

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Pengertian Sistem Rekomendasi**

Sistem rekomendasi adalah sistem yang dirancang dengan tujuan untuk membantu pengguna dengan cara memberikan rekomendasi kepada pengguna ketika pengguna dihadapkan dengan jumlah informasi yang besar. Rekomendasi yang diberikan diharapkan dapat membantu pengguna dalam proses pengambilan keputusan, seperti barang apa yang akan dibeli, buku apa yang akan dibaca, atau musik apa yang akan didengar, dan lainnya [8]. Mesin rekomendasi adalah bagian dari sistem rekomendasi yang merupakan core atau otak dari sistem rekomendasi yang terfokus pada penggunaan algoritma dan perhitungan matematika untuk mempelajari profil pengguna, profil item dan interaksi keduanya dalam memprediksi item yang akan direkomendasikan kepada user [9] karena sebuah sistem rekomendasi dirancang untuk memprediksi sekumpulan item yang sesuai dengan preferensi user dimana nantinya item tersebut akan direkomendasikan pada user [10].

#### **2.2 Rekomendasi Film**

Mesin rekomendasi film adalah bagian dari sistem rekomendasi dari sebuah website gudang film atau website ecommerce yang bergerak dibidang usaha penyewaan film dmn sistemnya saat ini telah berubah dari peminjaman ke ijin akses untuk meelihat film secara realtime atau ijin pengunduhan fillm secara temporary.

rekomendasi film juga menggunakan metode dan algoritma tertentu yang digunakan sebagai komputasi untuk menghasilkan rekomendasi yang bersifat personal. Berikut ini beberapa website gudang film dan website penyewaan film yang terkenal.

### **1. Netflix**

Netflix adalah pelopor layanan sewa film online. Didirikan sejak 1997, Netflix mengakomodasi arsip film paling lengkap dengan wilayah pengoperasian terbanyak. Namun, menyesuaikan dengan perkembangan teknologi, Netflix merubah layanan persewaan film menjadi layanan streaming yang memungkinkan pengguna menonton tayangan kesukaan di mana pun, kapan pun, dan hampir lewat medium apa pun (smartphone, smartTV, tablet, PC, dan laptop). Mirip langganan televisi berbayar (cable tv), Netflix bersih dari iklan, penonton tak perlu menunggu jadwal penayangan serial televisi, dan bisa menentukan sendiri konten yang ingin dinikmati. Untuk pembuatan rekomendasi, netflix meminta peilaian pengguna berupa rating 1-5 sebagai feedback atas film yang pernah ditonton pengguna tersebut. Kemudian hasil dari rating tersebut dijadikan masukan untuk pembuatan personal rekomendasi. Khusus untuk memperbaharui mesin rekomendasinya netflix mempekerjakan lebih dari 300 orang karena rekomendasi membawa peningkatan keuntungan Netflix. Kontribusi Netflix dalam bidang rekomendasi adalah netflix mengadakan sayembara untuk meningkatkan mesin rekomendasinya, namun hasil dari sayembara tersebut dipublikasikan secara umum.

Metode dan algoritma yang paling terkenal adalah matriks faktorisasi collaboration filtering dalam memperingkat film berdasarkan rating dan prediksi rating yang dihasilkan algoritma tersebut. Sampai saat ini algoritma tersebut banyak di gunakan dan dikembangkan pada perusahaan ecommerce raksasa di dunia seperti Amazon dan yahoo.

## **2. Rotten Tomatoes**

Rotten Tomatoes adalah situs web yang menyediakan informasi tentang film dari seluruh dunia, termasuk orang-orang yang terlibat di dalamnya mulai dari aktor/aktris, sutradara, penulis sampai penata rias dan musikus. Rotten Tomatoes juga dikenal sebagai review agregator (situs pengumpul review-review film dari situs lainnya) yang juga menyajikan berbagai ulasan terpercaya dari para kritikus film profesional dan penonton film (audience). Para kritikus film ini akan memberikan nilai terhadap sebuah film dalam Tomatometer. Sedangkan untuk para penonton biasa juga bisa menilai sebuah film melalui Audience Score. Ada banyak variabilitas dalam kualitas rekomendasi. Pada akhirnya penilaian dan pemeringkatan film dari Rotten Tomatoes menjadi referensi bagi pengguna dalam memilih film terbaik yang ingin ditontonnya.

