

**EFEKTIVITAS METODE**  
***USER-ITEM BASED COLLABORATIVE FILTERING***  
**PADA SISTEM PEREKOMENDASI RENTAL VCD**  
**DENGAN *RATING* IMPLISIT**

**SKRIPSI**

**Diajukan untuk memenuhi sebagian persyaratan mendapatkan gelar  
Strata Satu Jurusan Teknik Informatika**



**Disusun Oleh:**

**RETNO LAILA ARFIANI**

**M0508065**

**JURUSAN INFORMATIKA**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS SEBELAS MARET SURAKARTA**

**Desember, 2012**

*commit to user*

**SKRIPSI**  
**EFEKTIVITAS METODE**  
***USER-ITEM BASED COLLABORATIVE FILTERING***  
**PADA SISTEM PEREKOMENDASI RENTAL VCD**  
**DENGAN *RATING* IMPLISIT**

**Disusun Oleh:**


**RETNO LAILA ARFIANI**

**M0508065**


**Skripsi ini telah disetujui untuk dipertahankan di hadapan Dewan Penguji  
pada tanggal: 28 November 2012**

**Dosen Pembimbing 1**

**Dosen Pembimbing 2**



**Ristu Saptono, S.SI, M.T.**  
**NIP. 19790210 200212 1 001**



**Sari Widya Sihwi, S.Kom., M.TI.**  
**NIP. 19830412 200912 2 003**

*commit to user*

## SKRIPSI

**EFEKTIVITAS METODE**  
***USER-ITEM BASED COLLABORATIVE FILTERING***  
**PADA SISTEM PEREKOMENDASI RENTAL VCD**  
**DENGAN *RATING* IMPLISIT**

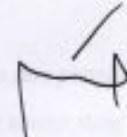
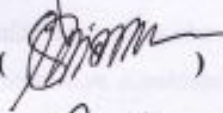
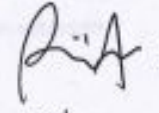
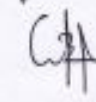
Disusun Oleh:

**RETNO LAILA ARFIANI**

**M0508065**

telah dipertahankan di hadapan Dewan Penguji  
pada tanggal : 11 Desember 2012

**Susunan Dewan Penguji**

- |  |              |   |
|--|--------------|---|
| 1. <u>Ristu Saptono, S.Si., M.T.</u>         | (Ketua)      | (  )  |
| NIP. 19790210 200212 1 001                   |              |   |
| 2. <u>Sari Widya Sihwi., S.Kom., M.Tl.</u>   | (Sekretaris) | (  ) |
| NIP. 19830412 200912 2 003                   |              |   |
| 3. <u>Rini Anggrainingsih, S.T., M.T.</u>    | (Anggota)    | (  ) |
| NIP. 19780909 200812 2 002                   |              |   |
| 4. <u>Meiyanto Eko Sulisty, S.T., M.Eng.</u> | (Anggota)    | (  ) |
| NIP. 19770513 200912 1 004                   |              |   |

Disahkan oleh

  
**Dekan FMIPA UNS**

**Prof. Ir. Ari Handono Ramelan, M.Sc.(Hons), Ph.D**

**NIP. 19610223 198601 1 001**

**Ketua Jurusan Informatika**

  
**Umi Salamah, S.Si., M.Kom.**

**NIP. 19700217 199702 2 001**

## MOTTO

*“Karena sesungguhnya sesudah kesulitan itu ada kemudahan, Sesungguhnya sesudah kesulitan itu ada kemudahan”*

(Q.S Al-Insyirah : 5-6)

*"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya..."*

(Q.S Al-Baqarah : 286)

*“Otak yang biasa – biasa saja dapat diperkuat dengan ilmu dan pengalaman. Usaha yang sungguh-sungguh dan sabar akan mengalahkan usaha yang biasa-biasa saja. Kalau bersungguh- sungguh akan berhasil, kalau tidak serius akan gagal. Kombinasi sungguh-sungguh dan sabar adalah keberhasilan. Kombinasi man jadda wajada dan man shabara zhafira adalah kesuksesan.”*

(Ranah 3 Warna)

*commit to user*

## PERSEMBAHAN

Karya ini Penulis persembahkan kepada:

*”My Beloved Family, Ibu, Bapak dan Adikku yang dalam setiap doanya ada namaku disebut, yang selalu ada disampingku, menyemangatiku dan mendukungku meski seandainya seluruh dunia memojokkanku”*

*”Kawan-kawan seperjuangan IF-08 khususnya sahabat-sahabat dekatku yang sudah kuanggap saudaraku sendiri yang tak pernah lelah mendengarkan keluhanku, mendoakan dan memberi semangat, bahkan membantuku bangkit kembali ketika aku jatuh, terimakasih tanpa kalian semua aku tak mungkin bisa sampai tahap ini, Semoga Allah selalu menjaga tali persaudaraan kita”*

*commit to user*

**EFEKTIVITAS METODE**  
***USER-ITEM BASED COLLABORATIVE FILTERING***  
**PADA SISTEM PEREKOMENDASI RENTAL VCD**  
**DENGAN RATING IMPLISIT**  
**RETNO LAILA ARFIANI**

Jurusan Informatika. Fakultas MIPA. Universitas Sebelas Maret.

**ABSTRAK**

Prinsip kerja dari algoritma *Collaborative Filtering* adalah memberikan rekomendasi atau prediksi *item* berdasarkan pada opini pengguna-pengguna yang mempunyai kemiripan. Opini yang diberikan bisa bersifat eksplisit maupun implisit. Pada opini yang bersifat implisit sistem secara otomatis memberikan *rating* berdasarkan *behaviour user* terhadap sistem sedangkan untuk kasus eksplisit membutuhkan tindakan dari *user* untuk melakukan *rating* pada suatu *item*. Pada penelitian ini penulis mencoba untuk menerapkan metode *User-Item Based Collaborative Filtering* pada sistem rekomendasi rental VCD menggunakan *rating* implisit dimana *rating user* diwakili oleh “0” bila *user* belum menyewa dan “1” bila *user* sudah menyewa. Pada penelitian ini juga dilibatkan unsur *content* untuk melakukan *cluster* pada matriks *user-item* berdasarkan *genre* yang sudah pernah disewa oleh *user*.

Berdasarkan hasil pengujian yang telah dilakukan, secara akurasi penggunaan *cluster genre* pada metode *User-Item Based Collaborative Filtering* menunjukkan hasil yang lebih baik ditunjukkan dengan penurunan nilai rata-rata *error* probabilitas sebesar 16,3% (*absolute MVA*) dan 12,7% (*MVA*). Sedangkan secara presisi hasil paling baik ditunjukkan pada training set 50% dengan nilai standar deviasi sebesar 0,01359. Penggunaan rumus prediksi *absolute MVA* tidak menunjukkan penurunan nilai kesalahan yang signifikan sehingga tidak cukup membantu meningkatkan efektivitas metode *User-Item Based Collaborative Filtering*.

Kata kunci: Genre, Implisit, *Missing Value Algorithm*, Rating, Rekomendasi, *User-Item Based Collaborative Filtering* .  
*commit to user*



**THE EFFECTIVENESS OF USER-ITEM BASED COLLABORATIVE  
FILTERING METHOD ON RECOMMENDER SYSTEM FOR FILM  
SELECTION AT VCD RENTAL USING IMPLICIT RATING**

**RETNO LAILA ARFIANI**

Department of Informatic, Mathematic and Science Faculty, Sebelas Maret  
University

**ABSTRACT**

The working principle of Collaborative Filtering algorithm is to make recommendations or predictions based on the subjective item of users that have similarities. Opinions can be given explicitly or implicitly. On implicit opinions, system automatically assigns ratings based on user behavior of the system while explicit opinions require action from the user to give a rating on an item. In this research, the authors try to apply the methods of User-Item Based Collaborative Filtering for VCD rental recommender system using implicit rating which the user's rating is represented by "0" if user has not rented an item and "1" if user has rented. In this research also included elements of content to perform cluster on user-item rating matrix based on genre that has been rented by user.

Based on the results of testing that has been done, the accuracy of the User-Item Based Collaborative Filtering method using cluster genre shows better results shown by the decrease in the value of the average error probability of 16.3% (absolute MVA) and 12.7% (MVA ). While the best precise outcome is shown in the training set 50% with a standard deviation of 0,01359. The use of absolute MVA prediction formula showed no significant errors reduction so it is not sufficient to help improve the effectiveness of User-Item Based Collaborative Filtering methods.

**Keywords:** Genre, Implicit, Missing Value Algorithm, Ratings, Recommendations, User-Item Based Collaborative Filtering.

*commit to user*

## KATA PENGANTAR

### *Bismillahirrahmaanirrahiim*

Puji syukur penulis panjatkan kehadirat Allah Subhanahu Wa Ta'ala yang senantiasa memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **"Efektivitas Metode *User-Item Based Collaborative Filtering* pada Sistem Rekomendasi Rental VCD dengan *Rating Implisit*"**, yang menjadi salah satu syarat wajib untuk memperoleh gelar Sarjana Informatika di Universitas Sebelas Maret (UNS) Surakarta.

Penulis menyadari bahwa dalam penyusunan laporan ini telah melibatkan banyak pihak, oleh karena itu dengan segala kerendahan hati penulis sampaikan rasa terima kasih kepada yang terhormat :

1. Bapak Ristu Saptono, S.Si., M.T. selaku Dosen Pembimbing I yang penuh kesabaran membimbing, mengarahkan, dan memberi motivasi kepada penulis selama proses penyusunan skripsi ini,
2. Ibu Sari Widya Sihwi, S.Kom., M.Ti. selaku Dosen Pembimbing II yang penuh kesabaran membimbing, mengarahkan, dan memberi motivasi kepada penulis selama proses penyusunan skripsi ini,
3. Ibu Umi Salamah, S.Si., M.Kom. selaku Ketua Jurusan Informatika FMIPA UNS atas saran, dukungan dan bimbingannya selama ini.
4. Bapak Wiharto selaku Pembimbing Akademik yang telah banyak memberi bimbingan dan pengarahan selama penulis menempuh studi di Jurusan Informatika FMIPA UNS,
5. Bapak dan Ibu dosen serta karyawan di Jurusan Informatika FMIPA UNS yang telah memberikan ilmu pengetahuan kepada penulis selama masa studi.
6. Bapak, Ibu dan Adikku serta teman-teman IF-08 yang telah memberikan dukungan doa, semangat dan bantuan sehingga penyusunan skripsi ini dapat terselesaikan,

*commit to user*



7. Serta semua pihak yang telah membantu dalam proses penyusunan skripsi ini.

Penulis berharap semoga skripsi ini dapat bermanfaat bagi semua pihak yang berkepentingan.

Surakarta, November 2012

Penulis



## DAFTAR ISI

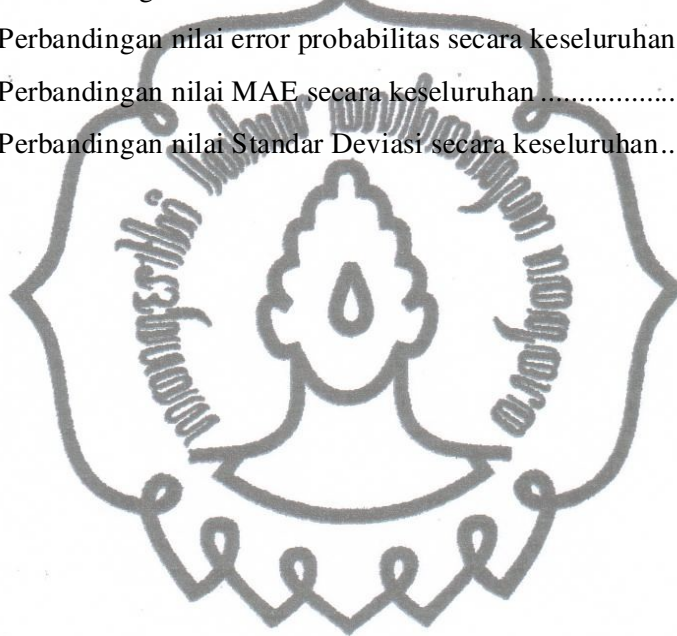
HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN MOTTO .....	iv
HALAMAN PERSEMBAHAN .....	v
ABSTRAK .....	vi
ABSTRACT .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Tugas Akhir .....	4
1.6 Sistematika Penulisan .....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Dasar Teori.....	6
2.1.1 Pengertian Sistem Rekomendasi.....	6
2.1.2 <i>Profile Generation dan Maintenance</i> .....	6
2.1.2.1 <i>Profile Representation</i> .....	6
2.1.2.2 <i>Initial Profile Generation</i> .....	8
2.1.2.3 <i>Profile Learning</i> .....	9
2.1.3 <i>Collaborative Filtering</i> .....	11
2.1.4 Probabilitas Untuk Tipe Data Biner .....	14

*commit to user*

2.1.6 User-Item Based Collaborative Filtering.....	14
2.1.7 Evaluasi Sistem .....	18
2.2 Penelitian Terkait .....	19
2.3 Rencana Penelitian .....	21
BAB III METODOLOGI PENELITIAN.....	22
3.1 Pengumpulan Data .....	22
3.1.1 Tempat dan Waktu Penelitian .....	22
3.1.2 Metode Pengumpulan Data .....	22
3.2 Pemodelan Masalah .....	23
3.2.1 Pembentukan Profil Pengguna .....	23
3.2.2 Perhitungan Nilai Prediksi ( <i>Prediction Generation</i> ).....	24
3.2.3 Rekomendasi .....	25
3.2.4 Contoh Perhitungan .....	25
3.3 Tahap Implementasi.....	34
3.4 Pengujian Metode.....	34
BAB IV PEMBAHASAN.....	38
4.1 Hasil Pembahasan .....	38
4.2 Pengujian Efektivitas Metode .....	39
4.2.1 Pengujian Secara Akurasi.....	40
4.2.2 Pengujian Secara Presisi.....	43
4.2.3 Analisis Hasil Pengujian .....	44
BAB V PENUTUP.....	47
5.1 Kesimpulan .....	47
5.2 Saran .....	48
DAFTAR PUSTAKA .....	50

