}$$

$$D_{(2,C0)} = \sqrt{(32)^2}$$

$$D_{(2,C0)} = \sqrt{1024}$$

$$D_{(2,C0)} = 32$$

- c) Jarak user 3 dengan centroid C0

$$D_{(4,C0)} = \sqrt{(24 - 21)^2}$$

$$D_{(4,C0)} = \sqrt{(3)^2}$$

$$D_{(4,C0)} = \sqrt{9}$$

$$D_{(4,C0)} = 3$$

- d) Jarak user 4 dengan centroid C0

$$D_{(5,C0)} = \sqrt{(33 - 21)^2}$$

$$D_{(5,C0)} = \sqrt{12^2}$$

$$D_{(5,C0)} = \sqrt{144}$$

$$D_{(5,C0)} = 12$$

- e) Jarak user 5 dengan centroid C0

$$D_{(6,C0)} = \sqrt{(42 - 21)^2}$$

$$D_{(6,C0)} = \sqrt{21^2}$$

$$D_{(6,C0)} = \sqrt{441}$$

$$D_{(6,C0)} = 21$$

Kelompok C1 :

- a) Jarak user id 1 dengan centroid C1

$$D_{(1,C1)} = \sqrt{(24 - 29)^2}$$

$$D_{(1,C1)} = \sqrt{(-5)^2}$$

$$D_{(1,C1)} = \sqrt{25}$$

$$D_{(1,C1)} = 5$$

- b) Jarak user 2 dengan centroid C1

$$D_{(2,C1)} = \sqrt{(53 - 29)^2}$$

$$D_{(2,C1)} = \sqrt{(24)^2}$$

$$D_{(2,C1)} = \sqrt{576}$$

$$D_{(2,C1)} = 24$$

- c) Jarak user 3 dengan centroid C1

$$D_{(4,C1)} = \sqrt{(24 - 29)^2}$$

$$D_{(4,C1)} = \sqrt{(-5)^2}$$

$$D_{(4,C1)} = \sqrt{25}$$

$$D_{(4,C1)} = 5$$

- d) Jarak user 4 dengan centroid C1

$$D_{(5,C1)} = \sqrt{(33 - 29)^2}$$

$$D_{(5,C1)} = \sqrt{4^2}$$

$$D_{(5,C1)} = \sqrt{16}$$

$$D_{(5,C1)} = 4$$

- e) Jarak user 5 dengan centroid C1

$$D_{(6,C1)} = \sqrt{(42 - 29)^2}$$

$$D_{(6,C1)} = \sqrt{13^2}$$

$$D_{(6,C1)} = \sqrt{169}$$

$$D_{(6,C1)} = 13$$

Kelompok C2 :

- a) Jarak user id 1 dengan centroid C2

$$D_{(1,C2)} = \sqrt{(24 - 47)^2}$$

$$D_{(1,C2)} = \sqrt{(-23)^2}$$

$$D_{(1,C2)} = \sqrt{529}$$

$$D_{(1,C2)} = 23$$

- b) Jarak user 2 dengan centroid C2

$$D_{(2,C2)} = \sqrt{(53 - 47)^2}$$

$$D_{(2,C2)} = \sqrt{(6)^2}$$

$$D_{(2,C2)} = \sqrt{(36)}$$

$$D_{(2,C2)} = 6$$

- c) Jarak user 3 dengan centroid C2

$$D_{(4,C2)} = \sqrt{(24 - 47)^2}$$

$$D_{(4,C2)} = \sqrt{(-23)^2}$$

$$D_{(4,C2)} = \sqrt{529}$$

$$D_{(4,C2)} = 23$$

- d) Jarak user 4 dengan centroid C2

$$D_{(5,C2)} = \sqrt{(33 - 47)^2}$$

$$D_{(5,C2)} = \sqrt{(14)^2}$$

$$D_{(5,C2)} = \sqrt{196}$$

$$D_{(5,C2)} = 14$$

- e) Jarak user 5 dengan centroid C2

$$D_{(6,C2)} = \sqrt{(42 - 47)^2}$$

$$D_{(6,C2)} = \sqrt{(-5)^2}$$

$$D_{(6,C2)} = \sqrt{25}$$

$$D_{(6,C2)} = 5$$

4. Mencari jarak terdekat (terpendek) dengan membandingkan hasil perhitungan di atas dengan tiap centroid.

- a) User id 1, C0=3, C1= 5, C2=23

Jarak min = 3 (C0)

- b) User id 2, C0=32, C1= 24, C2=6

Jarak min = 6 (C2)

- c) User id 3, C0=3, C1= 5, C2=23

Jarak min = 3 (C0)

- d) User id 4, C0=12, C1= 4, C2=14

Jarak min = 4 (C1)

- e) User id 5, C0=21, C1= 13, C2=5

Jarak min = 5 (C2)

Berdasarkan perhitungan tersebut di dapatkan hasil C0, C1, C2 :

Centroid 0	Centroid 1	Centroid 2	Jarak Terpendek	Hasil Cluster	User id
3	5	23	3	C0	1
32	24	6	6	C2	2
3	5	23	3	C0	3
12	4	14	4	C1	4
21	13	5	5	C2	5

Kelompok anggota :

anggota cluster		
C0	C1	C2
1	4	2
3	7	5

5. Memperbaharui nilai *centroid*, nilai *centroid* baru yang diperoleh dari rata-rata *cluster* yang baru saja terbentuk dengan rumus :

$$\frac{\text{Jumlah anggota tiap cluster pada data usia user}}{\text{jumlah anggota tiap cluster}}$$

Rata-rata iterasi 1 C0

$$\frac{\text{user id } 1 + 3 + 15 + 20 + 22 + 29 + 34 + 35 + 36 + 38 + 40 + 48 + 49 + 50}{14}$$

$$\frac{\text{user id } 1 + 3 + 15 + 20 + 22 + 29 + 34 + 35 + 36 + 38 + 40 + 48 + 49 + 50}{14}$$

$$= \frac{24+24+21+25+21+23+23+21+18+22+25+23+17+19}{14} = 21,86$$

Rata-rata iterasi 1 C1

$$\frac{\text{user id } 4 + 7 + 8 + 11 + 16 + 17 + 19 + 21 + 26 + 28 + 30 + 31 + 32 + 33 + 37 + 39 + 41 + 44 + 45 + 46}{20}$$

$$\begin{array}{r} \text{user id } 4 + 7 + 8 + 11 + 16 + 17 + 19 + 21 + 26 + 28 + 30 + 31 + 32 + 33 \\ + 37 + 39 + 41 + 44 + 45 + 46 \\ \hline 20 \end{array}$$

$$= \frac{33+36+29+28+30+35+26+30+32+28+30+29+26+29+26+37+27+27+31+32}{20} \\ = 30,05$$

Rata-rata iterasi 1 C2

$$\begin{array}{r} \text{user id } 2 + 5 + 6 + 9 + 10 + 12 + 13 + 14 + 18 + 23 + 24 + 25 + 27 + 42 + \\ + 43 + 47 \\ \hline 16 \end{array}$$

$$\begin{array}{r} \text{user id } 2 + 5 + 6 + 9 + 10 + 12 + 13 + 14 + 18 + 23 + 24 + 25 + 27 + 42 + \\ + 43 + 47 \\ \hline 16 \end{array}$$

$$= \frac{53+42+57+53+39+47+45+49+42+39+49+40+41+49+50+51}{16} = 46,625$$

6. Setelah menghitung ulang, hasil *cluster* C0 pada iterasi 2 sudah tidak ada yang berubah sehingga cluster akhir dapat ditentukan. Setelah itu, tiap *cluster* dihitung menggunakan DBI agar optimal.