## **3. Movielends**

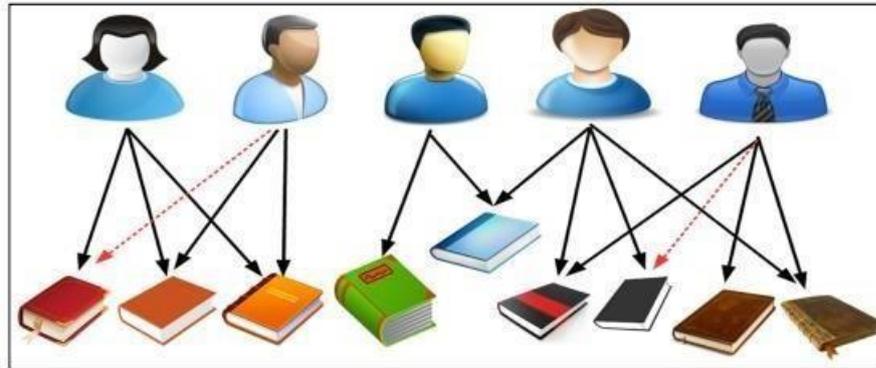
Movielends adalah situs rekomendasi film personal yang disusun berdasarkan rating yang diberikan pengguna terhadap sebuah film. MovieLens dijalankan oleh GroupLens, sebuah laboratorium penelitian di University of Minnesota. Dengan menggunakan MovieLens, pengguna

turut membantu GroupLens mengembangkan alat dan antarmuka eksperimental baru untuk eksplorasi data dan rekomendasi.

MovieLens tidak komersial, dan bebas dari iklanat. MovieLens menerbitkan datasets 16 nya secara Cuma-Cuma untuk pengembangan dan penelitian bidang sistem rekomendasi. Penelitian terpublikasi yang terdiksi menggunakan dataset dari MovieLens diindex dan dijadikan referensi bagi grup riset GroupLens dan masyarakat yang tertarik pada pengembangan dan penelitian sistem rekomendasi.

### **2.3 Collaborative Filtering**

Ide utama dalam collaborative filtering adalah untuk memanfaatkan opini atau penilaian pengguna lain yang ada untuk memprediksi item yang mungkin akan disukai/diminati oleh seorang pengguna [11]. Kualitas rekomendasi yang diberikan dengan menggunakan metode ini sangat bergantung dari penilaian pengguna lain terhadap suatu item. Seperti dikemukakan di bagian penjelasan sistem rekomendasi bahwa penilaian dapat berbentuk explicit maupun implicit dimana Pendekatan implicit, artinya, sistem menyimpan dan mempelajari perilaku pengguna terhadap item, contohnya item apa yang pernah dibeli pengguna, berapa kali pengguna melihat barang tersebut, dsb. Sementara pendekatan explicit, yaitu dengan menanyakan kepada pengguna secara langsung deskripsi item yang bagaimana yang ia sukai/minati contoh keluarannya berupa rating atau kuesioner [12].



**Gambar 2.1 Konsep collaborative filtering**

Dari Gambar 1 terlihat konsep dari collaborative filtering dalam mengolah profil pengguna didapatkan dari pengolahan profil pengguna terhadap ketertarikan terhadap suatu item, dua orang yang mempunyai ketertarikan yang sama dianggap memiliki selera yang sama terhadap suatu item, oleh karena itu item yang kemudian disukai oleh orang tersebut di rekomendasikan kepada orang yang memiliki selera yang sama dengan harapan orang tersebut dapat memiliki ketertarikan mengenai item tersebut sama dengan orang yang memiliki selera yang sama. Terdapat dua tahapan proses yang dilakukan pada teknik collaborative filtering dalam membuat rekomendasi, yaitu :

1. Prediksi

Prediksi opini akan diberikan oleh sistem. Pada tahapan ini dilakukan pelatihan (training, dengan tujuan untuk memperoleh suatu model yang paling optimal dalam artian memperoleh akurasi dengan nilai kesalahan (error) yang paling kecil.

2. Rekomendasi

Memberikan rekomendasi berupa daftar produk dengan nilai prediksi tertinggi yang mungkin akan disukai pengguna. Hal ini sering disebut Top N

recommendation dan personal recommendation. Pada tahap ini setelah mendapatkan model yang paling optimal, model tersebut diujikan pada data pengujian dengan tujuan untuk mendapatkan hasil peringkat sebagai output dari pemilihan data pengujian tersebut. 24 Untuk mengimplementasikan collaborative filtering terdapat dua jenis pendekatan yaitu yang berbasis memori (memory-based collaborative filtering) dan pendekatan berbasis model (model-based collaborative filtering).