**DAFTAR TABEL**

Tabel 3.1 Representasi Matrik Rating.....	25
Tabel 3.2 Representasi Matrik Genre Action.....	29
Tabel 3.3 Representasi Matrik Genre Drama.....	29
Tabel 3.4 Tabel Rating dan Prediksi MVA dan <i>Absolute</i> MVA.....	33
Tabel 4.1 Perbandingan nilai error probabilitas secara keseluruhan .....	40
Tabel 4.2 Perbandingan nilai MAE secara keseluruhan .....	42
Tabel 4.3 Perbandingan nilai Standar Deviasi secara keseluruhan.....	43



## DAFTAR GAMBAR

Gambar 2.1 Proses <i>Collaborative Filtering</i> .....	11
Gambar 2.2 Pembentukan <i>User-Neighborhood</i> .....	12
Gambar 2.3 Pembentukan <i>Item-Neighborhood</i> .....	13
Gambar 2.4 Proses perhitungan nilai prediksi <i>rating</i> .....	15
Gambar 2.5 Proses <i>User-Item Based Collaborative Filtering</i> .....	16
Gambar 3.1 Alur Penelitian.....	22
Gambar 3.2 Langkah kerja metode <i>User-Item Based Collaborative Filtering</i> .....	23
Gambar 4.1 Top-5 Rekomendasi pada training set 90% tanpa <i>cluster genre</i> .....	38
Gambar 4.2 Top-5 Rekomendasi pada training set 90% dengan <i>cluster genre</i> .	39
Gambar 4.3 Grafik perbandingan nilai <i>error</i> probabilitas secara keseluruhan.....	41
Gambar 4.4 Grafik perbandingan nilai MAE secara keseluruhan .....	42

## DAFTAR LAMPIRAN

LAMPIRAN A .....	52
LAMPIRAN B .....	59





## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Menurut Sarwar *et al.* (2001), *Collaborative Filtering* bekerja dengan membangun sebuah database dari preferensi *item-item* oleh *user*. Ide dasar dari *Collaborative Filtering* adalah menyediakan prediksi dan rekomendasi *item* berdasarkan opini *user* lain. Opini *user* dapat diperoleh secara eksplisit dari *user* atau dengan menggunakan beberapa pengukuran implisit.

Pendekatan *Collaborative Filtering* pada dasarnya dibagi menjadi dua kategori utama yaitu: *User Based Collaborative Filtering* (disebut juga *memory-based*), dan *Item Based Collaborative Filtering* (disebut juga *model-based*). Metode *memory based* atau *user based* memanfaatkan riwayat tentang pilihan pengguna. Pendekatan *user based* ini mempunyai kelebihan yaitu mampu menghasilkan rekomendasi berkualitas tinggi, namun mempunyai kelemahan yaitu kompleksitas perhitungan bertambah seiring dengan bertambahnya pengguna dan *item* (Sarwar *et al.*, 2001).

Untuk metode *model based* atau *item based* menggunakan kemiripan riwayat nilai *rating item*. Metode ini mengasumsikan kesamaan antara dua *item* berpusat pada satu nilai sehingga mudah menemukan *neighbor* pada sistem *offline*. Karena inilah metode *model based* atau *item based* ini mampu menghasilkan rekomendasi dengan cepat namun kualitas rekomendasi yang dihasilkan rendah (Sarwar *et al.*, 2001).

Menurut Saptono (2010), pada umumnya *Collaborative Filtering* belum memiliki fitur untuk mengakomodasi *item* baru. Hal ini dapat dipahami karena sebuah *item* baru yang belum pernah mendapatkan *rating* sebelumnya tidak dapat dihitung *similarity* atau kesamaannya dengan *item* lainnya. Oleh karena itu digunakan metode alternatif yang dapat menangani kelemahan dari dua metode utama di atas yang disebut dengan metode *User-Item Based Collaborative*

*commit to user*

*Filtering*. Metode ini dapat menghasilkan rekomendasi tanpa mencari formasi ketetanggaan karena menggunakan algoritma *missing value* untuk menghitung prediksi *rating* dan mampu merekomendasikan *item* baru pada sistem. Berdasarkan penelitian sebelumnya oleh Saptono (2006) hasil pengujian menunjukkan bahwa metode *User-Item Based Collaborative Filtering* juga memberikan performansi yang lebih baik dibandingkan dengan metode lain yaitu metode *item based Collaborative Filtering* dengan algoritma *adjusted cosine* dan *Pearson correlation*.

Karakter utama dari pendekatan *User-Item Based Collaborative Filtering* adalah menggunakan vektor kolom dan vektor baris pada *user-item ratings matrix* sebagai dasar perhitungan untuk memperoleh nilai prediksi *rating*. *User-Item Based Collaborative Filtering* pada tulisan ini dianalogikan sebagai sebuah *experimental design*. *Profile representation* yang digunakan yaitu *user-item ratings matrix* dianggap sebagai sebuah *randomized complete block design* (RCBD), dengan pengguna sebagai *treatment* dan *item* sebagai *block*. Perhitungan prediksi nilai *rating* memanfaatkan algoritma “*missing value*” pada RCBD. Operasi perhitungan yang dilakukan melibatkan nilai *rating* pada pengguna aktif, *item* target dan seluruh sel pada *user-item ratings matrix*.

Pada penelitian sebelumnya oleh Saptono (2010) dan Ati (2012) metode *User-Item Based Collaborative Filtering* diterapkan menggunakan *rating* eksplisit dimana data *rating* bersifat kontinu. Pada penelitian ini penulis menguji efektivitas metode dengan mengukur bagaimana kualitas presisi dan akurasi rekomendasi yang dihasilkan metode *User-Item Based Collaborative Filtering* pada sistem perekomendasi rental VCD dimana opini *user* diterapkan secara implisit karena pada *rating* implisit sistem secara otomatis memberikan *rating* berdasarkan *behaviour user* terhadap sistem sedangkan untuk kasus eksplisit membutuhkan tindakan dari *user* untuk melakukan *rating* pada suatu *item*. Kelebihan dari tipe eksplisit adalah tingkat kepercayaan sistem menjadi tinggi. Namun memiliki kelemahan yaitu pengguna menjadi terbebani dengan tugas perbaruan profil (Montaner *et al*, 2003). Oleh karena itu metode implisit

diterapkan dalam penelitian ini sehingga *user* tidak terbebani untuk melakukan tindakan *me-rating* secara langsung. *Rating user* diwakili oleh “0” bila *user* belum menyewa dan “1” bila *user* sudah menyewa.

Metode *User-Item Based Collaborative Filtering* diterapkan dalam penelitian ini menggunakan *Missing Value Algorithm (MVA)* yang merupakan modifikasi dari persamaan dasar *Randomized Complete Blocked Design (RCBD)*. Kemudian akan digunakan juga *Absoluted MVA* yang merupakan modifikasi dari *MVA* untuk meminimalkan anomali pada metode *User-Item Based Collaborative Filtering*. Pada penelitian ini juga melibatkan unsur *content* yaitu *genre* pada pembentukan matriks *user-item* untuk meningkatkan kualitas hasil rekomendasi metode *User-Item Based Collaborative Filtering*.

Pengujian kualitas metode *User-Item Based Collaborative Filtering* dilakukan terhadap presisi dan akurasi hasil prediksi pada data testing yang digunakan. Secara presisi diukur dengan menghitung nilai standar deviasi. Sedangkan secara akurasi diukur menggunakan *Mean Absolute Error (MAE)* dan *Error Probabilitas*.

## 1.2 Rumusan Masalah

Dari uraian di atas, dapat dibuat rumusan masalah yaitu bagaimana kualitas presisi dan akurasi rekomendasi yang dihasilkan oleh metode *User-Item Based Collaborative Filtering* pada sistem perekomendasi rental VCD menggunakan *Missing Value Algorithm (MVA)* dan *Absolute MVA*?

## 1.3 Batasan Masalah

Untuk tidak meluasnya lingkup pembahasan masalah dalam memberikan keterangan dalam tugas akhir ini, penulis akan memberikan batasan-batasan masalah yang akan dibahas. Adapun batasan masalah yang akan dibahas (diterangkan) dalam tugas akhir ini adalah:

- a. Data yang digunakan dalam sistem perekomendasi rental VCD ini adalah data transaksi yang diperoleh dari Toko Odiva Video Rental cabang Fajar Indah Solo selama 3 bulan yaitu januari, februari dan maret 2012.

*commit to user*

- b. Metode yang digunakan dalam sistem ini adalah *User-Item Based Collaborative filtering* dan data *rating* yang digunakan bersifat implisit.
- c. Pada penelitian ini hanya disajikan dua tolak ukur pengujian saja yaitu pengujian kualitas secara presisi dan akurasi.
- d. Pengguna akan mendapatkan rekomendasi judul film apabila telah menyewa paling sedikit 7 judul film.
- e. Pada pengujian dengan *error* probabilitas menggunakan *threshold* 0,065

#### 1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah menguji efektivitas metode dengan mengukur kualitas rekomendasi metode *User-Item Based Collaborative Filtering* pada sistem perekomendasi rental VCD dengan data *rating* yang bersifat implisit.

#### 1.5 Manfaat Tugas Akhir

Penelitian ini diharapkan dapat memberikan manfaat secara teoritis maupun praktis. Secara teoritis penelitian ini dapat digunakan sebagai referensi untuk penelitian selanjutnya terkait dengan penerapan metode *User-Item Based Collaborative Filtering*. Sedangkan manfaat secara praktis dapat memberikan rekomendasi rental VCD untuk para pelanggan.

#### 1.6 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini meliputi Bab I Pendahuluan, Bab II Tinjauan Pustaka, Bab III Metode Penelitian, Bab IV Pembahasan dan Bab V Penutup. Bab I Pendahuluan memuat latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan tugas akhir. Bab II Tinjauan Pustaka menguraikan dasar teori yang mendasari pembahasan secara detail, penelitian terkait yang pernah dilakukan serta rencana penelitian yang akan dilakukan dalam tugas akhir. Bab III Metode Penelitian menguraikan gambaran langkah-langkah yang dilakukan untuk melaksanakan dan menyelesaikan penelitian ini. Bab IV Pembahasan memaparkan hasil-hasil dari tahapan penelitian, yaitu tahap pemodelan data hingga dan serta hasil pengujian berupa penjelasan teoritik, baik secara kualitatif, kuantitatif, atau secara statistik. Bab V Penutup berisi kesimpulan dan saran. Kesimpulan berisi rumusan jawaban

*commit to user*

terhadap pertanyaan (perumusan masalah) dengan bukti-bukti yang ada dan telah dilakukan dalam penelitian ini. Saran merupakan sesuatu yang belum ditempuh dan layak untuk dilaksanakan pada penelitian selanjutnya.





## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Dasar Teori

##### 2.1.1 Pengertian Sistem Rekomendasi

Menurut Ricci *et al.* (2011), Sistem Rekomendasi adalah suatu teknik dan peralatan perangkat lunak yang menyediakan rekomendasi *item* yang berguna untuk *user*. Rekomendasi yang diberikan berhubungan dengan berbagai macam proses pengambilan keputusan seperti barang apa yang akan dibeli, musik apa yang akan didengarkan, atau berita online apa yang akan dibaca.

Sedangkan menurut Goldberg (2001), Sistem Rekomendasi merupakan model penyelesaian masalah yang menerapkan teknik-teknik tertentu pada pembuatan rekomendasi untuk pemilihan suatu informasi, produk dan jasa. Terdapat dua pendekatan dalam mengembangkan suatu *recommender system* yaitu *Content-Based* dan *Collaborative Filtering* (Kangas, 2002).

##### 2.1.2 Profile Generation dan Maintenance

*Profile generation* merupakan tahapan penting untuk membangun sistem rekomendasi. Dari proses ini, kebiasaan pengguna diperoleh, dipelajari, dan dieksploitasi untuk mengetahui ketertarikan pengguna. *Profile maintenance* lebih menekankan pada proses *profile update* berdasarkan masukan atau koreksi dari pengguna yang berupa *feedback*. Rancangan keputusan, yang digunakan untuk membangkitkan dan memelihara profil pengguna, antara lain: *profile representation*, *initial profile generation*, *profile learning*, dan *feedback*.

##### 2.1.2.1 Profile Representation

*Profile representation* adalah langkah pertama dalam sistem rekomendasi karena teknik lain bergantung pada langkah ini. Ketika langkah ini ditentukan, maka teknik yang lain dapat didefinisikan. Sistem rekomendasi tidak dapat mulai berfungsi sampai *user profile* dibuat. Selain itu, sistem perlu tahu sebanyak mungkin dari *user* untuk memberikan hasil yang memuaskan sejak awal. Oleh karena itu, sistem perlu melakukan *commit to user*.



karena itu sistem perlu menggunakan teknik yang sesuai dalam menghasilkan initial profile yang akurat (Montaner *et al*, 2003).