- a) Menghitung SSW

SSW = rata-rata dari jarak user dengan centroid dalam tiap cluster

SSW C0 , C1, dan C2 :

SSW	
C0	2,02
C1	2,57
C2	4,80

- b) Menghitung SSB

SSB = akar dari rata-rata tiap centroid pada iterasi 1

SSB	CENTROID		
	C0	C1	C2
C0	0	6,89	24,77
C1	6,89	0	18
C2	24,77	18	0

c) Menghitung Rasio

$$\text{Rasio} = \text{SSB C0} + \text{CENTROID C0} / \text{SSW C0}$$

ratio	
Rij C0, C1 =	0,67
R C0, C2 =	0,28
R C1, C2 =	0,42

d) Menghitung DBI

$$\text{DBI} = \frac{1}{K} \times \text{ratio}$$

DBI =	0,22	C0
	0,09	C1
	0,14	C2

3.1.4 Similarity

Setelah data dicluster, langkah selanjutnya adalah menghitung kemiripan demografi user untuk mendapatkan pola film yang paling direkomendasikan.

Berikut adalah contoh tampilan dari perhitungan *similarity*:

anggota cluster 0			
user id	age	gender	occupation
1	24	1	20
3	24	1	20
15	21	1	6
20	25	1	21
22	21	2	2
29	23	1	19
34	23	2	19
35	21	1	21
36	18	2	19
38	22	1	7
40	25	1	11
48	23	1	19
49	17	1	19
50	19	1	19

17-25
Definisi gender :
a) Gender 1 yaitu laki-laki.
b) Gender 2 yaitu perempuan.

RENTANG USIA		C0	C1	C2			
17-59		19-27	28-38	39-57			
asumsi data baru (x):							
user id	age	gender	occupation				
1	17	1	19	dilihat dari umur user 1 masuk C0			
2	37	2	9			C1	
3	46	2	18				C2

menghitung similarity user 1 (x) dan user 49 (y) di cluster 0				jumlah	sqrt	similarity
x	17	1	19			
y	17	1	19	651		
xy	289	1	361	651	25,515	1,000
x^2	289	1	361	651	25,515	
y^2	289	1	361			

menghitung similarity user 1 (x) dan user 50 (y) di cluster 0				jumlah	sqrt	similarity
x	17	1	19			
y	19	1	19	685		
xy	323	1	361	651	25,515	0,998
x^2	289	1	361	723	26,889	
y^2	361	1	361			

similarity (kemiripan) memiliki nilai 0-1 :
nilai 0 = tidak mirip
nilai 1 = sangat mirip

menghitung similarity user 2 (x) dan user 46 (y) di cluster 1				jumlah	sqrt	similarity
x	37	2	9			
y	32	1	4			
xy	1184	2	36	1222		
x^2	1369	4	81	1454	38,131	
y^2	1024	1	16	1041	32,265	0,993

menghitung similarity user 3 (x) dan user 27 (y) di cluster 2				jumlah	sqrt	sim
x	46	2	18			
y	41	1	15			
xy	1886	2	270	2158		
x^2	2116	4	324	2444	49,437	1,000
y^2	1681	1	225	1907	43,669	

3.1.5 Rekomendasi

Pada tahap ini, peneliti akan mengimplementasikan algoritma k-means clustering untuk mengelompokkan user berdasarkan usia. Setelah

perhitungan cluster dan similarity terbentuk, maka akan muncul rekomendasi film.

CLUSTER 0													
user (x)													
user id	age	gender	occupation	m1	...	m23	m24	m25	m26	m27	m28	m29	m30
1	17	1	19	0	...	0	4	4	0	0	0	0	0
user (y)				REKOMENDASI = m24 & m25 (di ambil dari 2-5 rekomendasi teratas)									

CLUSTER 1													
user (x)													
user id	age	gender	occupation	m1	...	m10	m11	...	m22	...	m28	m29	m30
2	37	2	9	0	...	5	4	...	4	...	4	0	0
user (y)				REKOMENDASI = m10, m11, m22 & m28 (di ambil dari 2-5 rekomendasi teratas)									

CLUSTER 2													
user (x)													
user id	age	gender	occupation	m1	...	m14	m15	...	m19	...	m28	m29	m30
3	46	2	18	0	...	0	5	...	0	...	5	3	0
user (y)				REKOMENDASI = m12, m15, m19, m28 & m8 (di ambil dari 2-5 rekomendasi teratas)									

3.1.6 Evaluasi

Pada penilitian ini evaluasi dilakukan untuk mengukur kemampuan dan keefektifan rekomendasi dan algoritma yang dihasilkan oleh sistem agar pengguna mendapatkan rekomendasi yang relevan sesuai dengan karakteristik pengguna itu sendiri.

3.2 Metode Pengembangan Perangkat Lunak Waterfall

3.2.1. Requirements Analysis and Definition

Pada tahap ini terdapat beberapa analisa kebutuhan yang berkaitan dengan penelitian ini, diantaranya:

1. Analisa Kebutuhan Data

Pada penelitian ini menggunakan dataset 100.000 dari *movielens* yang berisi 943 pengguna dan 1682 film yang

telah dirating sebanyak 100.000 kali. Variabel yang digunakan dalam penelitian ini yaitu:

- a. *User id*
- b. *Age*
- c. *Gender*
- d. *Occupation*
- e. *Movie id*
- f. *Rating*

2. Analisa Kebutuhan Perangkat Keras (*Hardware*)

Perangkat keras atau *hardware* yang digunakan pada penelitian ini adalah sebagai berikut:

- a. Laptop Acer Aspire E 14 E5-475G-73A3
- b. Processor Intel (R) Core (TM) i7-i7500U CPU @ 2.70GHz (4 CPUs), ~2.9GHz
- c. 8GB RAM
- d. Mouse