## **2.4 Sparsity**

Sparsity merupakan suatu kondisi yang terjadi akibat terdapat banyak kekosongan data rating pada item, sehingga menyebabkan nilai rating pengguna terhadap item cenderung renggang. Hal ini disebabkan karena pengguna cenderung hanya memberikan nilai ratingnya terhadap beberapa item saja. Masalah sparsity merupakan kendala yang ada pada sistem rekomendasi dengan metode collaborative filtering karena pada metode collaborative filtering rating merupakan hal yang paling penting dalam pemberian rekomendasi. Semakin banyak rating yang terkumpul maka semakin baik hasil rekomendasi yang diberikan. Semakin tinggi sparsity data rating maka akan memberikan poor accuracy (akurasi yang rendah) terhadap hasil rekomendasi (Xue, 2009) dalam (Sanjung, 2011).

Jumlah user dan jumlah item yang besar akan sangat membantu proses perekomendasi, namun tidak mungkin semua user akan aktif menggunakan system dan memberikan penilaian pada setiap item.

Terkadang user merasa malas untuk memberikan nilai rating pada item setelah memberikan rating terhadap beberapa item saja. Pada umumnya user hanya melakukan rating

terhadap item-item dalam jumlah yang lebih kecil dari jumlah item yang tersedia di database, sehingga menyebabkan matriks user- item mengalami kekosongan data yang sangat tinggi (Budianto,2012).

## 2.5 Implementasi Imputasi (Imputation)

Imputasi adalah proses pengisian nilai kosong terhadap data menggunakan metode tertentu [9]. Salah satu metode dasar dalam melakukan pengisian data adalah menggunakan mean. Nilai mean menunjukkan kecenderungan besarnya suatu nilai dalam kumpulan data, dan dapat digunakan untuk menggantikan kekosongan suatu nilai pada kumpulan data seperti pada [9]. Penggunaan mean pada Imputation-Boosted NBCF adalah dengan membentuk matriks preferensi yang berisi rating user terhadap genre. Nilai pada matriks preferensi merupakan nilai mean dari kumpulan rating user terhadap genre tertentu, menggunakan persamaan

(1).

$$pref_{u,g} = \frac{\sum_{i \in I_g} r_{ui}}{\|I_g\|}$$

Dimana,

$pref_{u,g}$  adalah nilai *rating* dari *user u* terhadap *genre g*,

$r_{ui}$  adalah *rating* milik *user u* terhadap *item i*,

$I_g$  adalah kumpulan *item* yang memiliki *genre g*.

Nilai imputasi *mean* yang akan digunakan pada pengisian *rating* sebuah *item* adalah nilai *mean* dari *rating user* tersebut terhadap *genre-genre* yang dimiliki *item* tersebut. Perhitungan imputasi *mean* terdapat pada persamaan (2) berikut.

(2)

$$impmean_{u,i} = \frac{\sum_{g \in G_i} pref_{u,g}}{\|G_i\|}$$

Dimana,

$impmean_{u,i}$  adalah nilai imputasi mean dari *user u* terhadap *item i*,

$pref_{u,g}$  adalah *rating* milik *user u* terhadap *genre g*,

$G_i$  adalah kumpulan *genre* yang dimiliki *item i*.

Sebuah metode lain untuk imputasi juga digunakan, yaitu modus. Modus menunjukkan kecenderungan dari kumpulan data yang berbentuk diskrit. Pada dasarnya, bentuk *rating* pada *Collaborative Filtering* bermacam-macam. Untuk dataset Movielens, data yang ada berbentuk numerik yang diskrit, sehingga metode modus juga dapat digunakan untuk imputasi. Terdapat nilai  $impmode_{u,i}$  yang menunjukkan nilai imputasi modus untuk *rating user u* pada *item i*. Nilai tersebut adalah modus dari kumpulan *rating* terhadap *item-item* yang setidaknya memiliki satu buah kesamaan *genre* dengan *item i*.

Selanjutnya, nilai imputasi yang digunakan adalah perpaduan kedua imputasi, menggunakan sebuah parameter bobot berupa  $\alpha$  dengan nilai 0-1. Parameter tersebut menunjukkan besarnya bobot imputasi *mean*, dan berbanding terbalik dengan bobot imputasi modus. Matriks *rating* yang padat (*dense*) akan terbentuk dengan mengisi nilai *rating* yang kosong menggunakan persamaan (3).

$$\underline{impvalue}_{u,i} = \alpha * impmean_{u,i} + (1 - \alpha) * impmode_{u,i} \quad (3)$$

Dimana,

$impvalue_{u,i}$  adalah nilai imputasi terhadap *rating user u* untuk *item i*,

$impmean_{u,i}$  adalah nilai imputasi *mean* dari *user u* terhadap *item i*,

$impmode_{u,i}$  adalah nilai imputasi modus dari *user u* terhadap *item i*.