Beberapa model yang sering digunakan untuk merepresatasikan *user profile* pada *recommender system* antara lain: *history-based model*, *vector space model*, dan *user-item ratings matrix* (Montaner *et al*, 2003).

a) *History-based model*

Model ini banyak digunakan dalam *e-commerce* dimana sistem menyimpan daftar produk yang dibeli dan *rating user* sebagai *user profile*. Contoh penerapan model ini pada *e-commerce* yaitu Amazon.com dan CDNow.com. Pendekatan serupa juga diterapkan pada WebSell, dimana profile didefinisikan dengan menggunakan dua daftar, pertama produk yang telah dibeli dan telah dirating sebagai *interesting* dan yang lainnya sebagai *uninteresting*.

b) *Vector space model*

Pada model *vector space*, *item* direpresentasikan dengan fitur vektor, biasanya kata-kata atau konsep dengan nilai yang berhubungan. Nilai ini dapat berupa *boolean* atau *real number*. Nilai *boolean* merepresentasikan adanya nilai dari fitur, dan *real number* menggambarkan frekuensi, relevansi atau probabilitas fitur yang dihitung menggunakan teknik *information indexing*.

c) *User-item Ratings Matrix*

Hampir sama dengan *history-based model*, pendekatan ini juga memanfaatkan riwayat interaksi pengguna, namun yang digunakan khusus *rating history* saja. *Rating history* disimpan dalam sebuah matrik yang biasa disebut *user-item ratings matrix* yang merupakan representasi nilai *rating* dari *item*. Matrik inilah yang disebut sebagai *user profile* atau profil pengguna. Setiap sel matriks  $(u,i)$  memuat nilai *rating* yang merupakan penilaian pengguna  $u$  terhadap *item*  $i$ . Sedangkan nilai 0 (nol) menandakan bahwa *item*  $i$  belum di-rating oleh pengguna (Montaner *et al*, 2003). Nilai *rating* yang diberikan umumnya merupakan bilangan *real* pada interval

tertentu. Besarnya nilai *rating* mewakili besaran minat atau ketertarikan pengguna pada *item* yang di-*rating*.

#### 2.1.2.2 Initial Profile Generation

Sangat diharapkan untuk mempelajari sebanyak mungkin dari *user* agar sistem perekomendasi memberikan hasil yang memuaskan sejak awal. Namun, *user* biasanya tidak ingin menghabiskan banyak waktu dalam menentukan ketertarikannya dengan membentuk profilnya sendiri. Selain itu, ketertarikan *user* dapat berubah sewaktu-waktu, sehingga membuat profil sulit untuk dipelihara. Karena alasan-alasan tersebut, memulai dan memelihara profil *user* adalah aspek yang sulit dalam perancangan dan pengembangan *intelligent agent systems*. Pembentukan profil pengguna pada sistem perekomendasi berbeda-beda, ada yang bersifat otomatis (dibentuk sendiri oleh sistem perekomendasi), semi-otomatis (menggunakan *training set* dan stereotyping) dan manual (Montaner *et al*, 2003). Berikut ini beberapa cara pembentukan profil menurut Montaner *et al* (2003), yaitu:

a. *Empty Profile*

Sistem yang menggunakan *empty profile* tidak mempunyai data awal dari pengguna dan secara otomatis akan membangkitkan profil pengguna sejak pertama pengguna berinteraksi dengan sistem.

b. *Manual*

Sistem manual mengharuskan pengguna untuk melakukan pendaftaran mengenai ketertarikan mereka dalam bentuk kata kunci, topik, dan sebagainya. Kelebihan dari sistem ini yaitu rekomendasi bersifat transparan sehingga pengguna bisa memprediksi *item* apa yang akan direkomendasikan. Kelemahannya yaitu terlalu banyak yang harus dilakukan oleh pengguna.

c. *Stereotyping*

*Stereotyping* menciptakan model pengguna diawali dengan melakukan klasifikasi pengguna dalam deskripsi secara *stereotype* yang menggambarkan fitur dari kelas pengguna. Data yang biasa digunakan

*commit to user*

antara lain data demografi, data record, data geografi, karakteristik pengguna, serta data psikografi. Kelemahan dari pendekatan ini adalah kesulitan untuk memperoleh data yang bersifat personal dari pengguna.

d. *Training set*

*Training set* merupakan koleksi dari contoh-contoh interaksi pengguna dengan sistem yang digunakan untuk memperoleh *initial profile*. Pada pendekatan ini pengguna akan diminta untuk memberikan penilaian terhadap beberapa *item* contoh, berupa *rating* yang menggambarkan ketertarikan pengguna. Dari *training set* diharapkan sistem dapat memprediksi ketertarikan pengguna terhadap *item-item* lainnya di luar *training set* (Goldberg *et al*, 2001). Keuntungan dari model ini adalah kesederhanaan penanganan. Kelemahan sekaligus bahaya dari model ini adalah pemilihan contoh *item* yang tidak mewakili keinginan atau ketertarikan dari pengguna sehingga menyebabkan hasil rekomendasi yang kurang akurat (Montaner *et al*, 2003).

### 2.1.2.3 Profile Learning

Teknik *learning profile* membangun profil *user* melalui data *training set*. Teknik ini dapat dilihat sebagai langkah *preliminary* dalam merepresentasikan profil *user*. Beberapa sistem menerapkan *learning algorithm* untuk memperoleh *profile update* secara otomatis. Prinsipnya adalah profil pengguna menggambarkan ketertarikan pengguna pada waktu yang sama (Montaner *et al*, 2003).

Beberapa sistem mempunyai fase *offline* untuk mempelajari sebuah model dari kebiasaan pengguna dan selanjutnya sebuah fase *online* digunakan untuk mengaplikasikan model dalam waktu nyata. Namun kebanyakan sistem menggunakan pendekatan *lazy learning (online)*, sehingga model di bangun dan diperbaharui secara real time. Metode *offline learning* sangat cocok pada lingkungan yang perubahan pada pilihan pengguna berubah secara lambat terhadap waktu, tetapi tidak sesuai untuk sistem yang membutuhkan perubahan pilihan pengguna secara cepat dan sering (Montaner *et al*, 2003).

*commit to user*

#### 2.1.2.4 Feedback

Minat pengguna berganti setiap waktu. Hal ini akan mempengaruhi profil pengguna yang telah ada. Sistem perekomendasi perlu melakukan *update* pada profil pengguna sehingga diperoleh profil terkini dari pengguna. Tujuannya adalah untuk meningkatkan kepuasan pengguna terhadap rekomendasi yang diberikan. Oleh karena itu diperlukan *feedback* dari pengguna yang sesuai dengan ketertarikannya.

Beberapa tipe *user feedback* menurut Montaner *et al* (2003), antara lain: *no feedback*, *explicit feedback*, dan *implicit feedback*.

a) *No Feedback*

Sistem yang melakukan *update* profil pengguna secara manual tidak memerlukan *feedback* dari pengguna. Dengan kata lain sistem ini tidak akan mengubah profil penggunanya selama pengguna tidak melakukan perubahan (Montaner *et al*, 2003).

b) *Explicit feedback*

Pengguna secara eksplisit memberitahukan kepada sistem *item-item* apa saja yang disukai dan tidak disukai (baik dengan skala nilai atau *boolean*). Kelebihan dari tipe ini adalah tingkat kepercayaan sistem menjadi tinggi. Namun memiliki kelemahan yaitu pengguna menjadi terbebani dengan tugas perbaruan profil (Montaner *et al*, 2003).

c) *Implicit feedback*

Beberapa sistem tidak meminta pengguna untuk memberitahukan ketertarikan atau minatnya, namun cukup mempelajari kebiasaan pengguna saat mengakses sistem. Misalnya menganalisis *link* yang diikuti pengguna, *navigaton history*, waktu yang dihabiskan untuk suatu halaman web tertentu, ataupun kebiasaan pembelian. Kelebihan dari model ini adalah tidak membebani pengguna. Sedangkan kekurangannya yaitu pada tingkat kepercayaan sistem yang rendah (Montaner *et al*, 2003).

### 2.1.3 Collaborative Filtering

Prinsip kerja dari algoritma *Collaborative Filtering* adalah memberikan rekomendasi atau prediksi *item* berdasarkan pada opini pengguna-pengguna yang mempunyai kemiripan. Opini yang diberikan bisa bersifat eksplisit maupun implisit (Sarwar, 2001).

Ada dua hal utama yang dilakukan *recommender system* yang menggunakan algoritma *Collaborative Filtering* (Sarwar et al, 2001), yaitu:

- Prediksi, melakukan prediksi opini yang akan diberikan oleh pengguna dalam skala bilangan yang sama.
- Rekomendasi, memberikan rekomendasi berupa daftar *item* dengan nilai prediksi tertinggi. Hal yang perlu dicatat adalah *item-item* yang direkomendasikan belum pernah dibeli, dilihat atau di-*rating* oleh pengguna tersebut. Antarmuka ini pada *recommender system* yang menggunakan algoritma *Collaborative Filtering* disebut dengan *Top-N recommendation*.

Gambar dibawah ini menunjukkan skema kerja dari *recommender system* yang menggunakan pendekatan *Collaborative Filtering* dengan profile berbentuk *user-item ratings matrix* berukuran  $m \times n$ , dimana  $m$  adalah jumlah pengguna dan  $n$  adalah jumlah *item* sebagai input (Sarwar et al, 2001).



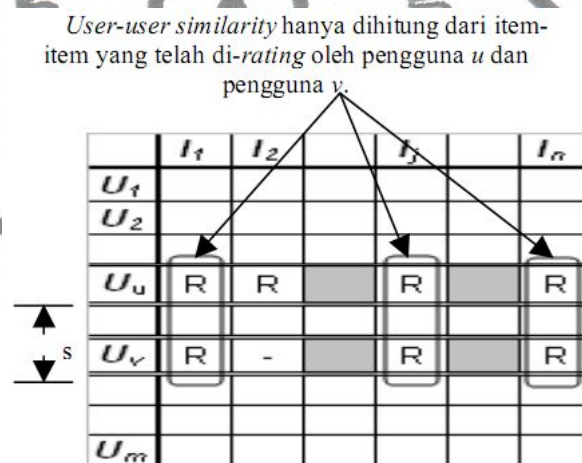
Gambar 2.1 Proses *Collaborative Filtering* (Sarwar et al, 2001)



Pendekatan *Collaborative Filtering* pada dasarnya dibagi menjadi dua kategori yaitu *User Based Collaborative Filtering* disebut juga *memory-based*, dan *Item Based Collaborative Filtering* yang disebut juga *model-based* (Sarwar et al, 2001).

a. *User-Based Collaborative Filtering*

Pada pendekatan *User-Based Collaborative Filtering*, proses pembentukan *user-neighborhood* dilakukan dengan terlebih dahulu melakukan perhitungan kemiripan antara pengguna aktif dengan pengguna lainnya (*user-user similarity*). Perhitungan hanya diberlakukan pada himpunan *item-item* yang telah di-rating oleh kedua pengguna yang dibandingkan seperti ditunjukkan gambar 2.2



Gambar 2.2 Pembentukan *User-Neighborhood* (Sarwar et al, 2001)

Untuk menghitung *user-user similarity* bisa digunakan persamaan *cosine-based similarity*. Pengguna yang dibandingkan misalnya  $u$  dan  $v$ , dianggap sebagai vektor baris dengan anggotanya adalah nilai *rating* yang terdapat pada baris tersebut. Dua vektor dikatakan sama jika membentuk sudut 0 derajat atau nilai cosinusnya 1. Dengan kata lain dua pengguna dikatakan mirip jika nilai cosinus dari perhitungan mendekati 1.

$$s(u, v) = \cos(\vec{R}(u, *), \vec{R}(v, *)) = \frac{\vec{R}(u, *) \cdot \vec{R}(v, *)}{\|\vec{R}(u, *)\| * \|\vec{R}(v, *)\|} \quad (2.1)$$

*commit to user*



di mana :

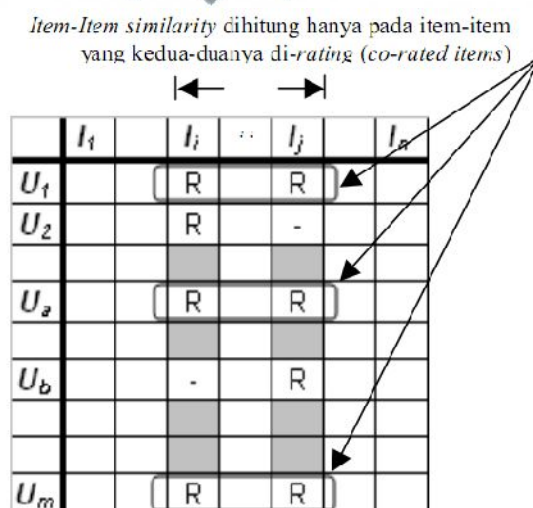
$R(u, *)$  adalah vektor baris dengan anggota nilai *rating* pada pengguna  $u$ .

$R((u, *), (v, *))$  adalah nilai cosinus sudut yang dibentuk vektor baris *rating* pengguna  $u$  dan  $v$ .

b. *Item-Based Collaborative Filtering*

Pendekatan *Item-Based Collaborative Filtering* menjadikan *item-item* yang telah di-*rating* oleh pengguna menjadi dasar perhitungan. Algoritma ini melakukan perhitungan bagaimana kemiripan *item-item* yang telah di-*rating* dengan *item-item* lain dan selanjutnya dipilih sekelompok *item* yang mempunyai kemiripan dengan *item* yang sudah di-*rating*. Nilai kemiripan tersebut dicatat untuk dijadikan nilai bobot untuk memprediksi nilai *rating* pada *item* target.

Dasar perhitungan kemiripan antara dua *item*  $i$  dan  $j$  adalah terlebih dahulu mengisolasi pengguna-pengguna yang telah menilai keduanya kemudian teknik *item similarity* diterapkan untuk memperoleh nilai kemiripan seperti ditunjukkan gambar dibawah ini



Gambar 2.3 Pembentukan *Item-Neighborhood* (Sarwar et al, 2001)

Nilai *item-item similarity* bisa dihitung dengan menggunakan persamaan *cosine-based similarity* seperti halnya menghitung *user-user similarity*. Yang berbeda adalah orientasinya, yaitu membandingkan antara dua *item*, misalnya *item i* dan *item j*. Sehingga *item-item* tersebut dianggap sebagai vektor kolom dengan anggota nilai *rating* pada kolom tersebut (Deshpande, 2004). Persamaan dibawah ini menunjukkan hal tersebut.

$$s(i, j) = \cos(\vec{R}(*, i), \vec{R}(*, j)) = \frac{\vec{R}(*, i) \cdot \vec{R}(*, j)}{\|\vec{R}(*, i)\| \cdot \|\vec{R}(*, j)\|} \quad (2.2)$$

di mana :

$R(i, *)$  adalah vektor kolom dengan anggota nilai *rating* pada pengguna *i*.

