

## BAB V

### PENGUJIAN DAN ANALISIS

#### 5.1 Pengujian

Pada subbab ini akan dilakukan pembahasan mengenai pengujian yang telah dilakukan pada sistem dan hasil evaluasi dari sistem tersebut. Pada pengujian sistem rekomendasi film ini digunakan *dataset* film yang berasal dari *GroupLens*.

#### 5.2 Hasil Uji

Seperti yang telah dibahas pada Bab 3, terdapat 3 macam pengujian yang dilakukan. Pengujian yang pertama dilakukan untuk mengetahui pengaruh jumlah *cluster* terhadap tingkat akurasi hasil prediksi rating. Pengujian yang kedua dilakukan untuk mengetahui pengaruh koefisien  $c$  terhadap tingkat akurasi hasil prediksi rating. Dan pengujian yang ketiga dilakukan untuk mengetahui pengaruh jumlah data latih dan data uji terhadap tingkat akurasi hasil prediksi rating. Pengujian tingkat akurasi prediksi rating dilakukan terhadap *MAE* (*Mean Absolute Error*).

##### 5.2.1 Hasil pengujian dan analisis pengaruh jumlah *cluster* terhadap tingkat akurasi hasil prediksi rating

Pada pengujian pertama yaitu pengaruh jumlah *cluster* (pada proses *clustering*) terhadap tingkat akurasi hasil prediksi rating dilakukan sepuluh pengujian yaitu pengujian pertama dengan jumlah *cluster* 10, pengujian kedua dengan jumlah *cluster* 20, pengujian ketiga dengan jumlah *cluster* 30, pengujian keempat dengan jumlah *cluster* 40, pengujian kelima dengan jumlah *cluster* 50, pengujian keenam dengan jumlah *cluster* 60, pengujian ketujuh dengan jumlah *cluster* 70, pengujian kedelapan dengan jumlah *cluster* 80, pengujian kesembilan dengan jumlah *cluster* 90 dan pengujian kesepuluh dengan jumlah *cluster* 100. Kesepuluh pengujian tersebut ditetapkan nilai koefisien  $c$  sebesar 0,5. Pengujian dilakukan terhadap data latih dan data uji yang sama pada setiap jumlah *cluster* dan terhadap data latih dan data uji yang berbeda pada setiap jumlah *cluster*.

Statistik dari data latih dan data uji yang digunakan pada pengujian pengaruh jumlah *cluster* (pada proses *clustering*) terhadap tingkat akurasi hasil prediksi rating dengan jumlah data latih dan data uji yang sama pada setiap jumlah *cluster* ditampilkan pada Tabel 5.1

**Tabel 5.1** Statistik Data Latih dan Data Uji yang Sama Pada Setiap Jumlah *Cluster* Pada Pengujian Pengaruh Jumlah *Cluster* Terhadap *MAE*

	User	Item	Rating
<b>Data Latih</b>	943	1652	80000
<b>Data Uji</b>	943	1413	20000

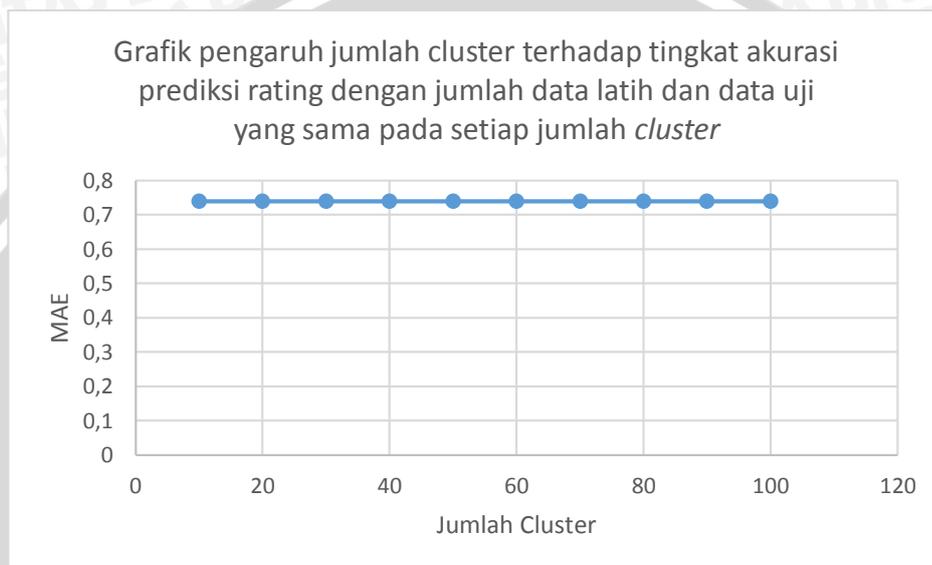
Dari kesepuluh hasil pengujian dengan data latih dan data uji yang sama pada setiap jumlah *cluster*, didapatkan tingkat akurasi atau nilai *MAE* sebesar 0,74; 0,74; 0,74; 0,74; 0,74; 0,74; 0,74; 0,74 dan 0,74.