3. Analisa Kebutuhan Perangkat Lunak (*Software*)

Perangkat lunak atau *software* yang digunakan pada penelitian ini adalah sebagai berikut:

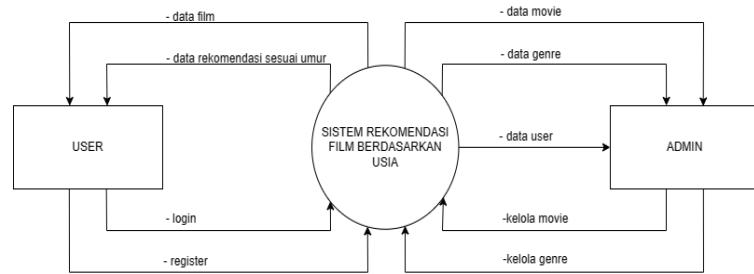
- a. Sistem Operasi Windows 10 Pro 64-bit
- b. PHP
- c. MySQL
- d. XAMPP

3.2.2. System and Software Design

Pada tahap ini dilakukan beberapa tahapan perancangan desain yang sesuai dengan kebutuhan sistem. Penulis menggunakan diagram konteks, *DFD (Data Flow Diagram)* dan *user interface* untuk merancang sistem. Perancangan tersebut dijelaskan sebagai berikut :

1) Diagram Konteks

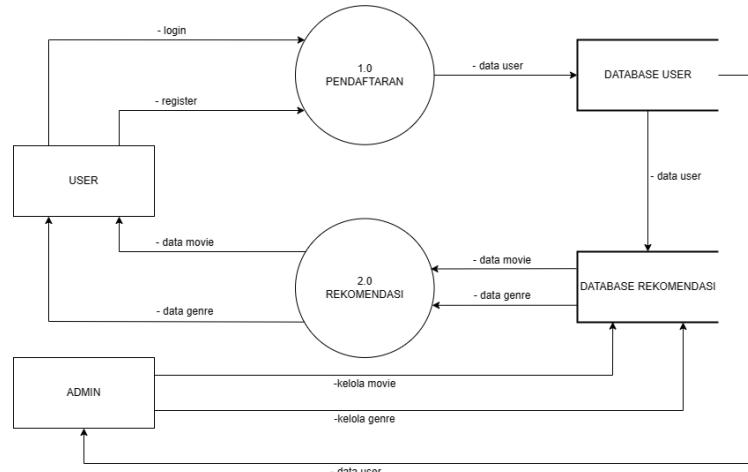
Diagram konteks pada sistem rekomendasi film berdasarkan usia dapat dilihat pada gambar 3.6.



Gambar 3.6 Diagram Konteks

2) DFD (Data Flow Diagram)

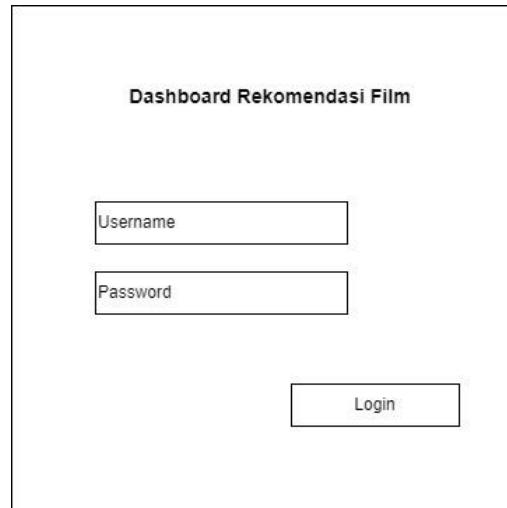
Rancangan DFD pada penelitian ini dapat dilihat pada gambar 3.7.



Gambar 3.7 DFD (Data Flow Diagram)

3) User Interface Login Admin

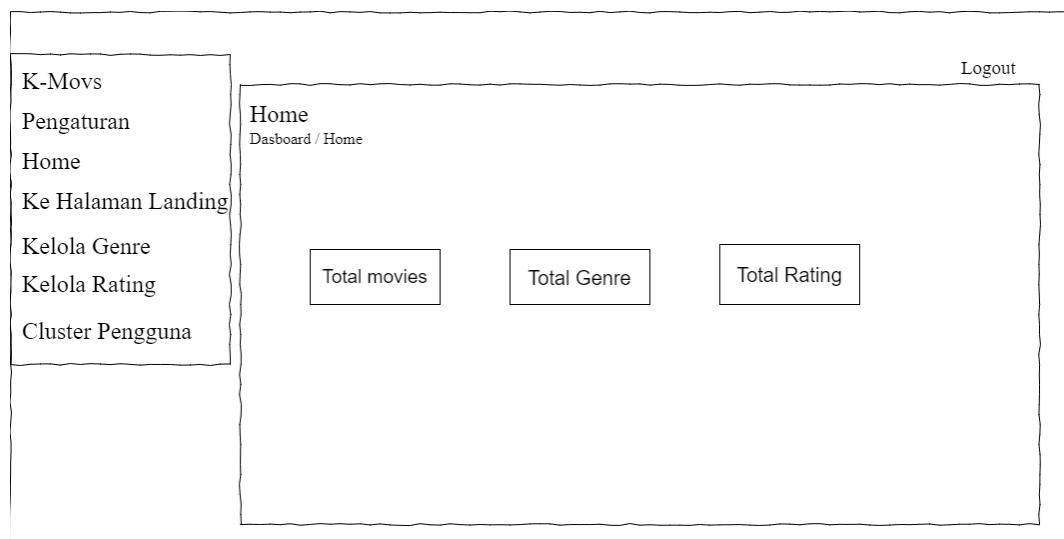
Pada *user interface login admin* menampilkan rancangan tampilan halaman *login admin* yang dapat dilihat pada gambar 3.8.



Gambar 3.8 User Interface Login Admin

4) User Interface Home Admin

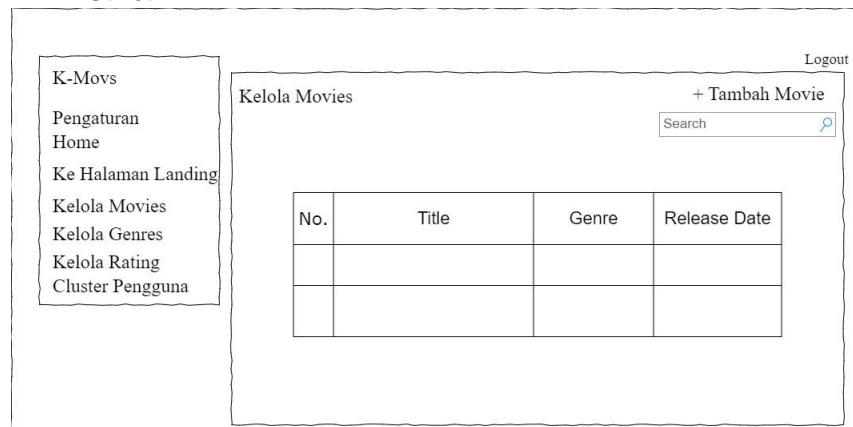
Pada *user interface home admin* menampilkan rancangan tampilan halaman *home admin* yang dapat dilihat pada gambar 3.9.