#### 2.1.4 Probabilitas Untuk Tipe Data Biner

Dalam situasi di mana kita tidak bisa memperoleh kepastian akan terjadinya suatu peristiwa, kita perlu menghitung peluang atau probabilitas terjadinya peristiwa tersebut. Untuk tipe data biner yang hanya terdiri dari “0” atau “1” diterapkan teori probabilitas frekuensi relatif. Dimana probabilitas frekuensi relatif menghitung probabilitas berdasarkan data historis. Seandainya pada sebuah eksperimen yang dilakukan sebanyak  $N$  kali terjadi kejadian  $A$  sebanyak  $f_A$  kali, maka jika eksperimen tersebut dilakukan tak terhingga kali banyaknya ( $N$  mendekati tak hingga), nilai limit dari frekuensi relatif  $f_A/N$  didefinisikan sebagai probabilitas kejadian  $A$  atau  $P(A)$  (Harinaldi, 2005).

$$P(A) = \lim_{N \rightarrow \infty} \frac{f_A}{N} \quad (2.3)$$

Berdasarkan persamaan diatas maka besarnya probabilitas tipe data biner dapat didefinisikan sebagai sebuah eksperimen yang dilakukan sebanyak  $N$  kali terjadi kejadian bernilai “1” sebanyak  $f_A$  kali.

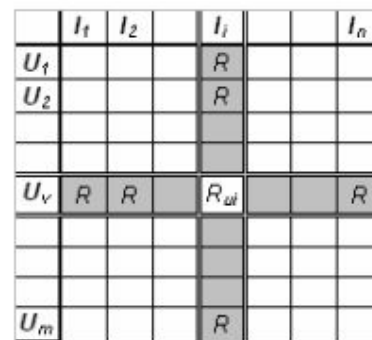
#### 2.1.6 User-Item Based Collaborative Filtering

Secara umum, klasifikasi *Collaborative Filtering* ditentukan oleh orientasi yang dijadikan sebagai dasar perhitungan untuk memprediksi nilai *rating*.  
*commit to user*

Pendekatan *user based Collaborative Filtering* menggunakan vektor  $r$  baris (yang merupakan sisi pengguna) pada *user-item ratings matrix* sebagai dasar perhitungan. Sedangkan *Item Based Collaborative Filtering* menggunakan vektor kolom (sisi *item*) sebagai dasar perhitungan.

Karakter utama dari pendekatan *User-Item Based Collaborative Filtering* adalah menggunakan vektor kolom dan vektor baris pada *user-item ratings matrix* sebagai dasar perhitungan untuk memperoleh nilai prediksi *rating*.

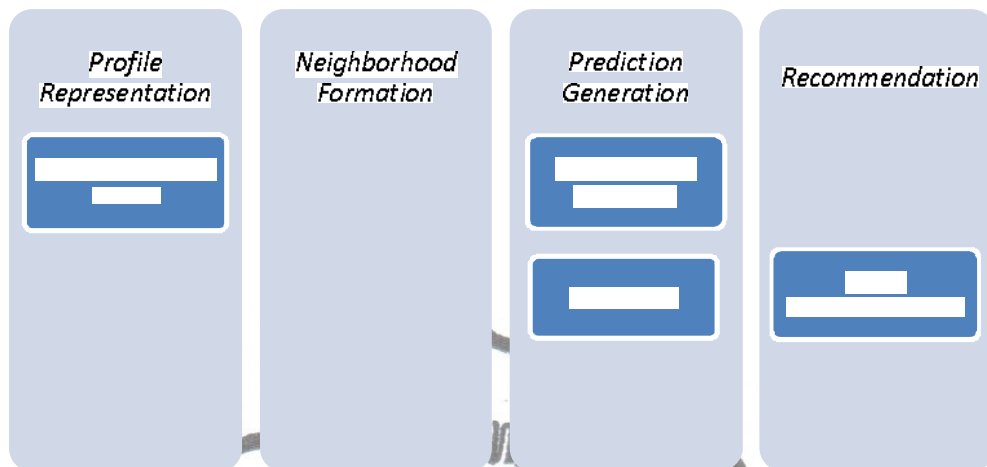
*User-Item Based Collaborative Filtering* pada tulisan ini dianalogikan sebagai sebuah *experimental design*. *Profile representation* yang digunakan yaitu *user-item ratings matrix* dianggap sebagai sebuah *randomized complete block design* (RCBD), dengan pengguna sebagai *treatment* dan *item* sebagai *block*. Perhitungan prediksi nilai *rating* memanfaatkan algoritma “missing value” pada RCBD. Operasi perhitungan yang dilakukan melibatkan nilai *rating* pada pengguna aktif, *item* target dan seluruh sel pada *user-item ratings matrix* seperti pada gambar 2.4



	$I_1$	$I_2$		$I_i$			$I_n$
$U_1$				$R_i$			
$U_2$				$R_i$			
$U_v$	$R$	$R$		$R_{ui}$			$R$
$U_m$				$R_i$			

Gambar 2.4 Proses perhitungan nilai prediksi *rating* (Saptono, 2006)

Secara umum, ada empat proses yang dilakukan pada metode *Collaborative Filtering* yaitu: *profile representation*, *neighborhood formation*, *prediction generation* dan *recommendation*. Proses pada metode *User-Item Based Collaborative filtering* terdiri dari *profile representation*, *prediction generation*, dan *recommendation*.



Gambar 2.5 Proses *User-Item Based Collaborative Filtering* (Saptono, 2006)

Keempat proses tersebut selama ini menjadi proses baku dalam metode *collaborative filtering*. Padahal sebenarnya proses utama yang harus dilakukan adalah memprediksi dan merekomendasikan (Goldberg, 2001) dan (Sarwar, 2001) seperti pada gambar 2.5. Oleh karena itu pendekatan *User-Item Based Collaborative Filtering* memberikan solusi untuk tidak melakukan pembentukan *neighborhood*, baik *user-neighborhood* maupun *item-neighborhood*. Dengan kata lain, pada pendekatan *User-Item Based Collaborative Filtering* melakukan efisiensi dalam perhitungan dengan tidak melakukan perhitungan kemiripan *item* maupun pengguna.

Proses-proses yang dilakukan pada *User-Item Based* seperti pada gambar 2.5, dengan proses keempat yaitu pemberian rekomendasi dalam bentuk antarmuka *top-N recommendation* sama dengan pendekatan lainnya.

#### a) *Profile Representation*

Seperti pendekatan lainnya pada metode *Collaborative Filtering*, pendekatan *User-Item Based Collaborative Filtering* juga memanfaatkan *user-item ratings matrix* sebagai *profile representation*. Pembentukan *user profile* dilakukan secara semi -otomatis dengan memanfaatkan *gauge set* seperti pada (Sarwar *et al*, 2001). Pengguna memberikan *feedback* sebagai opini secara implisit ketika *user* menyewa suatu *item*.

*commit to user*

b) *Neighborhood Formation*

Pendekatan *User-Item Based Collaborative Filtering* melewati proses ini, karena pada perhitungan prediksinya tidak diperlukan *neighborhood*. Dengan kata lain, pendekatan *User-Item Based Collaborative Filtering* tidak melakukan perhitungan kemiripan antar *item* maupun pengguna.

c) *Prediction Generation*

Nilai prediksi *rating item-item* di luar *gauge set* diperoleh dengan cara memanfaatkan algoritma perhitungan "*missing value*". Sel-sel yang kosong pada pengguna aktif dianggap sebagai suatu "*missing value*". Nilai prediksi *rating* oleh pengguna *i* pada *item j* pada matrik berukuran  $m \times n$ , adalah  $P(i, j)$

$$P(i, j) = \frac{n * R'(i, *) + m * R'(*, j) - R'(*, *)}{(m - 1)(n - 1)} \quad (2.4)$$

Di mana

$R'(i, *)$  : jumlah nilai *rating* pada baris ke-*i*.

$R'(*, j)$  : jumlah nilai *rating* pada kolom ke-*j*.

$R'(*, *)$  : jumlah nilai *rating* seluruh sel pada matrik.

$m$  : jumlah baris, sebagai representasi jumlah pengguna .

$n$  : jumlah kolom, sebagai representasi jumlah *item*.

Pada tipe data biner dikembangkan berdasarkan teori probabilitas frekuensi relatif yang telah dijelaskan pada dasar teori, besamya  $m$  dan  $n$  dianggap mendekati tak hingga ( $\sim$ ) sehingga persamaan (2.4) menjadi :

$$P(i, j) = P(i, *) + P(*, j) - P(*, *) \quad (2.5)$$

Dengan  $P(i, j)$  adalah nilai Probabilitas dari pasangan matrik *user i* dan *item j*,  $P(i, *)$  adalah nilai Probabilitas dari *user i*,  $P(*, j)$  adalah nilai Probabilitas dari *item j*, dan  $P(*, *)$  adalah nilai Probabilitas dari seluruh matrik *user-item*.

*commit to user*



Modifikasi rumus prediksi MVA menjadi *absolute MVA* dilakukan untuk meminimalkan anomali pada metode *User-Item Based Collaborative Filtering* seperti yang terjadi pada penelitian sebelumnya oleh Ati (2012). Modifikasi pada *absolute MVA* yaitu dengan melakukan pemutlakan pada rumus MVA pada persamaan (2.3)

$$P(i, j) = |P(i, *) + P(*, j) - P(*, *)| \quad (2.6)$$

### 2.1.7 Evaluasi Sistem

#### a) Mean Absolute Error (MAE)

Menurut Sarwar *et al.* (2001) dalam mengevaluasi kualitas hasil rekomendasi ada 2 cara yaitu *statistical accuracy metrics* dan *decision support accuracy metrics*. Cara yang paling sering digunakan adalah *statistical accuracy metrics* yaitu dengan mengevaluasi keakuratan sistem dengan membandingkan angka nilai rekomendasi dengan *rating* pasangan *user-item* yang sebenarnya pada pengujian dataset. *Mean Absolute Error (MAE)* antara *rating* dan prediksi adalah metrik yang umumnya digunakan. MAE adalah pengukuran deviasi rekomendasi dari nilai spesifik *user* yang sebenarnya. Untuk setiap pasangan prediksi  $p_i$  dan *rating*  $q_i$   $\langle p_i, q_i \rangle$  metrik ini menghitung *absolute error* antara  $|p_i - q_i|$ . MAE dihitung dengan menjumlahkan *absolute error* dari  $N$  pasangan prediksi *rating* yang saling berhubungan, lalu menghitung rata-ratanya. Semakin rendah MAE, makin akurat prediksi rekomendasi yang dihasilkan.

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (2.7)$$

#### b) Standar Deviasi

Menurut Frederick *et al.* (2010) standar deviasi menggunakan rata-rata distribusi sebagai referensi utama dan mengukur variabilitas dengan mempertimbangkan jarak antara tiap nilai dengan rata-rata hitung. Dengan kata lain standar deviasi menyediakan pengukuran standar atau rata-rata, jarak dari



rata-rata hitung dan menggambarkan apakah nilai menyebar disekitar rata-rata hitung atau menyebar secara meluas.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (2.8)$$

Dengan:

$x_i = \{x_1, x_2, \dots, x_N\}$  : Nilai data dari sampel

$\bar{x}$  : Rata-rata dari sampel

## 2.2 Penelitian Terkait

Penelitian sebelumnya yang pernah dilakukan berkaitan tentang sistem rekomendasi dengan metode *User-Item Based Collaborative Filtering* antara lain :

### a. *Recommender System* untuk Pencarian Buku dengan *User-Item Based Collaborative Filtering* (Saptono, 2006)

Dalam Penelitian ini diterapkan metode *User-Item Based Collaborative Filtering* untuk sistem rekomendasi pencarian buku di perpustakaan. Pada penelitian ini metode *User-Item Based Collaborative Filtering* dibandingkan dengan metode lain yaitu *Item-Based Collaborative Filtering* dengan teknik perhitungan *similarity adjusted cosine* dan *pearson correlation*. Dari hasil percobaan menggunakan sample data dari *Movielens* diperoleh kesimpulan bahwa metode *User-Item Based Collaborative Filtering* memberikan rekomendasi dengan kualitas yang lebih baik dibandingkan dengan metode *Item Based Collaborative Filtering* baik yang menggunakan *adjusted cosine* maupun *pure correlation*. Penelitian ini menggunakan *rating* eksplisit. Dengan mengacu pada penelitian ini penulis ingin menguji efektivitas metode *User-Item Based Collaborative Filtering* namun dengan *rating* yang bersifat implisit, pada sistem rekomendasi rental VCD.

*commit to user*

**b. Peningkatan Efektivitas Metode *User-Item Based Collaborative Filtering* pada Sistem Rekomendasi Wisata Kuliner Kota Solo (Ati, 2012)**

Penelitian ini memberikan rekomendasi *item* berdasarkan urutan teratas nilai prediksi *item* yang dihasilkan. Prediksi *rating* dihitung menggunakan metode *User-Item Based Collaborative Filtering*. Metode ini menggunakan algoritma *missing value* pada perhitungan prediksi *rating* dan tidak memerlukan proses pembentukan *neighborhood*. Untuk meminimalkan terjadinya anomali, pada penelitian ini dilakukan perbaikan atau modifikasi pada *missing value algorithm* (MVA) menjadi *adjusted* MVA. Perbaikan yang dilakukan pada *adjusted* MVA yaitu melakukan proporsi terhadap jumlah rata-rata dalam perhitungan prediksi nilai *rating* pada sistem rekomendasi wisata *item* Kota Solo. *Rating* pada penelitian ini juga diperoleh secara eksplisit. Pada penelitian ini disimpulkan bahwa penggunaan *adjusted* MVA lebih efektif dalam meningkatkan kualitas rekomendasi. Mengacu pada penelitian ini penulis akan menguji efektivitas metode *User-Item Based Collaborative Filtering* namun dengan *rating* yang bersifat implisit, pada sistem rekomendasi rental VCD, dan untuk meminimalkan terjadinya anomali penulis akan melakukan perbaikan atau modifikasi pada *missing value algorithm* (MVA) menjadi *absoluted* MVA serta penerapan *cluster genre* pada matriks *user-item* nya.