Hasil nilai *MAE* dari jumlah *cluster* sebanyak 10 sampai dengan 100 dengan koefisien  $c$  sebesar 0,5 serta waktu komputasi yang dihasilkan ditampilkan pada Tabel 5.2.

**Tabel 5.2** Hasil Pengujian Pengaruh Jumlah *Cluster* Terhadap *MAE* dengan Data Latih dan Data Uji yang Sama Pada Setiap Jumlah *Cluster*

Jumlah cluster	Koefisien $c$	Nilai <i>MAE</i>	Waktu Komputasi
10	0,5	0,74	5 jam 03 menit
20	0,5	0,74	5 jam 10 menit
30	0,5	0,74	5 jam 28 menit
40	0,5	0,74	5 jam 53 menit
50	0,5	0,74	6 jam 26 menit
60	0,5	0,74	6 jam 50 menit
70	0,5	0,74	7 jam 23 menit
80	0,5	0,74	7 jam 59 menit
90	0,5	0,74	8 jam 23 menit
100	0,5	0,74	9 jam 03 menit

Hasil pengujian pengaruh jumlah *cluster* (pada proses *clustering*) terhadap tingkat akurasi hasil prediksi rating dengan data latih dan data uji yang sama pada setiap jumlah *cluster* juga dapat dilihat dalam bentuk grafik pada Gambar 5.1. Grafik tersebut menampilkan nilai *MAE* pada koefisien  $c$  0,5 dan mewakili jumlah *cluster* 10, 20, 30, 40, 50, 60, 70, 80, 90 dan 100.



**Gambar 5.1** Grafik Pengaruh Jumlah *Cluster* Terhadap Tingkat Akurasi Prediksi Rating dengan Data Latih dan Data Uji yang Sama Pada Setiap Jumlah *Cluster*

Berdasarkan hasil pengujian pengaruh jumlah *cluster* terhadap tingkat akurasi hasil prediksi rating dengan data latih dan data uji yang sama pada setiap jumlah *cluster*, diketahui bahwa penambahan jumlah *cluster* atau semakin banyak *cluster* maka nilai *MAE* tidak berubah atau tetap. Dari jumlah *cluster* 10 sampai dengan 100 yang diujikan memberikan nilai *MAE* yang tetap yaitu sebesar 0,74. Hal ini dikarenakan pada perhitungan keterkaitan atau peluang antar *item* terhadap *cluster* (*Pro*) dengan data latih dan data uji yang sama pada setiap jumlah *cluster*, nilai dari *counter-similarity* dan *max counter-similarity* juga mengikuti seiring bertambahnya jumlah *cluster* sehingga mendapatkan nilai *Pro* yang hampir sama (perbedaan yang sangat sedikit) dan hasil yang didapatkan pada proses *similarity* dan prediksi rating juga hampir sama. Sedangkan waktu komputasi yang lama dikarenakan pada proses *similarity* terdapat tiga perhitungan *similarity* yaitu *similarity item-rating*,

*similarity group-rating* dan *linier combination of similarity* yang melakukan perhitungan dari tiap item terhadap item yang lain.