Gambar 3.9 User Interface Home Admin

5) User Interface Kelola Movies

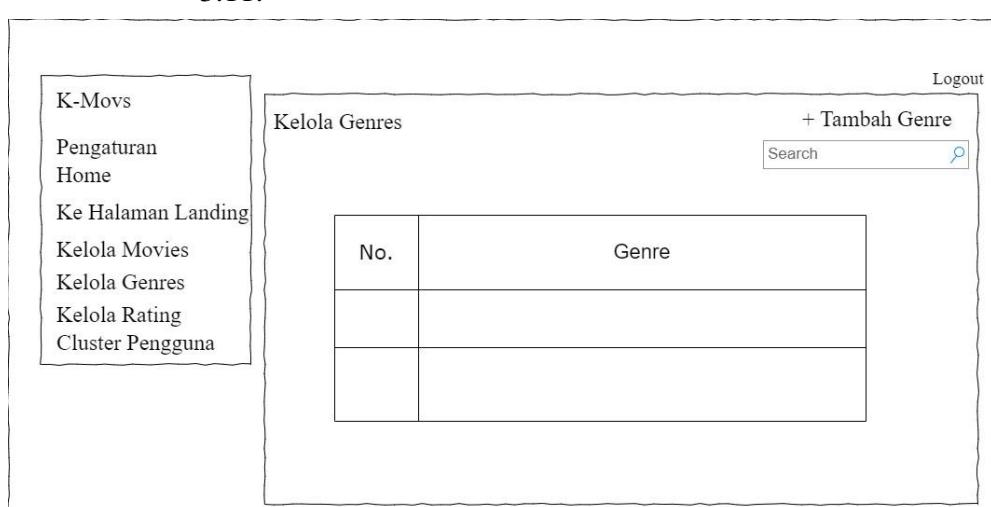
Pada *user interface* kelola movies menampilkan rancangan tampilan halaman kelola movies yang dapat dilihat pada gambar 3.10.



Gambar 3.10 User Interface Kelola Movies

6) User Interface Kelola Genre

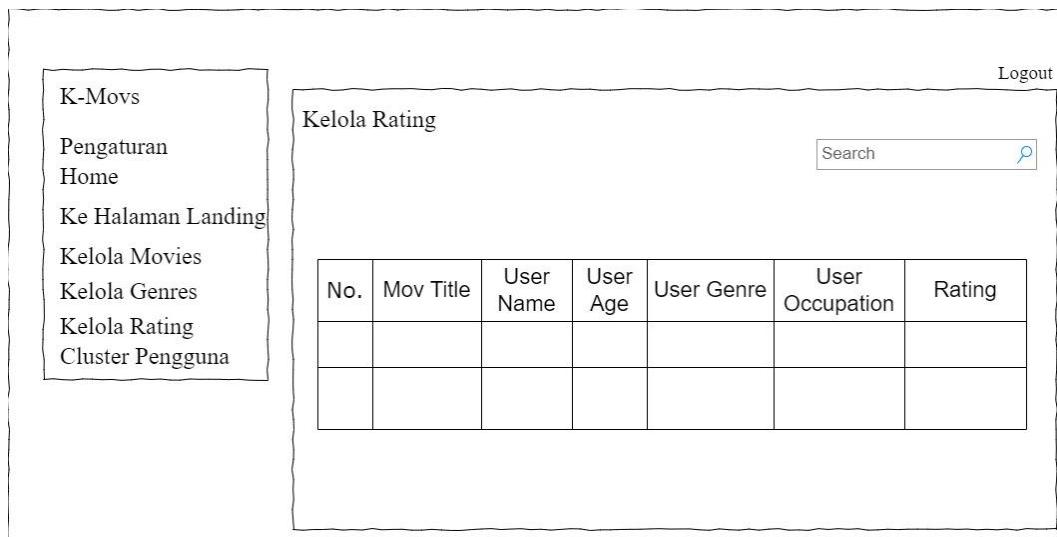
Pada *user interface* kelola genre menampilkan rancangan tampilan halaman kelola genre yang dapat dilihat pada gambar 3.11.



Gambar 3.11 User Interface Kelola Genre

7) *User Interface Kelola Rating*

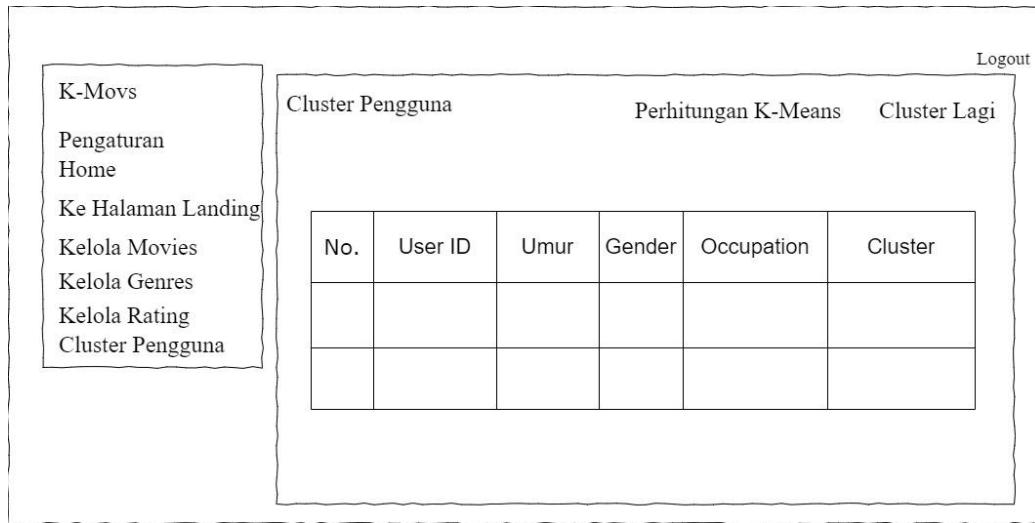
Pada *user interface* kelola rating menampilkan rancangan tampilan halaman kelola rating yang dapat dilihat pada gambar 3.12.