**c. *Exploiting Various Implicit Feedback for Collaborative Filtering* yang dimuat dalam *Proceedings of the 21st International Conference Companion on World Wide Web* (Byong Ju Yang et al., 2012).**

Dalam penelitian ini ditunjukkan bahwa perbedaan *user implicit feedback* dapat digunakan untuk meningkatkan akurasi rekomendasi secara signifikan. Penelitian ini mengklasifikasikan bermacam-macam

*commit to user*

perilaku *user* (misalnya pencarian *item*, *skip*, *add to playlist*, dll) ke dalam kelompok feedback positif atau negatif dan membangun *pseudo-rating matrix* yang lebih akurat. Evaluasi metode ini diterapkan menggunakan log dataset dari sebuah website yang menyediakan 20.514 lagu untuk *user*. Dari penelitian yang dilakukan diperoleh kesimpulan bahwa *Exploiting Various Implicit Feedback for Collaborative Filtering* menunjukkan performansi yang lebih baik daripada beberapa metode dasar dan perilaku “*play*” lebih mempengaruhi performansi dibandingkan perilaku (*behaviour*) *user* yang lain. Dengan mengacu pada penelitian ini penulis kemudian ingin menguji efektivitas metode *User-Item Based Collaborative Filtering* namun dengan *rating* yang bersifat implisit, pada sistem rekomendasi rental VCD, tetapi *behaviour user* yang dianggap sebagai *rating* dibatasi yaitu bernilai “1” bila sudah pernah menyewa suatu *item* dan “0” bila belum pernah menyewa suatu *item*.

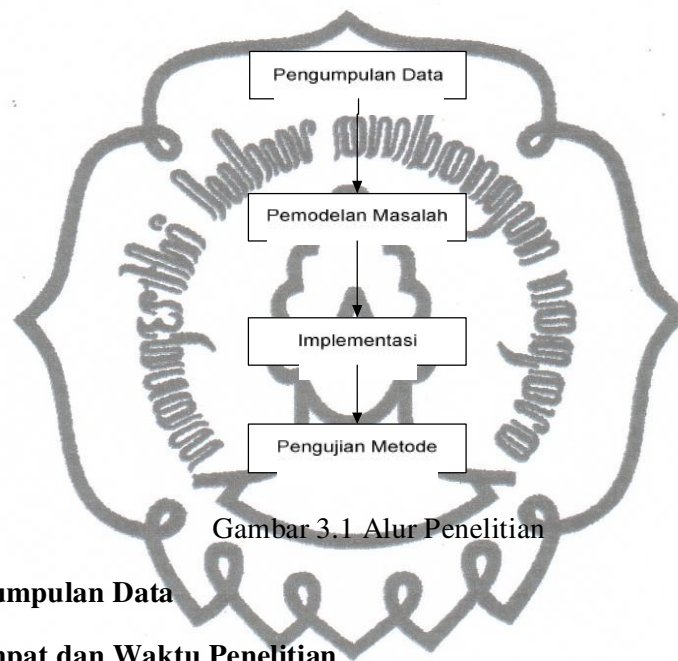
### 2.3 Rencana Penelitian

Rencana dari penelitian Tugas Akhir ini akan berfokus pada pengujian efektivitas metode *User-Item Based Collaborative Filtering* pada sistem rekomendasi rental VCD. Pengujian efektivitas metode dilakukan dengan mengukur kualitas rekomendasi secara presisi dan akurasi. Pada tahap pembentukan matrik *user-item* nya akan dipecah menjadi beberapa cluster berdasarkan *genre item* yang sudah disewa oleh pengguna sehingga matrik yang terbentuk berukuran lebih kecil. Dalam penerapannya akan digunakan rumus prediksi *Missing Value Algorithm (MVA)* dan *Absolute MVA* untuk menghasilkan prediksi nilai *rating*. Nilai *rating* yang digunakan dalam penelitian ini bersifat implisit yaitu diwakili dengan angka biner, *rating* bernilai “0” bila *user* belum pernah menyewa suatu *item* dan *rating* bernilai “1” bila *user* sudah pernah menyewa suatu *item*.

### BAB III

#### METODOLOGI PENELITIAN

Penelitian ini akan dilaksanakan berdasarkan rancangan penelitian seperti yang ditunjukkan pada gambar 3.1



Gambar 3.1 Alur Penelitian

#### 3.1 Pengumpulan Data

##### 3.1.1 Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Toko Odiva Video Rental cabang Fajar Indah Solo. Data yang digunakan adalah data transaksi penyewaan VCD selama 3 bulan Januari, Februari dan Maret 2012.

##### 3.1.2 Metode Pengumpulan Data

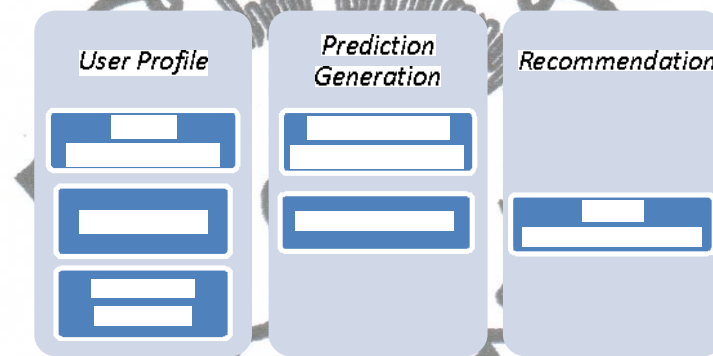
Pengumpulan data meliputi mengumpulkan data transaksi penyewaan dan studi literatur. Pengumpulan data transaksi penyewaan dilakukan di Odiva Video Rental cabang Fajar Indah Surakarta selama bulan Januari, Februari, dan Maret 2012. Pada pengujian metode *User-Item Based Collaborative Filtering* sampel data yang digunakan sebanyak 208 *item*, 33 *user* yang telah melakukan transaksi sewa terhadap tujuh atau lebih *item* dan 443 transaksi yang melibatkan *user* dan *item* terpilih di atas

*commit to user*

Studi literatur dilakukan dengan mencari dan mengumpulkan referensi berupa jurnal, skripsi, serta tesis yang berkaitan dengan sistem perekomendasi atau *Recommender System*, *Collaborative Filtering*, dan metode *User-Item Based Collaborative Filtering*. Keluaran yang diperoleh dari tahap ini adalah rangkuman dasar teori dan penelitian terkait yang pernah dilakukan sebelumnya.

### 3.2 Pemodelan Masalah

Pemodelan masalah meliputi penjabaran alur langkah kerja atau proses yang terdapat pada metode *User-Item Based Collaborative Filtering* yaitu :



Gambar 3.2 Langkah kerja metode *User-Item Based Collaborative Filtering*

#### 3.2.1 Pembentukan Profil Pengguna

##### 1. *Profile Representation*

Representasi profil pengguna menggunakan *Rating History* yang disimpan dalam sebuah matrik yang biasa disebut *user-item rating matrix* yang merupakan representasi nilai *rating* dari *item*. Setiap sel matriks( $u,i$ ) berisi nilai *rating* 1 atau 0. Nilai 0 terhadap suatu *item* berarti bahwa *user* yang bersangkutan belum pernah menyewa *item* tersebut, sedangkan nilai 1 berarti bahwa *user* yang bersangkutan sudah pernah menyewa *item* tersebut.

##### 2. *Initial Profile*

Pembentukan profil pengguna dalam sistem perekomendasi ini menggunakan *training set*. *Training set* dibentuk dari matriks *user-item* yang berasal dari basis data pengguna, transaksi penyewaaan *item* yang dilakukan oleh pengguna dan katalog *item*.

*commit to user*

### 3. *Relevance Feedback*

Pengguna memberikan *feedback* kepada sistem secara implisit atau secara tidak langsung yaitu ketika pengguna yang menyewa satu atau lebih *item*.

#### 3.2.2 Perhitungan Nilai Prediksi (*Prediction Generation*)

##### 1. Perhitungan Prediksi Menggunakan MVA

Perhitungan prediksi nilai *rating* menggunakan rumus prediksi *Missing Value Algorithm* (MVA) diadopsi dari persamaan dasar *Randomize Complete Block Design* (RCBD) yang memiliki *missing value* dan dikembangkan menggunakan teori probabilitas frekuensi relatif.

$$P(u, i) = P(u, *) + P(*, i) - P(*, *)$$

Dimana  $P(u, i)$  menyatakan nilai probabilitas dari pasangan matrik *user*  $u$  dan *item*  $i$ ,  $P(u, *)$  menyatakan nilai probabilitas dari *user*  $u$ ,  $P(*, i)$  menyatakan nilai probabilitas dari *item*  $i$ ,  $P(*, *)$  menyatakan nilai probabilitas dari seluruh matrik *user-item*.

Modifikasi rumus prediksi MVA menjadi *absoluted* MVA perlu dilakukan untuk meminimalkan anomali pada metode *User-Item Based Collaborative Filtering* seperti yang terjadi pada penelitian sebelumnya oleh Ati (2012).

##### 2. Perhitungan Prediksi Menggunakan *Absoluted* MVA

Perhitungan prediksi nilai *rating* dengan memodifikasi rumus MVA yaitu dengan melakukan pemutlakan pada rumus MVA menjadi

$$P(u, i) = |P(u, *) + P(*, i) - P(*, *)|$$

##### 3. Penggunaan *Cluster Genre* Pada Matriks *User-Item*

Pada pembentukan matriks *user-item* pada metode *User-Item Based Collaborative Filtering* diklasifikasikan atau di-*cluster* berdasarkan *genre item*. Pengklasifikasian *item* terhadap *genre* dibatasi ke dalam

*commit to user*



lima macam *genre*, yaitu: *Action/Adventure*, *Drama/Romance*, *Horror/Mystery*, *Comedy*, dan *Kids/Animation*. Perhitungan nilai prediksi tetap menggunakan MVA dan *absolute* MVA, perbedaannya hanya terletak pada matriks yang digunakan lebih kecil karena matriks yang terbentuk diklasifikasikan berdasarkan *genre item* yang sudah pernah disewa oleh *user*.

Pada saat perhitungan prediksi dan rekomendasi dengan *User-Item Based Collaborative Filtering*, *genre item* yang akan diprediksi terlebih dahulu dicocokkan dengan kelompok *genre* milik *user* aktif atau kelompok *genre* yang sudah pernah disewa oleh *user* aktif. Apabila tidak cocok, maka *item* tersebut tidak akan dihitung prediksinya sehingga tidak mungkin pula masuk ke dalam daftar *item* yang direkomendasikan untuk *user* aktif.

### 3.2.3 Rekomendasi

*Top-5 Recommendation*, dari perhitungan prediksi akan dipilih lima *item* yang memiliki nilai prediksi yang paling besar untuk direkomendasikan kepada pengguna.

### 3.2.4 Contoh Perhitungan

Untuk mempermudah memahami proses perhitungan prediksi *rating* dan pemberian rekomendasi menggunakan MVA maupun *absolute* MVA maka diberikan contoh kasus sebagai berikut.

Tabel 3.1 Representasi Matrik Rating

	M1	M2	M3	M4	M5
U1	0	1	0	1	0
U2	0	1	1	1	0
U3	0	1	1	0	0
U4	1	1	0	1	0
U5	1	1	1	0	0

*commit to user*

Terdapat sebuah sistem rekomendasi yang mengelola sebuah himpunan pengguna yang terdiri dari lima orang yaitu  $U = \{U1, U2, U3, U4, U5\}$  dan sebuah himpunan *item* yang terdiri dari lima macam *item*  $M = \{M1, M2, M3, M4, M5\}$ . Representasi *user profile* yang digunakan adalah *user-item ratings matrix*. Berdasarkan Tabel 3.1 di atas dapat diketahui sel yang kosong yaitu sel (1,1), (1,3), (1,5), (2,1), (2,5), (3,1), (3,4), (3,5), (4,3), (4,5), (5,4) dan (5,5). Dari tabel tersebut juga dapat diketahui bahwa *item* M5 merupakan *item* baru karena nilai *rating* pada kolom ke-5 kosong. Dengan menggunakan rumus MVA dan *absolute* MVA akan diprediksi nilai *rating*-nya pada sel-sel yang kosong tersebut.