## 2.6 Metode Borda

Rekomendasi dalam bentuk *ranking* juga dihasilkan oleh metode Borda. Metode tersebut awalnya merupakan metode pemilihan. Dalam pemilihan tersebut, masing- masing pengguna memberikan peringkat kepada semua produk sesuai urutan preferensi [15]. Posisi di daftar preferensi menjadi dasar untuk memberikan poin produk. Poin 1 akan diberikan kepada produk dengan posisi terakhir. Poin 2 diberikan kepada produk dengan posisi sebelum posisi terakhir, dan seterusnya sampai posisi pertama mendapat nilai maksimum yaitu  $n$  poin. Produk selanjutnya diurutkan dari nilai tertinggi ke nilai terendah. Produk dengan nilai tertinggi dinyatakan sebagai pemenang [16].

Berikut ini dijelaskan contoh perhitungan menggunakan metode Borda:

### 1. *User preference profiles* dan Poin

Proses ini dilakukan dengan melihat sejauh mana suatu produk lebih disukai oleh pengguna. Misalnya pada pengguna ke-1 =  $A > B > C > D$ . Hasil *user preference profiles* dapat dilihat pada Tabel 2.1

**Tabel 2.1 *User Preference Profiles dan Poin***

Jumlah pengguna	Poin			
	4	3	2	1
1	A	B	C	D
1	B	D	A	C
1	C	D	B	A

2. Mengalikan jumlah pengguna yang memiliki preferensi sama dengan poinnya. Hasil perkalian dapat dilihat pada Tabel 2.2

**Tabel 2.2 Hasil Perkalian Jumlah Pengguna dengan Poin**

<b>A</b> (1 x 4)	<b>B</b> (1 x 3)	<b>C</b> (1 x 2)	<b>D</b> (1 x 1)
<b>B</b> (1 x 4)	<b>D</b> (1 x 3)	<b>A</b> (1 x 2)	<b>C</b> (1 x 1)
<b>C</b> (1 x 4)	<b>D</b> (1 x 3)	<b>B</b> (1 x 2)	<b>A</b> (1 x 1)

3. Hasil akhir metode Borda diperoleh dari menjumlahkan hasil perkalian antara jumlah pengguna dengan poinnya untuk setiap produk. Total poin dapat dilihat pada Tabel 2.3

**Tabel 2.3 Hasil Borda**

<b>Produk</b>	<b>Menjumlahkan hasil perkalian</b>	<b>Total</b>
A	(4 + 0 + 2 + 1)	= 7
B	(4 + 3 + 2 + 0)	= 9
C	(4 + 0 + 2 + 1)	= 7
D	(0 + 6 + 0 + 1)	= 7

Berdasarkan hasil dari proses Borda nilai terbesar dimiliki oleh film B sehingga saat di *ranking* akan dihasilkan urutan film yaitu B A C D.

## **2.7 Penelitian yang berkaitan**

Dalam tabel 2.4 ini adalah penelitian tentang rekomendasi yang menggunakan metode collaborative filtering dengan dataset yang juga diambil dari dataset yang

telah tersedia di lembaga riset.