Gambar 3.12 *User Interface Kelola Rating*

8) *User Interface Cluster Pengguna*

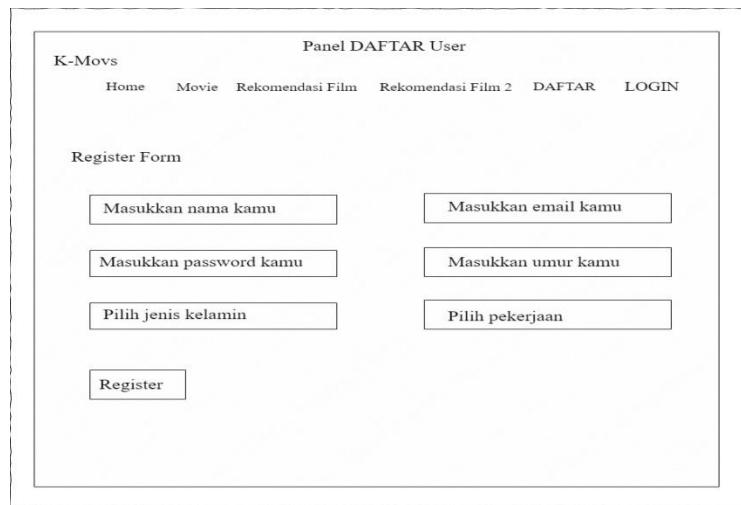
Pada *user interface* cluster pengguna menampilkan rancangan tampilan halaman cluster pengguna yang dapat dilihat pada gambar 3.13.



Gambar 3.13 User Interface Kelola Cluster

9) User Interface Daftar User

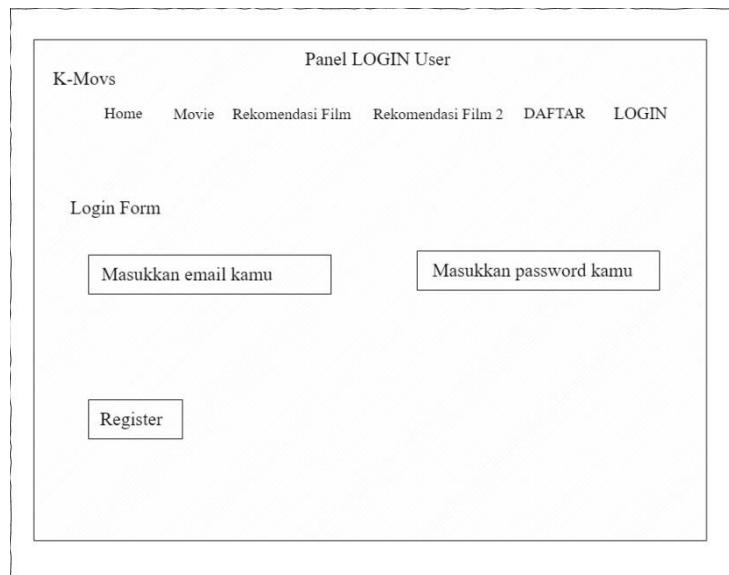
Pada *user interface* daftar *user* menampilkan rancangan tampilan halaman daftar *user* yang dapat dilihat pada gambar 3.14.



Gambar 3.14 User Interface Daftar User

10) User Interface Login User

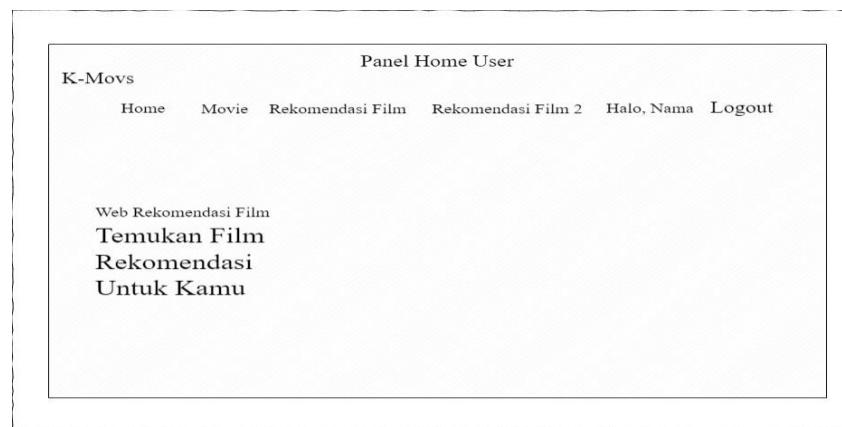
Pada *user interface login user* menampilkan rancangan tampilan halaman *login user* yang dapat dilihat pada gambar 3.15



Gambar 3.15 User Interface Login User

11) User Interface Home

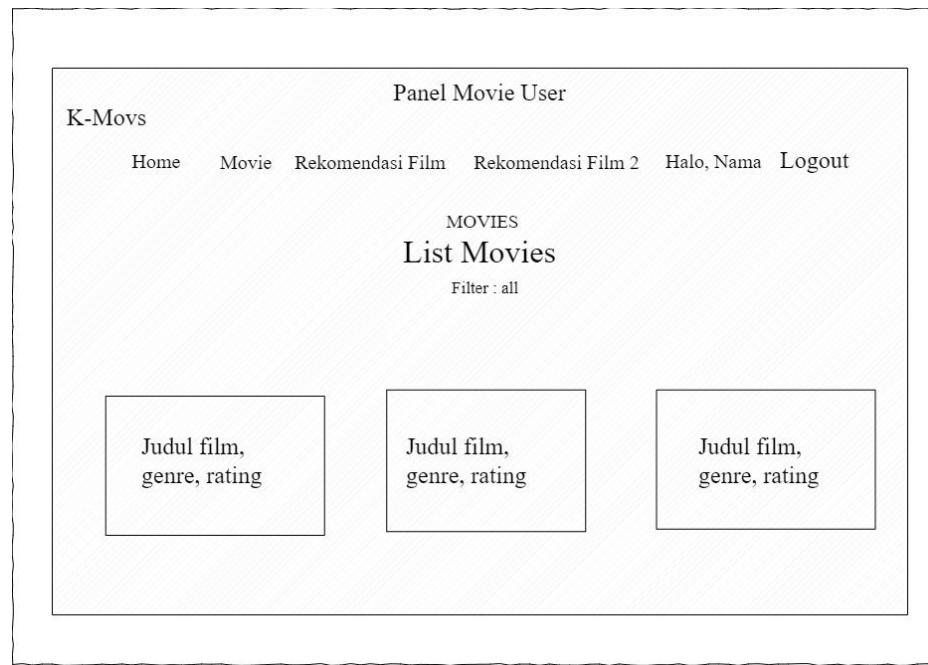
Pada *user interface home* menampilkan rancangan tampilan halaman *home* yang dapat dilihat pada gambar 3.16.