Untuk menghasilkan rekomendasi menggunakan MVA, langkah-langkah yang dilakukan sebagai berikut:

1. Menghitung jumlah pengguna  $m = 5$
2. Menghitung jumlah *item*  $n = 5$
3. Menghitung jumlah sel dalam tiap-tiap baris

$$R(1,*) = 0 + 1 + 0 + 1 + 0 = 2$$

$$R(2,*) = 0 + 1 + 1 + 1 + 1 = 3$$

$$R(3,*) = 0 + 1 + 1 + 0 + 0 = 2$$

$$R(4,*) = 1 + 1 + 0 + 1 + 0 = 3$$

$$R(5,*) = 1 + 1 + 1 + 0 + 0 = 3$$

4. Menghitung jumlah sel dalam tiap-tiap kolom

$$R(*,1) = 0 + 0 + 0 + 1 + 1 = 2$$

$$R(*,2) = 1 + 1 + 1 + 1 + 1 = 5$$

$$R(*,3) = 0 + 1 + 1 + 0 + 1 = 3$$

$$R(*,4) = 1 + 1 + 0 + 1 + 0 = 3$$

$$R(*,5) = 0 + 0 + 0 + 0 + 0 = 0$$

5. Menghitung jumlah seluruh sel  $R(*,*) = 13$

6. Menghitung nilai prediksi untuk sel-sel yang kosong menggunakan MVA

*commit to user*

$$P(1,1) = \frac{2}{5} + \frac{2}{5} - \frac{13}{25} = 0,28$$

$$P(1,3) = \frac{2}{5} + \frac{3}{5} - \frac{13}{25} = 0,48$$

$$P(1,5) = \frac{2}{5} + \frac{0}{5} - \frac{13}{25} = -0,12$$

$$P(2,1) = \frac{3}{5} + \frac{2}{5} - \frac{13}{25} = 0,48$$

$$P(2,5) = \frac{3}{5} + \frac{0}{5} - \frac{13}{25} = 0,08$$

$$P(3,1) = \frac{2}{5} + \frac{2}{5} - \frac{13}{25} = 0,28$$

$$P(3,4) = \frac{2}{5} + \frac{3}{5} - \frac{13}{25} = 0,48$$

$$P(3,5) = \frac{2}{5} + \frac{0}{5} - \frac{13}{25} = -0,12$$

$$P(4,3) = \frac{3}{5} + \frac{3}{5} - \frac{13}{25} = 0,68$$

$$P(4,5) = \frac{3}{5} + \frac{0}{5} - \frac{13}{25} = 0,08$$

$$P(5,4) = \frac{3}{5} + \frac{0}{5} - \frac{13}{25} = 0,68$$

$$P(5,5) = \frac{3}{5} + \frac{0}{5} - \frac{13}{25} = 0,08$$

7. Mengurutkan nilai prediksi *rating item* dari yang tertinggi hingga terendah untuk tiap-tiap pengguna sehingga didapatkan *Top-N Recommendation*.

$$U1 \rightarrow \{M3, M1\} , U2 \rightarrow \{M1, M5\} , U3 \rightarrow \{M4, M1\}$$

$$U4 \rightarrow \{M3, M5\} , U5 \rightarrow \{M4, M5\}$$

*commit to user*

Sedangkan untuk menghasilkan rekomendasi menggunakan *absolute* MVA, langkah-langkah yang dilakukan sebagai berikut.

1. Langkah 1-5 sama seperti langkah 1-5 menggunakan MVA
2. Menghitung jumlah seluruh sel *rating mn* = 25
3. Menghitung nilai prediksi untuk sel-sel yang kosong menggunakan *absolute* MVA

$$P(1,1) = \left| \frac{2}{5} + \frac{2}{5} - \frac{13}{25} \right| = 0,28$$

$$P(1,3) = \left| \frac{2}{5} + \frac{3}{5} - \frac{13}{25} \right| = 0,48$$

$$P(1,5) = \left| \frac{2}{5} + \frac{0}{5} - \frac{13}{25} \right| = 0,12$$

$$P(2,1) = \left| \frac{3}{5} + \frac{2}{5} - \frac{13}{25} \right| = 0,48$$

$$P(2,5) = \left| \frac{3}{5} + \frac{0}{5} - \frac{13}{25} \right| = 0,08$$

$$P(3,1) = \left| \frac{2}{5} + \frac{2}{5} - \frac{13}{25} \right| = 0,28$$

$$P(3,4) = \left| \frac{2}{5} + \frac{3}{5} - \frac{13}{25} \right| = 0,48$$

$$P(3,5) = \left| \frac{2}{5} + \frac{0}{5} - \frac{13}{25} \right| = 0,12$$

$$P(4,3) = \left| \frac{3}{5} + \frac{3}{5} - \frac{13}{25} \right| = 0,68$$

$$P(4,5) = \left| \frac{3}{5} + \frac{0}{5} - \frac{13}{25} \right| = 0,08$$

$$P(5,4) = \left| \frac{3}{5} + \frac{0}{5} - \frac{13}{25} \right| = 0,68$$

$$P(5,5) = \left| \frac{3}{5} + \frac{0}{5} - \frac{13}{25} \right| = 0,08$$

4. Mengurutkan nilai prediksi *rating item* dari yang tertinggi hingga terendah untuk tiap-tiap pengguna sehingga didapatkan *Top-N Recommendation*.

$$U1 \rightarrow \{M3, M1, M5\}, U2 \rightarrow \{M1, M5\},$$

$$U3 \rightarrow \{M4, M1, M5\}, U4 \rightarrow \{M3, M5\}, U5 \rightarrow \{M4, M5\}$$

Terlihat bahwa *Item M5* yang sebelumnya tidak direkomendasikan untuk *user U1* dan *U3* setelah dihitung menggunakan *Absolute MVA* menjadi *item* yang direkomendasikan untuk *user U1* dan *U3*.

Untuk menghasilkan rekomendasi menggunakan *cluster genre* pada pembentukan matriks dengan MVA, langkah-langkah yang dilakukan sebagai berikut.

1. Membentuk matriks sesuai *genre item* dan mengelompokkan *user* berdasarkan *genre* yang pernah disewa, misalnya *item M1* dan *M4* bergenre *Action*, dan *item M2*, *M3*, dan *M5* bergenre *Drama*, lebih jelasnya seperti tabel 3.2 dan tabel 3.3 dibawah ini

Tabel 3.2 Representasi Matrik *Genre Action*

	M1	M4
U1	0	1
U2	0	1
U4	1	1
U5	1	0

Tabel 3.3 Representasi Matrik *Genre* Drama

	M2	M3	M5
U1	1	0	0
U2	1	1	0
U3	1	1	0
U4	1	0	0
U5	1	1	0

2. Menghitung Jumlah pengguna  $m$  Action=4 dan  $m$  Drama=5
3. Menghitung Jumlah *item*  $n$  Action=2 dan  $n$  Drama=3
4. Menghitung jumlah sel dalam tiap-tiap baris pada matriks *genre*

Action :

$$R(1,*) = 0 + 1 = 1$$

$$R(2,*) = 0 + 1 = 1$$

$$R(4,*) = 1 + 1 = 2$$

$$R(5,*) = 1 + 0 = 1$$

Drama :

$$R(1,*) = 1+0+0=1$$

$$R(2,*) = 1+1+0=2$$

$$R(3,*) = 1+1+0=2$$

$$R(4,*) = 1 + 0 + 0=1$$

$$R(5,*) = 1 + 1 + 0=2$$

5. Menghitung jumlah sel dalam tiap-tiap kolom pada matriks *genre*

Action :

$$R(*,1) = 0+0+1+1=2$$

$$R(*,4) = 1 + 1 + 1 + 0 = 3$$



Drama :

$$R(*,2) = 1 + 1 + 1 + 1 + 1 = 5$$

$$R(*,3) = 0 + 1 + 1 + 0 + 1 = 3$$

$$R(*,5) = 0 + 0 + 0 + 0 + 0 = 0$$

6. Menghitung jumlah seluruh sel *rating* matriks *genre* Action = 5 dan *genre* Drama = 8
7. Menghitung nilai prediksi untuk sel-sel yang kosong menggunakan MVA

$$P(1,1) = \frac{1}{2} + \frac{2}{4} - \frac{5}{8} = 0,375$$

$$P(1,3) = \frac{1}{3} + \frac{3}{5} - \frac{8}{15} = 0,4$$

$$P(1,5) = \frac{1}{3} + \frac{0}{5} - \frac{8}{15} = -0,2$$

$$P(2,1) = \frac{1}{2} + \frac{2}{4} - \frac{5}{8} = 0,375$$

$$P(2,5) = \frac{2}{3} + \frac{0}{5} - \frac{8}{15} = 0,13$$

$$P(3,5) = \frac{2}{3} + \frac{0}{5} - \frac{8}{15} = -0,13$$

$$P(4,3) = \frac{1}{3} + \frac{3}{5} - \frac{8}{15} = 0,4$$

$$P(4,5) = \frac{1}{3} + \frac{0}{5} - \frac{9}{15} = -0,2$$

$$P(5,4) = \frac{1}{2} + \frac{3}{4} - \frac{5}{8} = 0,625$$

$$P(5,5) = \frac{2}{3} + \frac{0}{5} - \frac{8}{15} = 0,13$$

8. Mengurutkan nilai prediksi *rating item* dari yang tertinggi hingga terendah untuk tiap-tiap pengguna sehingga didapatkan *Top-N Recommendation*.

*commit to user*

$$U1 \rightarrow \{M3, M1\}, U2 \rightarrow \{M1, M5\}, U3 \rightarrow \{M5\}$$

$$U4 \rightarrow \{M3\}, U5 \rightarrow \{M4, M5\}$$

Untuk menghasilkan rekomendasi menggunakan *cluster genre* pada pembentukan matriks dengan *absolute MVA*, langkah-langkah yang dilakukan sebagai berikut.

1. Untuk langkah 1-6 sama seperti perhitungan menggunakan *cluster genre* pada pembentukan matriks dengan MVA diatas
2. Menghitung nilai prediksi untuk sel-sel yang kosong menggunakan *absolute MVA*

$$P(1,1) = \left| \frac{1}{2} + \frac{2}{4} - \frac{5}{8} \right| = 0,375$$

$$P(1,3) = \left| \frac{1}{3} + \frac{3}{5} - \frac{8}{15} \right| = 0,4$$

$$P(1,5) = \left| \frac{1}{3} + \frac{0}{5} - \frac{8}{15} \right| = 0,2$$

$$P(2,1) = \left| \frac{1}{2} + \frac{2}{4} - \frac{5}{8} \right| = 0,375$$

$$P(2,5) = \left| \frac{2}{3} + \frac{0}{5} - \frac{8}{15} \right| = 0,13$$

$$P(3,5) = \left| \frac{2}{3} + \frac{0}{5} - \frac{8}{15} \right| = 0,13$$

$$P(4,3) = \left| \frac{1}{3} + \frac{3}{5} - \frac{8}{15} \right| = 0,4$$

$$P(4,5) = \left| \frac{1}{3} + \frac{0}{5} - \frac{9}{15} \right| = 0,2$$

$$P(5,4) = \left| \frac{1}{2} + \frac{3}{4} - \frac{5}{8} \right| = 0,625$$

$$P(5,5) = \left| \frac{2}{3} + \frac{0}{5} - \frac{8}{15} \right| = 0,13$$

*commit to user*

3. Mengurutkan nilai prediksi *rating item* dari yang tertinggi hingga terendah untuk tiap-tiap pengguna sehingga didapatkan *Top-N Recommendation*.

$U1 \rightarrow \{M3, M1, M5\}$  ,  $U2 \rightarrow \{M1, M5\}$  ,  $U3 \rightarrow \{M5\}$

$U4 \rightarrow \{M3, M5\}$  ,  $U5 \rightarrow \{M4, M5\}$

Tabel 3.4 Tabel Rating dan Prediksi MVA dan *Absolute MVA*

<i>User</i>	<i>Item</i>	<i>Rating</i>	Tanpa <i>Cluster genre</i>		Dengan <i>Cluster genre</i>	
			Prediksi MVA	Prediksi <i>Absolute MVA</i>	Prediksi MVA	Prediksi <i>Absolute MVA</i>
U1	M1		0,28	0,28	0,375	0,375
U1	M2	1				
U1	M3		0,48	0,48	0,4	0,4
U1	M4	1				
U1	M5		-0,12	0,12	-0,2	0,2
U2	M1		0,48	0,48	0,375	0,375
U2	M2	1				
U2	M3	1				
U2	M4	1				
U2	M5		0,08	0,08	0,13	0,13
U3	M1		0,28	0,28	-	-
U3	M2	1				
U3	M3	1				
U3	M4		0,48	0,48	-	-
U3	M5		-0,12	0,12	0,13	0,13
U4	M1	1				
U4	M2	1				
U4	M3		0,68	0,68	0,4	0,4
U4	M4	1				
U4	M5		0,08	0,08	-0,2	0,2
U5	M1	1				
U5	M2	1				
U5	M3	1				
U5	M4		0,68	0,68	0,625	0,625
U5	M5		0,08	0,08	0,13	0,13

Dari Tabel 3.4 pada kolom tanpa *cluster genre* prediksi MVA diperoleh prediksi nilai *rating* di luar *range* yaitu pada sel (U1,M5 dan sel (U3,M5) dengan nilai prediksi -0.12. Pada kolom prediksi *absolute* MVA untuk sel (U1,M5) dan (U3,M5) nilai prediksi di luar *range* tersebut dapat diperbaiki.

Pada kolom dengan *cluster genre* prediksi MVA diperoleh prediksi nilai *rating* di luar *range* yaitu pada sel (U1,M5) dan sel (U4,M5) dengan nilai prediksi -0.2. Pada kolom prediksi *absolute* MVA untuk sel (U1,M5) dan (U4,M5) nilai prediksi di luar *range* tersebut dapat diperbaiki. Kemudian untuk sel (U3,M1) dan (U3,M4) tidak dihitung nilai prediksinya atau tidak direkomendasikan karena M1 dan M4 merupakan *genre* Action sedangkan *user* U3 belum pernah meminjam *genre* Action.

### 3.3 Tahap Implementasi

Tahap implementasi ini dimulai dengan analisis dan perancangan sistem yang akan dibuat. Selanjutnya tahap implementasi sistem dengan tujuan menghasilkan *prototype* program yang sesuai dengan hasil perancangan yang telah ditentukan sebelumnya, yaitu penulisan kode program (*coding*) sesuai dengan metode *User-Item Based Collaborative Filtering* menggunakan lingkungan bahasa pemrograman PHP. Data yang digunakan akan disimpan dalam *Database Management System* (DBMS) MySQL. *Debugging* dilakukan untuk menguji aplikasi serta mencari *error* yang masih terdapat pada program untuk diperbaiki.