**Tabel 2.4 Penelitian yang berkaitan dengan metode collaborative filtering**

No	Judul dan Peneliti	Metode, Algoritma dan datasets	Hasil yang didapatkan
1	A new collaborative filtering metric that improves the behavior of recommender systems. ( J. Bobadilla , F. Serradilla, J. Bernal)	Collaborative filtering – Pearson Correlation Dengan tambahan Jaccard measure dan traditional metric. Menggunakan dataset MovieLens (1M ratings) dan Netflix (10M ratings)	MAE pada datasets MovieLens : pearson correlation 1.9 sementara matriks baru dengan penambahan jaccard measure dan traditional matriks MAE turun menjadi 1.3. MAE pada datasets MovieLens : pearson correlation 0,75 sementara matriks baru dengan penambahan jaccard measure dan traditional matriks MAE naik menjadi 0,95
2	Item-Based Collaborative Filtering Recommendation Algorithms. (Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl)	Collaborative filtering (CF)- Matrix Faktorisasi dengan model-based method. Menggunakan datasets movieLens 100K	Memperbaiki hasil metode memory-based collaboration filtering yang menggunakan algoritma KNN yaitu pada sparsity dan skalabilitas. Dengan matrik faktorisasi prediksi untuk seluruh item yang telah mendapat rating dibuat dan direkomendasikan. Didapatkan MAE 0,75.
3	An Approach for Recommender System by Combining Collaborative Filtering with User Demographics and Items Genres (SaurabhKumar Tiwari and Shailendra Kumar)	Collaborative filtering menggunakan KNN dan menggunakan cosine similarity untuk mengaitkan dengan genre dan demografic (feature content). Menggunakan movieLens	Menghasilkan RMSE 1,18 untuk collaborative filtering menggunakan KNN., namun cold start problem untuk pengguna baru dapat dipecahkan berdasarkan demografik pengguna,

	Shrivastava)	datasets(100K ratings)	sedangkan item cold start dipecahkan dengan item cluster
4	An Improved Switching Hybrid Recommender System Using Naive Bayes Classifier and Collaborative Filtering (Mustansar Ali Ghazanfar and Adam Prügel-Bennett)	Hybrid recommender system dengan menggunakan collaborative filtering (KNN) dan content based filtering (naive bayes classifier) dengan teknik switching. Menggunakan data set movielends (100K ratings) dan Filmtrust (28K ratings)	Hasil dari collaborative filtering didapatkan MAE 1.421 dan hasil dari hybrid didapatkan ROC 0,657. Dapat memecahkan masalah cold start problem pada item baru walaupun MAE dinilai masih Tinggi
5	Large-scale Parallel Collaborative Filtering for the Netflix Prize. (Yunhong Zhou, Dennis Wilkinson, Robert Schreiber and Rong Pan	Memperkenalkan algoritma ALS-WR. Menggunakan netflix dataset (1,4M ratings)	Dapat mengatasi masalah sparsity, scalability dan overfitting dibandingkan dengan algoritma collaboratif filtering lainnya dengan pendekatan tetangga. Menghasilkan RMSE 0,935.
6	Machine Learning at Scale (Ondřej Fiedler)	Membandingkan performansi antara k-NN dan ALS pada mesin pembelajaran Mlib Spark untuk pembuatan rekomendasi film dengan metode collaborative filtering. Menggunakan Netflix dataset (100M ratings)	ALS menghasilkan RMSE yang lebih baik yaitu 0,45 dibandingkan dengan KNN yaitu 1,4 . Untuk training, ALS membutuhkan waktu lebih lama daripada KNN namun runtime untuk menghasilkan rekomendasi ALS lebih cepat dibandingkan dengan KNN.

## 2.8 NDCG (Normalized Discount Cumulative Gain)

Normalized Discount Cumulative Gain (NDCG) merupakan salah satu metode yang banyak digunakan untuk mengevaluasi hasil prediksi rekomendasi pada pendekatan ranking-based collaborative filtering. NDCG berfungsi untuk

mengukur kinerja dari sistem rekomendasi berdasarkan relevansi yang dinilai dari entitas yang direkomendasikan. NDCG digunakan untuk mengetahui kualitas dari ranking yang dihasilkan sistem rekomendasi. Metode NDCG merupakan ukuran evaluasi pada n produk teratas yang berasal dari hasil ranking. Adapun persamaan NDCG dapat dituliskan seperti Persamaan (4) dan Persamaan (5):

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i-1}}{\log_2(i+1)} \quad (4)$$

$$NDCG_p = \frac{DCG_p}{IDCG_p} \quad (5)$$

Persamaan (1) adalah Discounted Cumulative Gain (DCG) yang diperoleh dengan menjumlahkan hasil dari  $\frac{2^{rel_i-1}}{\log_2(i+1)}$  reli adalah nilai relevansi (rating) yang diberikan oleh pengguna pada peringkat ke-i dan p adalah banyaknya data (produk) yang akan dievaluasi. Selanjutnya adalah menghitung NDCG  $\frac{DCG_p}{IDCG_p}$  (Persamaan (2) yang diperoleh dari *Ideal Discounted Cumulative Gain* (IDCG) merupakan data yang sudah diurutkan berdasarkan nilai rating sehingga nilainya ideal. Nilai NDCG berkisar dari 0 sampai 1. Semakin tinggi nilai, semakin baik peringkat.