Sedangkan statistik dari data latih dan data uji yang digunakan pada pengujian pengaruh jumlah *cluster* (pada proses *clustering*) terhadap tingkat akurasi hasil prediksi rating dengan jumlah data latih dan data uji yang berbeda pada setiap jumlah *cluster* ditampilkan pada Tabel 5.3

**Tabel 5.3** Statistik Data Latih dan Data Uji yang Berbeda Pada Setiap *Cluster* Pada Pengujian Pengaruh Jumlah *Cluster* Terhadap *MAE*

Cluster	Data Latih		Data Uji	
	User	Item	User	Item
10	943	1651	943	1396
20	943	1642	941	1417
30	943	1648	942	1417
40	943	1652	943	1408
50	943	1648	943	1408
60	943	1651	942	1427
70	943	1650	942	1404
80	943	1652	942	1409
90	943	1648	941	1426
100	943	1646	942	1418

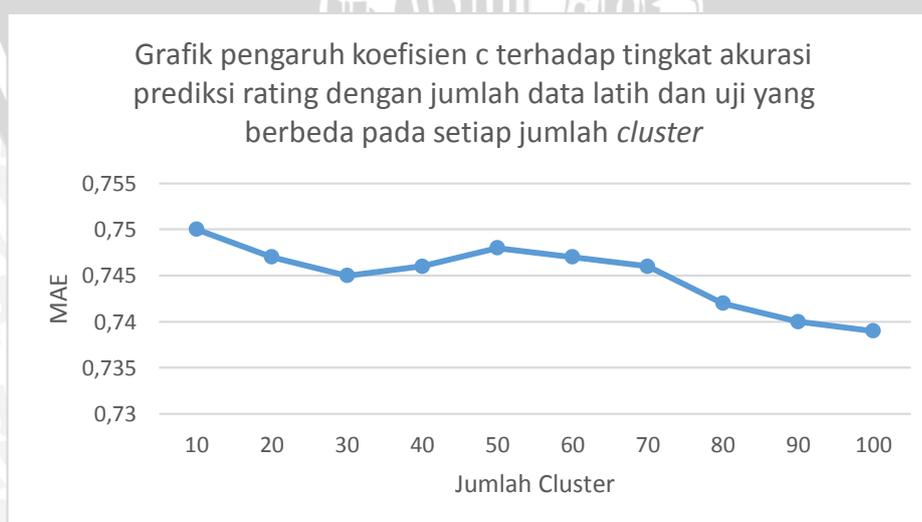
Dari kesepuluh hasil pengujian dengan data latih dan data uji yang berbeda pada setiap jumlah *cluster*, didapatkan tingkat akurasi atau nilai *MAE* sebesar 0,75; 0,747; 0,745; 0,746; 0,748; 0,747; 0,746; 0,742; 0,74 dan 0,739.

Hasil nilai *MAE* dari jumlah *cluster* sebanyak 10 sampai dengan 100 dengan koefisien *c* sebesar 0,5 serta waktu komputasi yang dihasilkan ditampilkan pada Tabel 5.4.

**Tabel 5.4** Hasil Pengujian Pengaruh Jumlah *Cluster* Terhadap *MAE* dengan Data Latih dan Data Uji yang Berbeda Pada Setiap Jumlah *Cluster*

Jumlah cluster	Koefisien $c$	Nilai MAE	Waktu Komputasi
10	0,5	0,75	5 jam 03 menit
20	0,5	0,747	5 jam 10 menit
30	0,5	0,745	5 jam 28 menit
40	0,5	0,746	5 jam 53 menit
50	0,5	0,748	6 jam 26 menit
60	0,5	0,747	6 jam 50 menit
70	0,5	0,746	7 jam 23 menit
80	0,5	0,742	7 jam 59 menit
90	0,5	0,74	8 jam 23 menit
100	0,5	0,739	9 jam 03 menit

Hasil pengujian pengaruh jumlah *cluster* (pada proses *clustering*) terhadap tingkat akurasi hasil prediksi rating dengan data latih dan data uji yang berbeda pada setiap jumlah *cluster* juga dapat dilihat dalam bentuk grafik pada Gambar 5.2. Grafik tersebut menampilkan nilai *MAE* pada koefisien  $c$  0,5 dan mewakili jumlah *cluster* 10, 20, 30, 40, 50, 60, 70, 80, 90 dan 100.