Gambar 3.16 User Interface Home

12) *User Interface Movies*

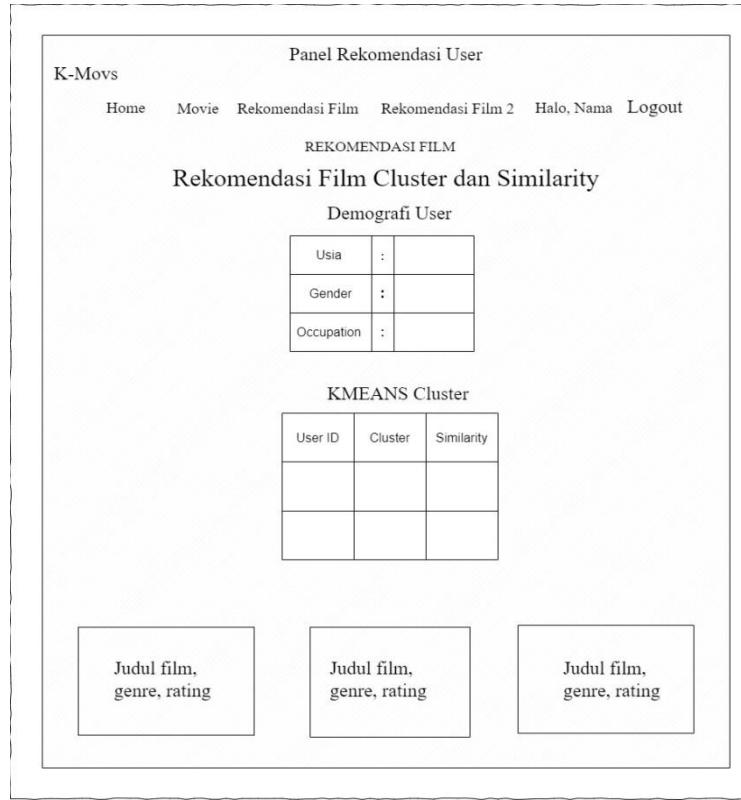
Pada *user interface movies* menampilkan rancangan tampilan halaman *movies* yang dapat dilihat pada gambar 3.17.



Gambar 3.17 *User Interface Movies*

13) *User Interface Rekomendasi Film*

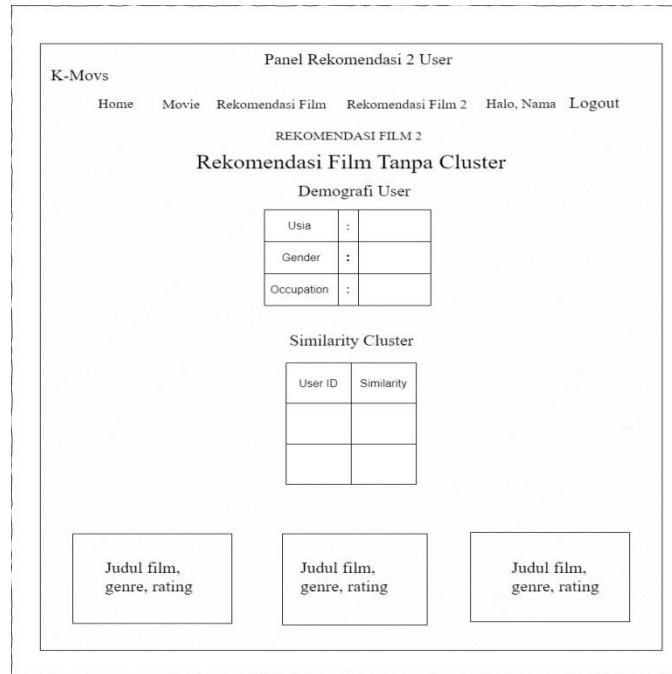
Pada *user interface* rekomendasi film menampilkan rancangan tampilan halaman rekomendasi film yang dapat dilihat pada gambar 3.18.



Gambar 3.18 User Interface Rekomendasi Film

14) User Interface Rekomendasi Film 2

Pada *user interface* rekomendasi film 2 menampilkan rancangan tampilan halaman rekomendasi film 2 yang dapat dilihat pada gambar 3.19.



Gambar 3.19 User Interface Rekomendasi Film 2

3.2.3. *Implementation and Unit Testing*

Pada tahap ini, peneliti akan menerapkan perhitungan ke dalam kode-kode dengan menggunakan bahasa pemrograman.

Menggunakan *cluster* :

```
<h6>Demografi User</h6>
<table border="1">
    <tr>
        <td style="padding:10px">Usia</td>
        <td style="padding:10px">:</td>
        <td style="padding:10px"><?php echo $this->session->userdata('age') ?></td>
    </tr>
    <tr>
        <td style="padding:10px">Gender</td>
        <td style="padding:10px">:</td>
        <td style="padding:10px"><?php echo $this->session->userdata('gender') ?></td>
    </tr>
    <tr>
```

```

        <td style="padding:10px">Occupation</td>
        <td style="padding:10px">:</td>
        <td style="padding:10px"><?php echo $this->session-
>userdata('occname') ?></td>
    </tr>
</table>
<br><br>
<h6>KMEANS Cluster</h6>
<table border="1">
    <tr>
        <td style="padding:10px">User ID</td>
        <td style="padding:10px">Cluster</td>
        <td style="padding:10px">Similarity</td>
    </tr>
<?php
$myage = $this->session->userdata("age");
if($this->session->userdata("gender")=="M"){
    $mygender = 1;
}
else {
    $mygender = 2;
}
$myoccupation_id = $this->session-
userdata("occupation_id");
$iduser = array;
foreach($user as $data) {
    $cluster = $this->kmeans_model->getClusterByUser($data-
>id)->row->cluster;
    $age = $data->age;
    if($data->gender=="M"){
        $gender = 1;
    }
    else {
        $gender = 2;
    }
    $occupation_id = $data->occupation_id;
    $xy_age = (int)$myage * (int)$age;
    $xy_gender = (int)$mygender * (int)$gender;
    $xy_occupation_id = (int)$myoccupation_id *
(int)$occupation_id;
    $sum_xy = (int)$xy_age + (int)$xy_gender +
(int)$xy_occupation_id;
    $sqrt_sum_xy = sqrt($sum_xy);
    $xv2_age = pow($myage,2);
    $xv2_gender = pow($mygender,2);
}

