BAB III

METODOLOGI PENELITIAN

Di dalam bab ini akan dijelaskan metode dan cara kerja yang akan digunakan dalam penelitian, sehingga dapat memberikan gambaran bagaimana membagun rekomendasi dengan genre berdasarkan metode collaborative filtering.

3.1 Diagram Alur Penelitian

Alur penelitian yang digunakan dalam penelitian ini terdiri dari 5 tahap yang digambarkan pada Gambar 3.1 Pada tahapan ini memaparkan kegiatan yang dilakukan pada pengembangan metode Collaborative Filtering dengan penjelasan sebagai berikut :

1. Pengambilan Data Set (MovieLens 100k)

Tahapan ini dilakukan dengan mengumpulkan *dataset movieLens 100k* sebagai bahan evaluasi terhadap metode yang dikembangkan. Dataset yang digunakan pada penelitian ini adalah data Movie yang diakses dari website GroupLens (https://grouplens.org/datasets/movielens/). Dataset yang digunakan adalaah MovieLens 100K.

2. **Pra- prosesing *Dataset***

Bertujuan untuk eliminasi data yang tidak digunakan dalam penelitian ini. Adapun Dataset MoviLens merupakan data yang valid dan terupdate secara berkala, pra-pengelolahan dataset tetap perlu dilakukan untuk meningkatkan performa sistem rekomendasi.

3. Mengatasi Sparsity menggunakan Implementasi Imputasi

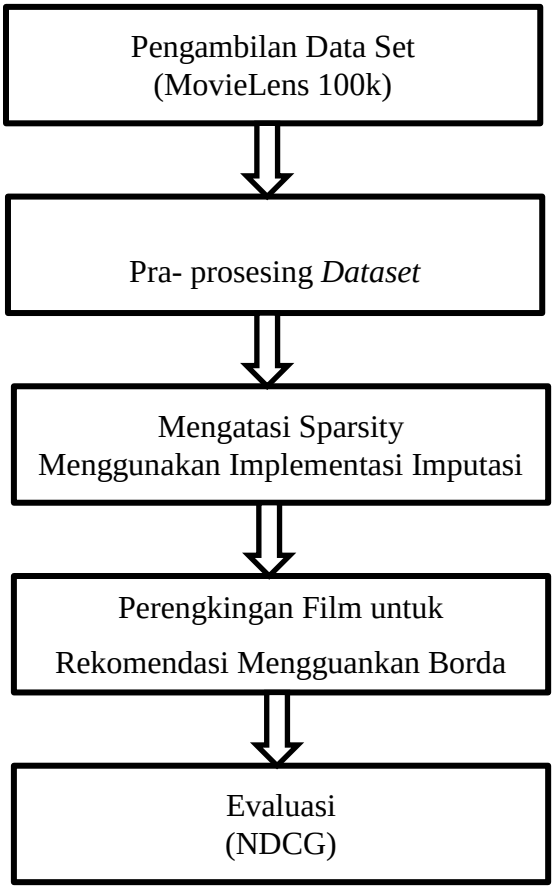
Proses mengatasi Sparsity merupakan kendala yang ada pada sistem rekomendasi dengan metode collaborative filtering karena pada metode collaborative filteringrating merupakan hal yang paling penting dalam pemberian rekomendasi. Hal tersebut Pengunaan Implementasi Imputasi merupakan suatu teknik dalam aljabar linear yang memiliki banyak fungsi dalam pengolahan Spasity dalam dataset.

4. Perengkingan Film untuk rekomendasi menggunakan Borda

Proses perengkingan film ini bertujuan untuk merekomendasikan kepada pengguna tentang produk tersebut sehingga pengguna mendapatkan rekomendasi sesuai preferensi user.

5. Evaluasi (NDCG)

Langkah selanjutnya adalah melakukan evaluasi hasil penelitian yang dilakukan dengan beberapa teknik evaluasi.



Gambar 3.1 Diagram Alur Penelitian

3.2 Alat dan Bahan

1. Pena dan Kertas

Untuk melakukan pencatatan dan keterangan-keterangan yang berkaitan dengan penelitian.

2. Perangkat Keras (hardware)

    a. Laptop Asus, Memori 320GB, Corei3

    b. Flashdisk Sandisk 8GB

    c. Printer Canon IP 2700

    d. Modem Portebel Smart Fren – M5

3. Perangkat Lunak (software)

   a. OS (Windows 7 Ultimate 64-Bit)

   c. Microsoft Word 2007

   d. Microsoft Power Point 2007

   e. Web Browser (Mozilla Firefox)

Penelitian ini menggunakan dataset yang umum dan banyak digunakan dalam sistem rekomedasi collaborative filtering yaitu dataset movie dari MovieLens [17]. Dataset tersebut dapat diakses di website GroupLens (https://grouplens. org/datasets/movielens/). Dataset MovieLens merupakan dataset open source yang dapat digunakan secara bebas namun harus mengikuti aturan yang telah dibuat oleh GroupLens, yang merupakan laboratorium penelitian yang berada di University of Minnesota, USA. Dataset yang digunakan dalam eksperimen ini yaitu MovieLens 100K. Dataset 100K memiliki sekitar 1700 movie, 1000 pengguna dan 100,000 rating. Dataset tersebut memiliki karakteristik umum yaitu informasi demografis pengguna (user id, gender, age, occupation, dan zipcode).