**Gambar 5.2** Grafik Pengaruh Jumlah *Cluster* Terhadap Tingkat Akurasi Prediksi Rating dengan Data Latih dan Data Uji yang Berbeda Pada Setiap Jumlah *Cluster*

Berdasarkan hasil pengujian pengaruh jumlah *cluster* terhadap tingkat akurasi hasil prediksi rating dengan data latih dan data uji yang berbeda pada setiap jumlah *cluster*, diketahui bahwa penambahan jumlah *cluster* atau semakin banyak *cluster* memberikan nilai *MAE* yang cenderung menurun. Nilai *MAE* yang menurun menandakan prediksi rating yang dihasilkan semakin mendekati rating sebenarnya atau hasil prediksi semakin akurat seiring dengan bertambahnya jumlah *cluster*. Nilai *MAE* terendah didapat pada *cluster* dengan jumlah 100 dan memberikan nilai *MAE* sebesar 0,739.

### 5.2.2 Hasil pengujian dan analisis pengaruh koefisien $c$ terhadap tingkat akurasi hasil prediksi rating

Pada pengujian kedua yaitu pengaruh koefisien  $c$  terhadap tingkat akurasi hasil prediksi rating dilakukan sembilan pengujian yaitu pengujian pertama dengan koefisien  $c$  0,1, pengujian kedua dengan koefisien  $c$  0,2, pengujian ketiga dengan koefisien  $c$  0,3, pengujian keempat dengan koefisien  $c$  0,4, pengujian kelima dengan koefisien  $c$  0,5, pengujian keenam dengan koefisien  $c$  0,6, pengujian ketujuh dengan koefisien  $c$  0,7, pengujian kedelapan dengan koefisien  $c$  0,8 dan pengujian kesembilan dengan koefisien  $c$  0,9 . Kesembilan pengujian tersebut ditetapkan jumlah *cluster* sebanyak 10.

Statistik dari data latih dan data uji yang digunakan pada pengujian pengaruh koefisien  $c$  terhadap tingkat akurasi hasil prediksi ditampilkan pada Tabel 5.5

**Tabel 5.5** Statistik Data Latih dan Data Uji Pada Pengujian Pengaruh Koefisien  $c$  Terhadap *MAE*

	User	Item	Rating
<b>Data Latih</b>	943	1652	80000
<b>Data Uji</b>	943	1413	20000

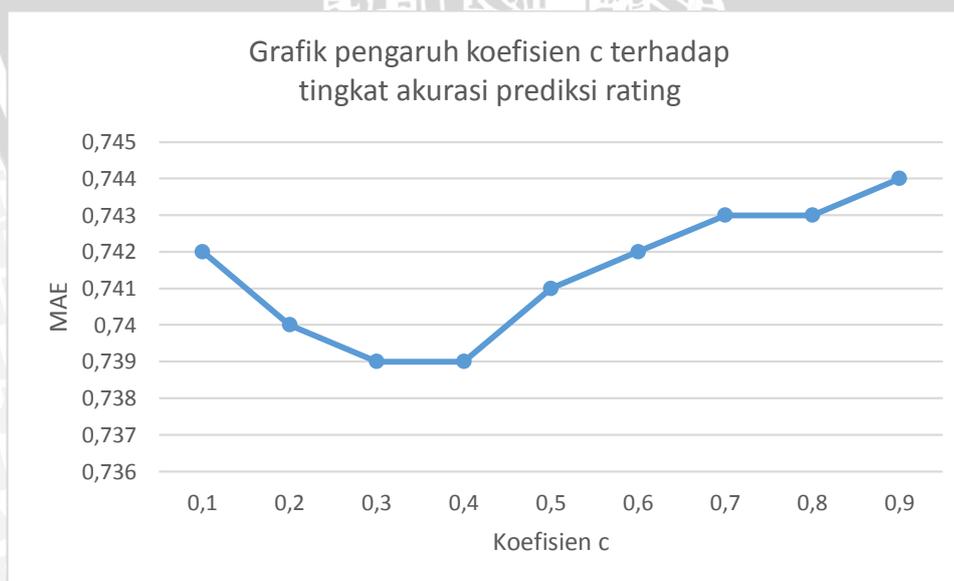
Dari kesembilan hasil pengujian tersebut didapatkan tingkat akurasi atau nilai *MAE* sebesar 0,742; 0,74; 0,739; 0,739; 0,74; 0,742; 0,743; 0,743 dan 0,744.