```

```

        $xv2_occupation_id = pow($myoccupation_id,2);
        $sum_xv2 = (int)$xv2_age + (int)$xv2_gender +
(int)$xv2_occupation_id;
        $sqrt_sum_xv2 = number_format((float)sqrt($sum_xv2), 3,
'.', '');
        $yv2_age = pow($age,2);
        $yv2_gender = pow($gender,2);
        $yv2_occupation_id = pow($occupation_id,2);
        $sum_yv2 = (int)$yv2_age + (int)$yv2_gender +
(int)$yv2_occupation_id;
        $sqrt_sum_yv2 = number_format((float)sqrt($sum_yv2), 3,
'.', '');
        $similarity = (int)$sum_xy / ((float)$sqrt_sum_xv2 *
(float)$sqrt_sum_yv2);
        if(number_format((float)$similarity, 3, '.', '')==1.000)
{
    $iduser[] = $data->id;
    if($cluster!="") {
    ?>
    <tr>
        <td style="padding:10px"><?php echo $data->id ?></td>
        <td style="padding:10px"><?php echo $cluster ?></td>
        <td style="padding:10px"><?php echo
number_format((float)$similarity, 3, '.', '') ?></td>
    </tr>
    <?php } } ?>
    </table>
    </center><br><hr>
    <h6><center>Hasil Rekomendasi</center></h6><br>
    <div class="row">
    <?php
        $recommend_movie = $this->movie_model-
>getMovieByUser($iduser)->result;
        foreach($recommend_movie as $data) { ?>
            <div class="col-xl-3 col-lg-4 col-sm-6 grid-item grid-
sizer cat-two">
                <div class="movie-item mb-60">
                    <div class="movie-poster">
                        <a href="#">
alt=""></a>
                    </div>
                    <div class="movie-content">
                        <div class="top">

```

```

        <h5 class="title"><a href="#"><?php
echo $data->title ?></a></h5>
                <span class="date"><?php echo
date("Y",strtotime($data->release_date)) ?></span>
                </div>
                <div class="bottom">
                    <ul>
                        <li><span class="quality"><?php
echo $data->genre_mov ?></span></li>
                        <li>
                            <?php
                                $ratarating = $this->rating_model-
>getAvgRating($data->movie_id)->row->ratarating;
                            ?>
                            <span class="rating"><i
class="fas fa-star"></i> <?php echo number_format($ratarating, 1,
'.', ''); ?></span>
                        </li>
                    </ul>
                </div>
            </div>
        <?php } ?>
    </div>

```

Tanpa *cluster* :

```

<center>
    <h6>Demografi User</h6>
    <table border="1">
        <tr>
            <td style="padding:10px">Usia</td>
            <td style="padding:10px">:</td>
            <td style="padding:10px"><?php echo $this->session-
>userdata('age') ?></td>
        </tr>
        <tr>
            <td style="padding:10px">Gender</td>
            <td style="padding:10px">:</td>
            <td style="padding:10px"><?php echo $this->session-
>userdata('gender') ?></td>
        </tr>

```

```

<tr>
    <td style="padding:10px">Occupation</td>
    <td style="padding:10px">:</td>
    <td style="padding:10px"><?php echo $this->session-
>userdata('occname') ?></td>
</tr>
</table>
<br><br>
<h6>Similarity tanpa cluster</h6>
<table border="1">
    <tr>
        <td style="padding:10px">User ID</td>
        <td style="padding:10px">Similarity</td>
    </tr>
<?php
$myage = $this->session->userdata("age");
if($this->session->userdata("gender")=="M"){
    $mygender = 1;
}
else {
    $mygender = 2;
}
$myoccupation_id = $this->session-
userdata("occupation_id");
$iduser = array;
foreach($user as $data) {
    $cluster = $this->kmeans_model->getClusterByUser($data-
>id)->row->cluster;
    $age = $data->age;
    if($data->gender=="M"){
        $gender = 1;
    }
    else {
        $gender = 2;
    }
    $occupation_id = $data->occupation_id;
    $xy_age = (int)$myage * (int)$age;
    $xy_gender = (int)$mygender * (int)$gender;
    $xy_occupation_id = (int)$myoccupation_id *
(int)$occupation_id;
    $sum_xy = (int)$xy_age + (int)$xy_gender +
(int)$xy_occupation_id;
    $sqrt_sum_xy = sqrt($sum_xy);
    $xv2_age = pow($myage,2);
    $xv2_gender = pow($mygender,2);
}

```

```

        $xv2_occupation_id = pow($myoccupation_id,2);
        $sum_xv2 = (int)$xv2_age + (int)$xv2_gender +
(int)$xv2_occupation_id;
        $sqrt_sum_xv2 = number_format((float)sqrt($sum_xv2), 3,
'.', '');
        $yv2_age = pow($age,2);
        $yv2_gender = pow($gender,2);
        $yv2_occupation_id = pow($occupation_id,2);
        $sum_yv2 = (int)$yv2_age + (int)$yv2_gender +
(int)$yv2_occupation_id;
        $sqrt_sum_yv2 = number_format((float)sqrt($sum_yv2), 3,
'.', '');
        $similarity = (int)$sum_xy / ((float)$sqrt_sum_xv2 *
(float)$sqrt_sum_yv2);
        if(number_format((float)$similarity, 3, '.', '')==1.000)
{
    $iduser[] = $data->id;
    if($cluster!="") {
        ?
        <tr>
            <td style="padding:10px"><?php echo $data->id ?></td>
            <td style="padding:10px"><?php echo
number_format((float)$similarity, 3, '.', '') ?></td>
        </tr>
        <?php } } ?>
    </table>
    </center><br><hr>
    <h6><center>Hasil Rekomendasi</center></h6><br>
    <div class="row">

        <?php
        $recommend_movie = $this->movie_model-
>getMovieByUser($iduser)->result;
        foreach($recommend_movie as $data) { ?>
            <div class="col-xl-3 col-lg-4 col-sm-6 grid-item grid-
sizer cat-two">
                <div class="movie-item mb-60">
                    <div class="movie-poster">
                        <a href="#">
alt=""></a>
                    </div>
                    <div class="movie-content">
                        <div class="top">

```

```

        <h5 class="title"><a href="#"><?php
echo $data->title ?></a></h5>
                <span class="date"><?php echo
date("Y",strtotime($data->release_date)) ?></span>
                </div>
                <div class="bottom">
                    <ul>
                        <li><span class="quality"><?php
echo $data->genre_mov ?></span></li>
                        <li>
                            <?php
                                $ratarating = $this->rating_model-
>getAvgRating($data->movie_id)->row->ratarating; ?>
                                <span class="rating"><i
class="fas fa-star"></i> <?php echo number_format($ratarating, 1,
', ', ''); ?></span>
                            </li>
                        </ul>
                    </div>
                </div>
            <?php } ?>
        </div>
    
```

3.2.4. Integration and System Testing

Pada tahap ini peneliti melakukan penyatuan terhadap beberapa kode program untuk membangun *website* rekomendasi film, dan mengujinya.

```

class Showcase extends CI_Controller
{
    var $navbarTitle;
    public function __construct
    {
        parent::__construct;
        $this->load->model('kmeans_model');
        $this->load->model('movie_model');
        $this->load->model('rating_model');
        $this->load->model('user_model');
        $this->load->helper('waktu_helper');
        $this->load->helper('slug_helper');
    }
}