Dataset ini juga memiliki 19 genre seperti pada Tabel 3.1. Selain itu setiap pengguna minimal memberikan rating pada 20 movie. Dataset tersebut mengandung sparsity 95,8% [18], hal tersebut disebabkan karena banyak pengguna tidak memberikan rating ke movie.

Dataset MovieLens 100K ketika pertama kali diunduh yang didapatkan adalah data yang masih terkompresi. Data yang terkompresi tersebut berisi data mentah yang perlu melalui proses pra-pengolahan. Terdapat tiga file utama yaitu Users,

Movies, dan Ratings seperti yang dicontohkan pada Tabel dibawah ini.

a. User

File ini berisi tentang informasi demografis pengguna yang terdapat dalam*dataset*. Terdapat data id-pengguna, jenis kelamin, umur, jenis pekerjaan dan kode pos/alamat pengguna. Id pengguna adalah data yang digunakan untuk mengidentifikasi pengguna. Tabel 5 merupakan contoh sebagian data yang berada pada file *users* dengan susunan *field* yaitu: UserID::Gender::Age::Occupation::Zip- code. Adapun daftar *occupation* dapat dilihat ada Tabel 6.

Tabel 3.1 Contoh Sebagian Data Pengguna

User Id	Gender	Age	Occupation	Zip Code
1	M	24	19	85711
2	F	53	13	94043
3	M	23	20	32067
4	M	24	19	43537
5	F	33	13	15213
6	M	42	6	98101
7	M	57	0	91344
8	M	36	0	05201
9	M	29	18	01002
10	M	52	9	90703

Tabel 3.2 Daftar Occupation

Id	Occupation
0	Administrator
1	Artist
2	Doctor
3	Educator
4	Engineer
5	Entertainment
6	Executive
7	Healthcare
8	Homemaker
9	Lawyer
10	Librarian
11	Marketing
12	None
13	Other
14	Programmer
15	Retired
16	Salesman
17	Scientist
18	Student
19	Technician
20	Writer

b. Movies

File *movies* berisi informasi tentang data produk atau movie. Tabel 3.3 merupakan contoh sebagian data yang berada pada file *movies* dengan susunan *field* yaitu: MovieID::Title::Genres. Adapun daftar genre *movie* dapat dilihat pada Tabel 3.3.

Tabel 3.3 Contoh Daftar Movie

MovieID	Title	Genres
1	Snatch (2000)	Action – Comedy
2	Big hero 6 (2014)	Animasi – adventure – comedy
3	50 first date (2004)	Romance
4	Ratatoville (2007)	Animasi – adventure – comedy
5	Kungfu hustle (2004)	Action – Comedy
6	The orphanage (2007)	Horror
7	Incredibles 2 (2018)	Animasi – adventure – comedy
8	500 days of summer	Romance
9	Zombieland (2009)	Action – Comedy
10	Scream (1996)	Horror

Tabel 3.4 Daftar Genre Movie

Id	**Genre**	**Id**	**Genre**
0	Unknown	10	Film-Noir
1	Action	11	Horror
2	Adventure	12	Musical
3	Animation	13	Mystery
4	Children's	14	Romance
5	Comedy	15	Sci-Fi
6	Crime	16	Thriller
7	Documentary	17	War
8	Drama	18	Western
9	Fantasy		

c. Movies

Files ini berisi 100,000 rating yang diberikan oleh 1000 pengguna ke 1700 movie yang ada. . File *movies* berisi data *rating* dengan urutan *field* yaitu UserID::MovieID::Rating:: Timestamp, seperti pada Tabel 3.5.

Tabel 3.5 File movies

User Id	Movie Id	Rating	Timestamp
1	242	3	881250949
2	302	3	891717742
3	377	1	878887116
4	346	1	886397596
5	474	4	884182806
6	265	2	881171488
7	465	5	891628467
8	451	3	886324817
9	257	2	879372434
10	1014	5	879781125

3.4 Contoh perhitungan menggunakan NDCG

Ditampilkan daftar film yang telah di rating oleh pengguna dengan skala dari 1 – 5. Urutan film tersebut adalah: F1, F2, F3, F4, F5, F6, F7 yang masing-masing telah di rating oleh pengguna yaitu 2, 1, 2, 1, 1, 0, 0. Selanjutnya film diurutkan berdasarkan rating terbesar ke rating terkecil (ranking). Hasil ranking tersebut selanjutnya dievaluasi menggunakan NDCG.