Hasil nilai *MAE* dengan jumlah *cluster* sebanyak 10 dengan koefisien  $c$  sebesar 0,1 sampai dengan 0,9 serta waktu komputasi yang dihasilkan ditampilkan pada Tabel 5.6.

**Tabel 5.6** Hasil Pengujian Pengaruh Koefisien  $c$  Terhadap *MAE*

Koefisien $c$	Jumlah cluster	Nilai <i>MAE</i>	Waktu Komputasi
0,1	10	0,742	5 jam 03 menit
0,2	10	0,74	5 jam 03 menit
0,3	10	0,739	5 jam 03 menit
0,4	10	0,739	5 jam 03 menit
0,5	10	0,74	5 jam 03 menit
0,6	10	0,742	5 jam 03 menit
0,7	10	0,743	5 jam 03 menit
0,8	10	0,743	5 jam 03 menit
0,9	10	0,744	5 jam 03 menit

Hasil pengujian dapat juga dilihat dalam bentuk grafik pada Gambar 5.3. Grafik tersebut menampilkan nilai *MAE* pada jumlah *cluster* 10 dan mewakili jumlah nilai koefisien  $c$  0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8 dan 0,9.



**Gambar 5.3** Grafik Pengaruh Koefisien  $c$  Terhadap Tingkat Akurasi Prediksi Rating

Berdasarkan hasil pengujian pengaruh koefisien  $c$  terhadap tingkat akurasi hasil prediksi rating, diketahui bahwa penambahan koefisien  $c$  memberikan nilai  $MAE$  yang cenderung meningkat. Nilai  $MAE$  terendah terdapat pada nilai koefisien  $c$  sebesar 0,3 dan 0,4 dengan nilai  $MAE$  0,739. Nilai  $MAE$  meningkat menandakan prediksi rating yang dihasilkan semakin menjauhi rating sebenarnya atau hasil prediksi rating semakin tidak akurat seiring bertambahnya koefisien  $c$ .

### 5.2.3 Hasil pengujian dan analisis pengaruh jumlah data latih dan data uji terhadap tingkat akurasi hasil prediksi rating

Pada pengujian ketiga yaitu pengaruh jumlah data latih dan data uji terhadap tingkat akurasi hasil prediksi rating dilakukan empat pengujian yaitu pengujian pertama dengan data latih 20% data uji 80%, pengujian kedua dengan data latih 40% data uji 60%, pengujian ketiga dengan data latih 60% data uji 40%, dan pengujian keempat dengan data latih 80% data uji 20%. Keempat pengujian tersebut ditetapkan jumlah *cluster* sebanyak 10 dan nilai koefisien  $c$  sebesar 0,5.

Statistik dari data latih dan data uji yang digunakan pada pengujian pengaruh jumlah data latih dan data uji terhadap tingkat akurasi hasil prediksi ditampilkan pada Tabel 5.7

**Tabel 5.7** Statistik Dataset Pengujian Pengaruh Jumlah Data Latih Dan Data Uji Terhadap  $MAE$

Dataset	Data Latih			Data Uji		
	User	Item	Rating	User	Item	Rating
20% data latih 80% data uji	942	1405	20000	942	1419	80000
40% data latih 60% data uji	943	1544	40000	941	1548	60000
60% data latih 20% data uji	943	1609	60000	942	1611	40000
80% data latih 20% data uji	943	1652	80000	943	1655	20000