```

```

        $this->load->helper('kmeans_helper');
    }

    // Showcase Home
    public function index{
        $this->navbarTitle = "Home";
        $data['navbarTitle'] = $this->navbarTitle;
        $data['movies'] = $this->movie_model->get5Movies->result;
        $data['toprated'] = $this->movie_model->topRatedMovies-
>result;
        $this->load->view('showcase/parts/header',$data);
        $this->load->view('showcase/parts/navbar');
        $this->load->view('showcase/home');
        $this->load->view('showcase/parts/footer');
    }

    // Register
    public function register{
        $this->navbarTitle = "Register";
        $data['navbarTitle'] = $this->navbarTitle;
        $data['occupation'] = $this->user_model->getOccupation-
>result;
        $this->load->view('showcase/parts/header',$data);
        $this->load->view('showcase/parts/navbar');
        $this->load->view('showcase/register');
        $this->load->view('showcase/parts/footer');
    }

    public function register_user {
        header('Content-Type: application/json; charset=utf-8');
        $return_array = array;
        $name = $this->input->post('name');
        $age = $this->input->post('age');
        $gender = $this->input->post('gender');
        $occupation_id = $this->input->post('occupation_id');
        $email = $this->input->post('email');
        $password = $this->input->post('password');
        $register = $this->user_model-
>register($name,$age,$gender,$occupation_id,$email,$password);
        if($register) {
            $return_array[] = array("status"=>"OK");
        }
        else {
            $return_array[] = array("status"=>"Failed");
        }
    }
}

```

```

        echo json_encode($return_array);
    }

    public function update_profil {
        header('Content-Type: application/json; charset=utf-8');
        $return_array = array;
        $iduser = $this->session->userdata('iduser');
        $name = $this->input->post('name');
        $age = $this->input->post('age');
        $gender = $this->input->post('gender');
        $email = $this->input->post('email');
        $password = $this->input->post('password');
        $update = $this->user_model-
>updateProfil($iduser,$name,$age,$gender,$email,$password);
        if($update) {
            $session['name'] = $name;
            $session['age'] = $age;
            $session['gender'] = $gender;

            $this->session->set_userdata($session);
            $return_array[] = array("status"=>"OK");
        }
        else {
            $return_array[] = array("status"=>"Failed");
        }
        echo json_encode($return_array);
    }

    // Login
    public function login{
        $this->navbarTitle = "Login";
        $data['navbarTitle'] = $this->navbarTitle;
        $this->load->view('showcase/parts/header',$data);
        $this->load->view('showcase/parts/navbar');
        $this->load->view('showcase/login');
        $this->load->view('showcase/parts/footer');
    }

    // Login
    public function profil{
        $this->navbarTitle = "Profil";
        $data['navbarTitle'] = $this->navbarTitle;
        $this->load->view('showcase/parts/header',$data);
        $this->load->view('showcase/parts/navbar');
        $this->load->view('showcase/profil');
    }
}

```

```

        $this->load->view('showcase/parts/footer');
    }

    // Cosine Similarity with cluster
    public function getRecommendation{
        $iduser = $this->session->userdata('iduser');
        if($iduser != "") {
            $this->navbarTitle = "Rekomendasi Film";
            $data['navbarTitle'] = $this->navbarTitle;
            $data['user'] = $this->user_model->getUser->result;
            $this->load->view('showcase/parts/header',$data);
            $this->load->view('showcase/parts/navbar');
            $this->load->view('showcase/rekomendasi');
            $this->load->view('showcase/parts/footer');
        }
        else {
            redirect(base_url.'showcase/login');
        }
    }

    // Cosine Similarity no cluster
    public function getRecommendation2{
        $iduser = $this->session->userdata('iduser');
        if($iduser != "") {
            $this->navbarTitle = "Rekomendasi Film 2";
            $data['navbarTitle'] = $this->navbarTitle;
            $data['user'] = $this->user_model->getUser->result;
            $this->load->view('showcase/parts/header',$data);
            $this->load->view('showcase/parts/navbar');
            $this->load->view('showcase/rekomendasi2');
            $this->load->view('showcase/parts/footer');
        }
        else {
            redirect(base_url.'showcase/login');
        }
    }

    // kMeans Test
    public function kmeans{
        $this->navbarTitle = "kMeans";
        $data['navbarTitle'] = $this->navbarTitle;
        $data['movies'] = $this->movie_model->getMovies->result;
        $this->load->view('showcase/kmeans',$data);
    }
}

```

```

// Movies
public function movies($key,$page) {
    if(!isset($page)) {
        redirect(base_url.'404');
    }
    else {
        $this->navbarTitle = "Movies";
        $batas = 8;
        $posisi = ($page - 1) * $batas;
        $data[ 'key' ] = $key;
        $data[ 'page' ] = $page;
        $data[ 'next' ] = $page+1;
        $data[ 'prev' ] = $page-1;

        $data[ 'totalmovies' ] = $this->movie_model-
>getMoviesPage($key,"0","10000")->num_rows;
        $data[ 'totalhalaman' ] = ceil($data[ 'totalmovies' ]/$batas);
    // 3

        $data[ 'movies' ] = $this->movie_model-
>getMoviesPage($key,$posisi,$batas)->result;
        $data[ 'navbarTitle' ] = $this->navbarTitle;
        $this->load->view('showcase/partials/header',$data);
        $this->load->view('showcase/partials/navbar');
        $this->load->view('showcase/movies');
        $this->load->view('showcase/partials/footer');
    }
}

}

```

The screenshot displays the k-Movs application's user interface. At the top, there is a navigation bar with links for HOME, MOVIES, REKOMENDASI FILM, REKOMENDASI FILM 2, HALO, EVANI, and LOGOUT. The main content area features a title 'Rekomendasi Film Cluster dan Similarity'. Below the title, there is a section titled 'Demografi User' containing a table with three rows: 'Usia : 22', 'Gender : F', and 'Occupation : Administrator'. Another section titled 'KMEANS Cluster' contains a table with three rows: 'User ID | Cluster | Similarity' followed by two entries: '922 | 1 | 1.000' and '920 | 1 | 1.000'.

Demografi User		
Usia	:	22
Gender	:	F
Occupation	:	Administrator

KMEANS Cluster		
User ID	Cluster	Similarity
922	1	1.000
920	1	1.000

3.2.5. Operation and Maintenance

Pada tahap terakhir akan dilakukan pengoperasian program dan melakukan pemeliharaan. Dalam pengoperasian sistem dilakukan pengujian dengan menggunakan *blackbox testing* untuk mengetahui kemampuan dan keefektifan sistem.