Berdasarkan uruaian tersebut dapat diilustrasikan sebagai berikut:

Tabel 3.6 Dataset Film Ratting

F1	F2	F3	F4	F5	F6	F7
2	1	2	1	1	0	0

→ Dataset film

→ Rating

Data Rating

F1	F2	F3	F4	F5	F6	F7
2	2	1	1	1	0	0

→ Urutan film

→ Rating

Data Rating

Langkah penyelesaian:

a. Menghitung DCG

Tabel 3.7 Menghitung DCG

Film	i	rel_i	Gains $2^{rel_i} - 1$	Posisi Discount $\frac{1}{log_2(i+1)}$	DCG $\sum_{i=1}^{p} \frac{2^{rel_i} - 1}{log_2(i+1)}$
F1	1	2	3	1,00	3,00
F2	2	1	1	0,63	3,63
F3	3	2	3	0,50	5,13
F4	4	1	1	0,43	5,56
F5	5	1	1	0,39	5,95
F6	6	0	0	0,36	5,95
F7	7	0	0	0,33	5,95

b. Menghitung IDCG (Ideal DCG)

Tabel 3.8 Menghitung IDCG

Film	i	rel_i	Gains $2^{rel_i} - 1$	Posisi Discount $\frac{1}{log_2(i+1)}$	IDCG $\sum_{i=1}^{p} \frac{2^{rel_i} - 1}{log_2(i+1)}$
F1	1	2	3	1,00	3,00
F3	2	2	3	0,63	4,89
F2	3	1	1	0,50	5,39
F4	4	1	1	0,43	5,82
F5	5	1	1	0,39	6,21
F6	6	0	0	0,36	6,21
F7	7	0	0	0,33	6,21

a. Menghitung NDCG

Tabel 3.9 Menghitung NDCG

$$NDCG_p = \frac{DCG_p}{IDCG_p}$$

$NDCG_1$	1,00
$NDCG_2$	0,74
$NDCG_3$	0,95
$NDCG_4$	0,96
$NDCG_5$	0,96
$NDCG_6$	0,96
$NDCG_7$	0,96

Berdasarkan hasil perhitungan NDCG menunjukkan peringkat (ranking) yang terbentuk bagus, karena mendekati nilai 1 sehingga relevan untuk direkomendasikan.

Hasil ranking Top-3 yaitu produk B, A, dan C.

Tabel 3.10 Hasil Perhitungan NDCG

Pengguna	Film		
	B	A	C
Pengguna 1	4	5	3
Pengguna 2	4	2	1
Pengguna 3	3	2	5

Berdasarkan data ranking produk tersebut dilakukan proses perhitungan NDCG sebagai berikut:

Perhitungan NDCG Pengguna ke-1

Tabel 3.11 Perhitungan NDCG dan IDCG

i	rel_i	Gains $2^{rel_i} - 1$	Posisi Discount $\frac{1}{log_2(i+1)}$	DCG $\sum_{i=1}^{p} \frac{2^{rel_i}-1}{log_2(i+1)}$
1	4	15	1,00	15,00
2	5	31	0,63	34,56
3	3	7	0,50	38,06

i	rel_i	Gains $2^{rel_i} - 1$	Posisi Discount $\frac{1}{log_2(i+1)}$	IDCG $\sum_{i=1}^{p} \frac{2^{rel_i}-1}{log_2(i+1)}$
2	5	31	0,63	19,56
1	4	15	1,00	34,56
3	3	7	0,50	38,06

$NDCG_1$	0,77
$NDCG_2$	1,00
$NDCG_3$	1,00

Perhitungan NDCG Pengguna ke-2

Tabel 3.12 Perhitungan NDCG dan IDCG

i	rel_i	Gains $2^{rel_i} - 1$	Posisi Discount $\frac{1}{log_2(i+1)}$	DCG $\sum_{i=1}^{p} \frac{2^{rel_i}-1}{log_2(i+1)}$
1	4	15	1,00	15,00
2	2	3	0,63	16,89
3	1	1	0,50	17,39

i	rel_i	Gains $2^{rel_i} - 1$	Posisi Discount $\frac{1}{log_2(i+1)}$	IDCG $\sum_{i=1}^{p} \frac{2^{rel_i}-1}{log_2(i+1)}$
1	4	15	1,00	15,00
2	2	3	0,63	16,89
3	1	1	0,50	17,39

$NDCG_1$	1,00
$NDCG_2$	1,00
$NDCG_3$	1,00

Perhitungan NDCG Pengguna ke-3

Tabel 3.13 Perhitungan NDCG dan IDCG

i	rel_i	Gains $2^{rel_i} - 1$	Posisj Discount $\frac{1}{log_2(i+1)}$	DCG $\sum_{i=1}^{p} \frac{2^{rel_i}-1}{log_2(i+1)}$
1	3	7	1,00	7,00
2	2	3	0,63	8,89
3	5	31	0,50	24,39

i	rel_i	Gains $2^{rel_i} - 1$	Posisi Discount $\frac{1}{log_2(i+1)}$	IDCG $\sum_{i=1}^{p} \frac{2^{rel_i}-1}{log_2(i+1)}$
3	5	31	0,50	15,50
1	3	7	1,00	22,50
2	2	3	0,63	24,39

$NDCG_1$	0,45
$NDCG_2$	0,40
$NDCG_3$	1,00

Dari perhitungan tersebut selanjutnya kita lakukan rata-rata, sehingga diperoleh nilai NDCG sebagai berikut:

rata-rata	$NDCG_1$	0,74
	$NDCG_2$	0,80
	$NDCG_3$	1,00