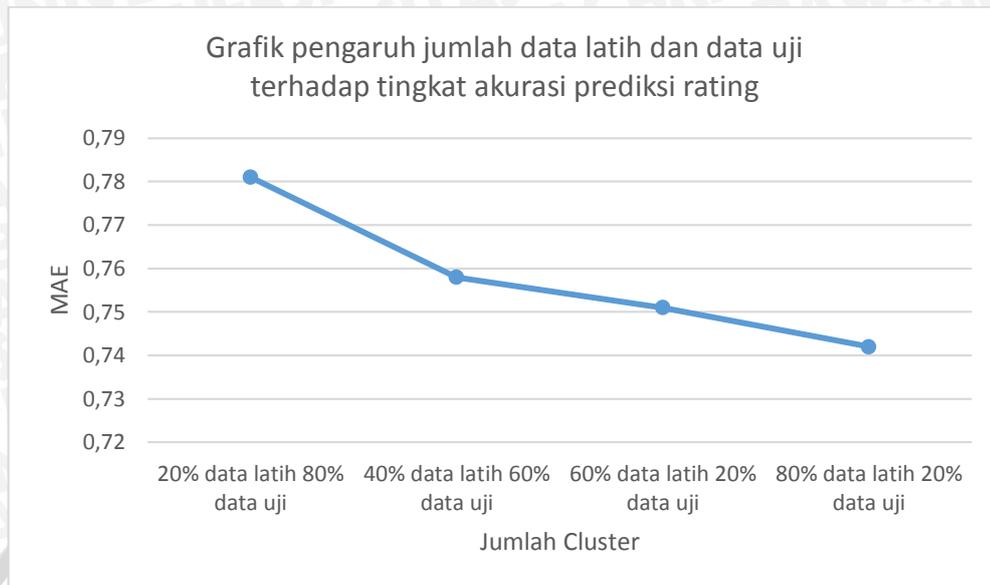
Dari keempat hasil pengujian tersebut didapatkan tingkat akurasi atau nilai *MAE* sebesar 0,781; 0,758; 0,751 dan 0,742

Hasil nilai *MAE* dengan jumlah *cluster* sebanyak 10 dengan koefisien *c* sebesar 0,1 sampai dengan 0,9 serta waktu komputasi yang dihasilkan ditampilkan pada Tabel 5.8.

**Tabel 5.8** Hasil Pengujian Pengaruh Jumlah Data Latih Dan Data Uji Terhadap *MAE*

Dataset	Cluster	Koefisien <i>c</i>	MAE	Waktu Komputasi
20 % data latih, 80 % data uji	10	0,5	0,781	4 jam 03 menit
40 % data latih, 60 % data uji	10	0,5	0,758	4 jam 23 menit
60 % data latih, 20 % data uji	10	0,5	0,751	4 jam 43 menit
80 % data latih, 20 % data uji	10	0,5	0,742	5 jam 03 menit

Hasil pengujian dapat juga dilihat dalam bentuk grafik pada Gambar 5.3. Grafik tersebut menampilkan nilai *MAE* pada jumlah *cluster* 10 dan koefisien *c* 0,5 serta mewakili 20% data latih 80% data uji, 40% data latih 60% data uji, 60% data latih 40% data uji dan 80% data latih 20% data uji.



**Gambar 5.3** Grafik Pengaruh Jumlah Data Latih dan Data Uji Terhadap Tingkat Akurasi Prediksi Rating

Berdasarkan hasil pengujian pengaruh jumlah data latih dan data uji terhadap tingkat akurasi hasil prediksi rating, diketahui bahwa semakin banyak jumlah data latih atau semakin sedikit jumlah data uji memberikan nilai *MAE* yang cenderung menurun. Nilai *MAE* menurun menandakan prediksi rating yang dihasilkan semakin mendekati rating sebenarnya atau hasil prediksi rating semakin akurat seiring bertambahnya data latih. Nilai *MAE* terendah sebesar 0,742 didapatkan pada jumlah 80% data latih 20% data uji.