### 3.4 Pengujian Metode

Untuk mengetahui keefektifan metode *Collaborative Filtering* yang diimplementasikan, dilakukan pengujian dengan sampel data yang diambil dari data transaksi penyewaan di Odiva Video Rental cabang Fajar Indah Surakarta selama bulan Januari, Februari, dan Maret 2012. Prosedur pengujian yang dilakukan adalah sebagai berikut:

1. Pengambilan Sampel Data
  - a. Dipilih 208 *item*

*commit to user*

- b. Dipilih 33 *user* yang telah melakukan transaksi sewa terhadap 7 atau lebih *item*. Transaksi sewa dibatasi 7 *item* mengingat besarnya *sparsity* dan data sample yang digunakan. Dimana suatu kondisi dikatakan *sparsity* bila *rating* yang tersedia cenderung sangat sedikit dan tersebar (*sparse*) karena *user* biasanya hanya memberi *rating* pada (atau hanya menyewa) sebagian kecil saja dari keseluruhan *item* yang ada. Sehingga semakin banyak *item* yang disewa atau dirating oleh *user* akan semakin mengurangi tingkat *sparsity* (nilai *error* prediksi makin mengecil) yang berakibat semakin baik pula kualitas rekomendasi yang dihasilkan. Selain itu transaksi sewa tidak dibatasi lebih banyak item lagi (lebih besar dari 7) karena semakin banyak item yang menjadi batasan transaksi, maka *user* akan memerlukan waktu semakin lama pula untuk mendapatkan rekomendasi.
- c. Dipilih 443 transaksi yang melibatkan *user* dan *item* terpilih di atas.

Dari pengambilan sampel data didapat tingkat *sparsity* sebesar 93,6% sehingga peluang *user* meminjam sembarang *item* adalah 6,4 %.

## 2. Pengosongan *Rating*

Dari data sampel yang didapat, dipilih nilai *rating* untuk dikosongkan dengan menghilangkan transaksi secara acak mulai dari 10%, 20%, 30%, 40%, dan 50%, kemudian transaksi yang dihilangkan tersebut dipindahkan ke tabel baru.

## 3. Perhitungan Prediksi

Untuk setiap data sampel yang telah dikosongkan secara acak, dihitung prediksi menggunakan metode *User-Item Based Collaborative Filtering* dengan *Missing Value Algorithm* (MVA) dan *Absolute MVA*.

## 4. Perulangan Percobaan

Untuk masing-masing pengosongan nilai *rating* dan perhitungan nilai prediksi pada tiap metode, dilakukan percobaan sebanyak 40 kali.

*commit to user*

Pengujian dilakukan terhadap presisi dan akurasi hasil *cluster* dengan data testing yang digunakan

#### 1. Pengukuran Kualitas Rekomendasi Secara Presisi

Presisi adalah ukuran yang menunjukkan derajat kesesuaian antara hasil uji individual yang diukur melalui penyebaran hasil individual dari rata-rata secara berulang pada sampel-sampel yang diambil. Presisi dapat dihitung dengan Standar Deviasi:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Dengan:

$x_i = \{x_1, x_2, \dots, x_N\}$  : Nilai data dari sampel dan

$\bar{x}$  : Rata-rata dari sampel.

#### 2. Pengukuran Kualitas Rekomendasi Secara Akurasi

Dari data sampel yang didapat, dipilih nilai *rating* untuk dikosongkan dengan menghilangkan transaksi secara acak mulai dari 10%, 20% , 30%, 40% dan 50%. Untuk setiap data sampel yang telah dikosongkan secara acak, dihitung nilai prediksinya untuk kemudian diukur kualitasnya dengan menghitung nilai rata-rata kesalahannya dengan metrik MAE dan *error* probabilitas.

- a. Nilai MAE dari 40 kali percobaan yang dilakukan dihitung nilai rata-ratanya untuk mengukur kualitas rekomendasi.
- b. Nilai probabilitas kesalahan prediksi dari 40 percobaan yang dilakukan dihitung nilai rata-ratanya untuk mengukur kualitas rekomendasi. Nilai *threshold* yang digunakan adalah 0,065. Nilai *threshold* tersebut dipilih karena syarat pemilihan *threshold* adalah harus lebih besar dari 0,064 atau 6,4% (peluang *user* meminjam sembarang *item*). Diambil nilai minimumnya yaitu 0,065 agar memperoleh *error minimum* pula atau terkecil. *Error* probabilitas

*commit to user*



dihitung dengan membandingkan antara jumlah prediksi yang salah (dibawah *threshold*) dengan jumlah prediksi keseluruhan. Jadi logikanya bila *threshold* dinaikkan/diperbesar jumlah prediksi yang salah (berada dibawah nilai *threshold*) makin besar sehingga *error* yang diperoleh juga makin besar.

Hasil prediksi kontinu dikonversikan ke dalam bilangan biner 0 atau 1, sebagai berikut

$$pred_{biner} = \begin{cases} 0 & : \text{if } pred_{kontinu} < 0,065 \\ 1 & : \text{if } pred_{kontinu} \geq 0,065 \end{cases}$$

Contoh cara perhitungan *error* nya, misal dari 100 data ada 25 data yang bernilai <0,065 sedangkan sisanya 75 data bernilai  $\geq 0,065$  berarti 25 data kita konversikan menjadi 0 dan 75 data dikonversikan menjadi 1. Kemudian untuk menghitung kesalahan prediksi kita bagi jumlah prediksi yang salah yaitu 25 data dengan jumlah prediksi keseluruhan, 25/100 berarti nilai kesalahan prediksi sebesar 0,25.

## BAB IV

### PEMBAHASAN

#### 4.1 Hasil Pembahasan

Data yang digunakan pada penelitian ini terdiri dari data *item* VCD, data *user* dan data transaksi penyewaan *item*. Untuk data *item* terdiri dari 208 buah *item*. Data *user* sebanyak 33 *user* yang telah melakukan transaksi sewa terhadap tujuh atau lebih *item* terpilih dan data transaksi yang terdiri dari 443 transaksi yang melibatkan *user* dan *item* terpilih sebelumnya. Dari 443 data transaksi nantinya akan dibagi lagi menjadi training set dan test set untuk pengujian dimana besarnya *training set* maupun *test set* tergantung pada besarnya pengosongan yang dilakukan. Misalnya pada pengosongan 10% maka 10% data tersebut akan menjadi *test set* dan sisanya 90% data akan menjadi *training set*.

Gambar 4.1 dan 4.2 merupakan *screenshoot* salah satu contoh rekomendasi *top-N* untuk salah satu pengguna yang dihasilkan metode *User-Item Based Collaborative Filtering* baik dengan *cluster genre* maupun tanpa *cluster genre* pada *training set* 90% (pengosongan *rating* 10%). Rekomendasi yang dihasilkan adalah *Top-5* (lima *item* dengan nilai prediksi teratas).

Berikut ini Rekomendasi Film untuk JUMADI B RAHDIAN

ID Item	Judul Item	Aksi
8994181168425	KNIGHT AND DAY	<a href="#">Lihat</a>
8994181166865	A-TEAM	<a href="#">Lihat</a>
8994181174167	THOR	<a href="#">Lihat</a>
8994765016456	TRUE LEGEND	<a href="#">Lihat</a>
8994765016357	LITTLE BIG SOLDIER	<a href="#">Lihat</a>

Gambar 4.1 Top-5 Rekomendasi pada training set 90% tanpa *cluster genre*

*commit to user*



ID Item	Judul Item	Aksi
8994765011086	TWILIGHT	<a href="#">Lihat</a>
8994181168425	KNIGHT AND DAY	<a href="#">Lihat</a>
8994181161952	ICE AGE 3: DAWN OF THE DINOSAURS	<a href="#">Lihat</a>
8994356166959	CONFESSIONS OF A SHOPAHOLIC	<a href="#">Lihat</a>
8994181168865	A-TEAM	<a href="#">Lihat</a>

Gambar 4.2 Top-5 Rekomendasi pada training set 90% dengan *cluster genre*

Dapat dilihat pada gambar 4.2 bahwa setelah dilakukan *cluster genre* hasil rekomendasi berubah karena *item* yang direkomendasikan telah disesuaikan dengan *genre* film yang sudah pernah dipinjam oleh *user*.

#### 4.2 Pengujian Efektivitas Metode

Untuk mengetahui efektivitas metode *User-Item Based Collaborative Filtering*, pengujian menggunakan data yang diperoleh dari Odiva Rental VCD Fajar Indah. Data ini merupakan *user-item rating matrix* yang terdiri dari 33 pengguna dan 208 *item* dengan *rating* yang terisi sebanyak 443 *rating*. Data yang digunakan untuk pengujian hanya data yang nilai *rating* nya terisi atau nilainya tidak sama dengan nol.

Pengujian metode *User-Item Based Collaborative Filtering* dilakukan dengan menghitung prediksi nilai *rating* menggunakan rumus prediksi MVA, *absolute MVA* dan kombinasi *cluster genre* pada matriks *user-item*.

Data yang digunakan adalah data transaksi dengan pengosongan nilai *rating* secara acak mulai dari 10%, 20%,...,50%. Pengujian tersebut dilakukan sebanyak 40 kali. Pengujian metode *User-Item Based Collaborative Filtering* dengan rumus prediksi MVA, *absolute MVA* dan kombinasi *cluster genre* pada matriks *user-*

*item* menggunakan metrik evaluasi MAE, probabilitas kesalahan dan standar deviasi.

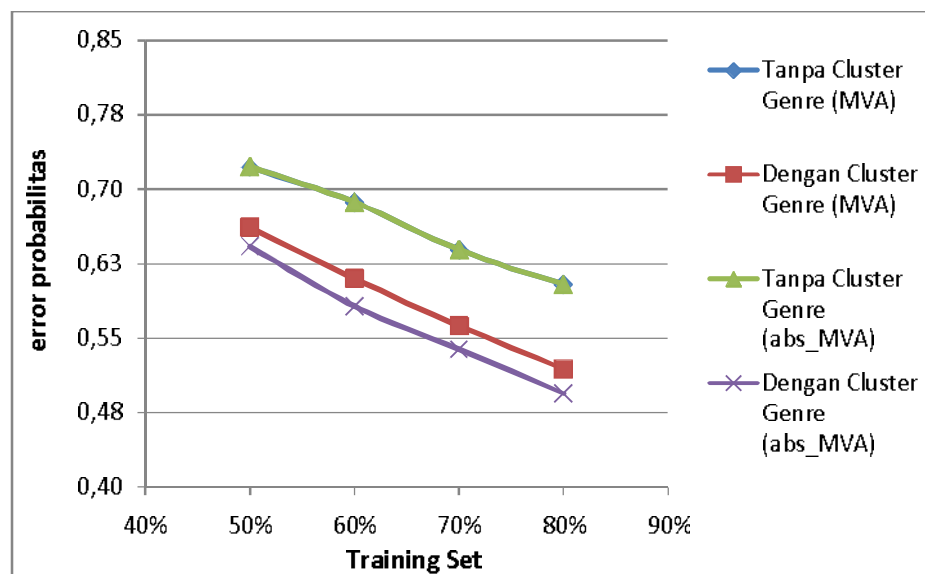
Seluruh hasil pengujian efektivitas yang terdiri dari perhitungan *error* probabilitas, MAE, dan standar deviasi dapat dilihat pada lampiran A. Dari data tersebut dapat dihitung rata-rata MAE, *error* probabilitas dan standar deviasi yang dihasilkan rumus prediksi MVA dan *absolute* MVA.

#### 4.2.1 Pengujian Secara Akurasi

Pengujian secara akurasi diukur dengan *error* probabilitas dan MAE. Akurasi menunjukkan kedekatan hasil pengukuran dengan nilai sesungguhnya pada penelitian ini akurasi diukur dengan menggunakan MAE dan *error* probabilitas. Pengujian dengan *error* probabilitas dan MAE masing-masing dilakukan sebanyak 40 kali pada *training set* 90% hingga 50%.

Tabel 4.1 Perbandingan nilai error probabilitas secara keseluruhan

Training Set	ERROR PROBABILITAS			
	MVA		ABSOLUTE MVA	
	Tanpa <i>cluster genre</i>	Dengan <i>cluster genre</i>	Tanpa <i>cluster genre</i>	Dengan <i>cluster genre</i>
90%	0,59833	0,48444	0,59833	0,46278
80%	0,60421	0,51910	0,60421	0,49466
70%	0,63929	0,56259	0,63929	0,53891
60%	0,68694	0,61011	0,68694	0,58244
50%	0,72218	0,66182	0,72275	0,64245
<b>Rata-rata</b>	0,65019	0,56762	0,65030	0,54425



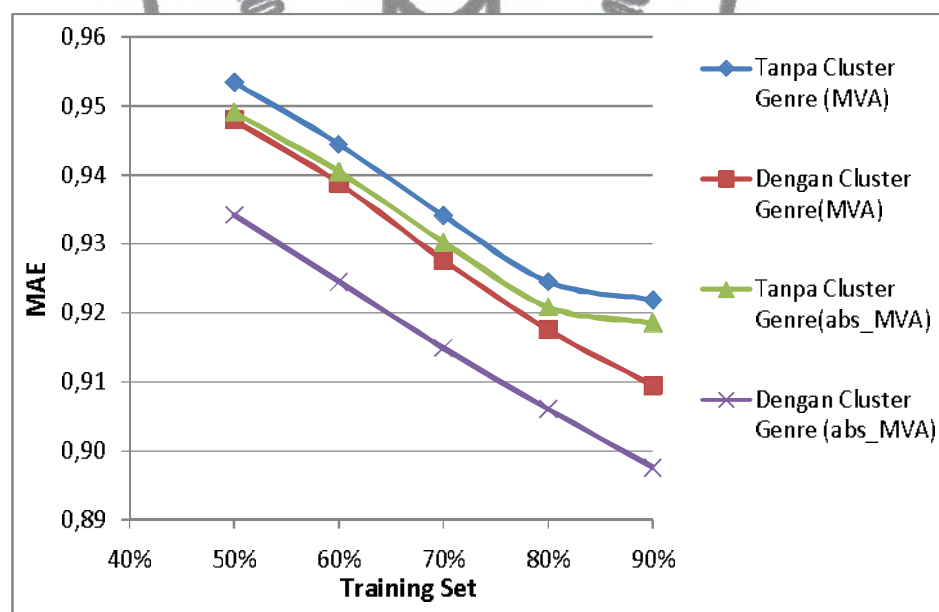
Gambar 4.3 Grafik perbandingan nilai *error probabilitas* secara keseluruhan

Dari tabel 4.1 dan gambar 4.3 menunjukkan bahwa nilai prediksi MVA maupun *absolute* MVA pada metode *User-Item Based Collaborative Filtering* dengan *cluster genre* menghasilkan kualitas rekomendasi yang lebih baik dibandingkan menggunakan metode *User-Item Based Collaborative Filtering* tanpa *cluster genre* dilihat dari kualitas prediksi yang dihasilkan. Hal ini bisa diketahui dari nilai *error probabilitas* yang dihasilkan oleh rumus prediksi MVA dan *absolute* MVA pada metode *User-Item Based Collaborative Filtering* dengan *cluster genre* lebih kecil daripada nilai *error probabilitas* yang dihasilkan oleh rumus prediksi MVA dan *absolute* MVA pada metode *User-Item Based Collaborative Filtering* tanpa *cluster genre*. Semakin kecil nilai *error probabilitas* menunjukkan semakin baik kualitas prediksi yang dihasilkan



Tabel 4.2 Perbandingan nilai MAE secara keseluruhan

Training set	MAE			
	MVA		ABSOLUTE MVA	
	Tanpa cluster genre	Dengan cluster genre	Tanpa cluster genre	Dengan cluster genre
90%	0,92192	0,90952	0,91856	0,89762
80%	0,92456	0,91761	0,92092	0,90611
70%	0,93414	0,92769	0,93031	0,91496
60%	0,94445	0,93881	0,94047	0,92455
50%	0,95346	0,94798	0,94911	0,93425
<b>Rata-rata</b>	0,93571	0,92832	0,93187	0,91550



Gambar 4.4 Grafik perbandingan nilai MAE secara keseluruhan

Dari tabel 4.2 dan gambar 4.4 di atas menunjukkan bahwa nilai prediksi MVA maupun *absolute MVA* pada metode *User-Item Based Collaborative Filtering* dengan *cluster genre* menghasilkan kualitas rekomendasi yang lebih baik dibandingkan menggunakan metode *User-Item Based Collaborative commit to user*.



*Filtering* tanpa *cluster genre* dilihat dari kualitas prediksi yang dihasilkan. Hal ini bisa diketahui dari nilai MAE yang dihasilkan oleh rumus prediksi MVA dan *absolute MVA* pada metode *User-Item Based Collaborative Filtering* dengan *cluster genre* lebih kecil daripada nilai MAE yang dihasilkan oleh rumus prediksi MVA dan *absolute MVA* pada metode *User-Item Based Collaborative Filtering* tanpa *cluster genre*. Semakin kecil nilai MAE menunjukkan semakin baik kualitas prediksi yang dihasilkan oleh sistem.

#### 4.2.2 Pengujian Secara Presisi

Setelah dilakukan pengosongan 10% hingga 50% sebanyak 40 kali, kemudian dihitung rata-ratanya untuk selanjutnya diukur presisinya dengan Standar Deviasi.

Tabel 4.3 Perbandingan nilai Standar Deviasi secara keseluruhan

Training set	Dengan <i>Cluster genre</i>				Tanpa <i>Cluster genre</i>				Rata-rata
	<i>Error Probb</i>		MAE		<i>Error Probb</i>		MAE		
	MVA	Abs MVA	MVA	Abs MVA	MVA	Abs MVA	MVA	Abs MVA	
90%	0,05851	0,05801	0,14756	0,16665	0,08649	0,08649	0,20785	0,21486	0,12830
80%	0,04869	0,04729	0,00899	0,03351	0,03351	0,13509	0,00556	0,00524	0,03974
70%	0,03310	0,03491	0,00498	0,00423	0,04162	0,04162	0,00368	0,00318	0,02091
60%	0,02967	0,03536	0,00355	0,00268	0,02799	0,02799	0,00239	0,00207	0,01646
50%	0,02725	0,02729	0,00212	0,00244	0,02323	0,02339	0,00152	0,00149	0,01359

Dari tabel Standar Deviasi diatas terlihat bahwa secara keseluruhan semakin kecil nilai training set, semakin kecil pula nilai standar deviasinya. Hal tersebut menunjukkan bahwa semakin kecil training set nilai presisinya semakin baik. Nilai presisi paling baik terdapat pada training set 50%.

#### 4.2.3 Analisis Hasil Pengujian

Pengujian metode *User-Item Based Collaborative Filtering* dengan rumus prediksi MVA dan *absolute MVA* pada sampel sebanyak 443 data transaksi yang dilakukan sebanyak 40 kali dengan *training set* secara acak dari 90% sampai dengan 50% sebagaimana terlihat dalam Tabel 4.2 diperoleh hasil bahwa penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi MVA) dengan *cluster genre* menurunkan nilai rata-rata MAE sebesar 0,00739 (dari 0,93571 menjadi 0,92832) atau sebesar 0,8% dari nilai rata-rata MAE yang dihasilkan oleh penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi MVA) tanpa *cluster genre*.

Sedangkan penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi *absolute MVA*) dengan *cluster genre* menurunkan nilai rata-rata MAE sebesar 0,01637 (dari 0,93187 menjadi 0,91550) atau sebesar 1,8% dari nilai rata-rata MAE yang dihasilkan oleh penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi *absolute MVA*) tanpa *cluster genre*.

Untuk penggunaan rumus prediksi *absolute MVA* pada metode *User-Item Based Collaborative Filtering* dengan *cluster genre* berhasil menurunkan nilai rata-rata MAE sebesar 0,01282 (dari 0,92832 menjadi 0,91550) atau sebesar 1,4% dari nilai rata-rata MAE yang dihasilkan dengan rumus prediksi MVA. Sedangkan penggunaan rumus prediksi *absolute MVA* pada metode *User-Item Based Collaborative Filtering* tanpa *cluster genre* menurunkan nilai rata-rata MAE sebesar 0,00383 (dari 0,93571 menjadi 0,93187) atau sebesar 0,4% dari nilai rata-rata MAE yang dihasilkan dengan rumus prediksi MVA.

Akan tetapi, MAE kemudian diketahui cenderung tidak sesuai digunakan untuk pengujian akurasi hasil prediksi pada sistem perekomendasi yang menggunakan data *rating* biner. Hal ini dikarenakan MAE menggunakan *rating* prediksi kontinu yang bernilai antara 0 hingga 1 untuk mengukur kesalahan prediksi, padahal penelitian ini lebih berbasis pada kejadian (*occurrence*) biner apakah seorang *user* menyewa atau belum menyewa *item* tertentu, yang dengan

kata lain lebih bersifat probabilistik. Oleh karena itu, metrik probabilitas kesalahan lebih tepat digunakan untuk mengukur akurasi prediksi.

Pengujian dengan menggunakan perhitungan probabilitas kesalahan ditunjukkan dalam Tabel 4.1 dengan hasil bahwa penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi MVA) dengan *cluster genre* menurunkan nilai rata-rata *error* probabilitas sebesar 0,08257 (dari 0,65019 menjadi 0,56762) atau sebesar 12,7% dari nilai rata-rata *error* probabilitas yang dihasilkan oleh penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi MVA) tanpa *cluster genre*.

Sedangkan penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi *absolute* MVA) dengan *cluster genre* menurunkan nilai rata-rata *error* probabilitas sebesar 0,10605 (dari 0,65030 menjadi 0,54425) atau sebesar 16,3% dari nilai rata-rata *error* probabilitas yang dihasilkan oleh metode *User-Item Based Collaborative Filtering* (rumus prediksi *absolute* MVA) tanpa *cluster genre*.

Untuk penggunaan rumus prediksi *absolute* MVA pada metode *User-Item Based Collaborative Filtering* tanpa *cluster genre* berhasil menurunkan *error* probabilitas sebesar 0,00011 (dari 0,65019 menjadi 0,65030) atau sebesar 0,02 % dari nilai rata-rata *error* probabilitas dengan rumus prediksi MVA. Sedangkan pada metode *User-Item Based Collaborative Filtering* dengan *cluster genre* rumus prediksi *absolute* MVA berhasil menurunkan sebesar 0,02337 (dari 0,56762 menjadi 0,54425) atau sebesar 4,1 % dari nilai rata-rata *error* probabilitas dengan rumus prediksi MVA.

Dari hasil pengujian diatas terlihat bahwa penggunaan metode *User-Item Based Collaborative Filtering* dengan *cluster genre* memberikan hasil yang lebih baik dibandingkan metode *User-Item Based Collaborative Filtering* tanpa *cluster genre*. Hal ini terbukti dengan penurunan nilai rata-rata kesalahan, dimana nilai rata-rata *error* probabilitas lebih menunjukkan penurunan nilai *error* yang signifikan dibanding MAE. Sedangkan untuk penggunaan rumus prediksi terlihat bahwa rumus prediksi *absolute* baik pada MAE maupun *error* probabilitas hanya

*commit to user*

menurunkan sedikit sekali nilai *error* atau tidak terlalu menunjukkan penurunan yang besar. Oleh karena itu dapat dikatakan bahwa rumus prediksi *absolute MVA* tidak menunjukkan penurunan yang signifikan baik pada *error* probabilitas maupun MAE.

Sedangkan secara presisi, pengujian pada training set 50% memberikan hasil yang paling baik ditinjau dari nilai rata-rata standar deviasi yang paling kecil dibandingkan pada level *training set* lainnya yaitu sebesar 0,01359.

Secara akurasi, metode *User-Item Based Collaborative Filtering* dengan *cluster genre* lebih efektif dalam memberikan hasil kualitas rekomendasi ditinjau dari rata-rata nilai MAE dan *error* probabilitas. Sedangkan penggunaan rumus prediksi *absolute MVA* tidak menunjukkan penurunan nilai kesalahan yang signifikan sehingga tidak cukup membantu meningkatkan akurasi metode *User-Item Based Collaborative Filtering*.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Dari hasil penelitian, dapat disimpulkan bahwa secara akurasi metode *User-Item Based Collaborative Filtering* dengan *cluster genre* lebih efektif dalam menghasilkan rekomendasi dibandingkan dengan metode *User-Item Based Collaborative Filtering* tanpa *cluster genre*. Keefektifan tersebut ditunjukkan dengan kualitas akurasi prediksi yang terukur dari rata-rata nilai MAE maupun *error* probabilitas. Namun, metrik MAE cenderung tidak sesuai digunakan untuk pengujian akurasi hasil prediksi pada sistem perekomendasi yang menggunakan data *rating* biner. Hal ini dikarenakan MAE menggunakan *rating* prediksi kontinu yang bernilai antara 0 hingga 1 untuk mengukur kesalahan prediksi, padahal penelitian ini lebih berbasis pada kejadian (*occurrence*) biner apakah seorang *user* menyewa atau belum menyewa *item* tertentu, yang dengan kata lain lebih bersifat probabilistik. Oleh karena itu, metrik probabilitas kesalahan lebih tepat digunakan untuk mengukur akurasi prediksi.

Pengujian dengan menggunakan perhitungan probabilitas kesalahan seperti pada Tabel 4.1 menunjukkan bahwa penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi MVA) dengan *cluster genre* menurunkan nilai rata-rata *error* probabilitas sebesar 0,08257 (dari 0,65019 menjadi 0,56762) atau sebesar 12,7% dari nilai rata-rata *error* probabilitas yang dihasilkan oleh penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi MVA) tanpa *cluster genre*. Sedangkan penggunaan metode *User-Item Based Collaborative Filtering* (rumus prediksi *absolute* MVA) dengan *cluster genre* menurunkan nilai rata-rata *error* probabilitas sebesar 0,10605 (dari 0,65030 menjadi 0,54425) atau sebesar 16,3% dari nilai rata-rata *error* probabilitas yang dihasilkan oleh metode *User-Item Based Collaborative Filtering* (rumus prediksi *absolute* MVA) tanpa *cluster genre*. Secara akurasi juga dapat disimpulkan bahwa performansi metode *User-Item Based Collaborative Filtering* tidak cukup baik

*commit to user*



diterapkan pada *rating* implisit ditunjukkan dengan nilai kesalahan paling kecil masih diatas 50 persen yaitu sebesar 0,54425 (*absolute MVA*).

Secara presisi, pengujian pada training set 50% memberikan hasil yang paling baik ditinjau dari nilai standar deviasi yang paling kecil dibandingkan pada level *training set* lainnya yaitu sebesar 0,01359.

Sedangkan penggunaan rumus prediksi *absolute MVA* menunjukkan penurunan nilai yang sangat sedikit sekali dibandingkan dengan *User-Item Based Collaborative Filtering* menggunakan MVA. Hal tersebut ditunjukkan dengan kualitas prediksi yang terukur dari rata-rata nilai MAE maupun *error* probabilitas yang dihasilkan *absolute MVA* tidak menunjukkan penurunan nilai kesalahan yang signifikan sehingga tidak cukup membantu meningkatkan akurasi metode *User-Item Based Collaborative Filtering*.

## 5.2 Saran

Berdasarkan kesimpulan di atas dapat diberikan saran sebagai berikut:

1. Ketahanan metode *User-Item Based Collaborative Filtering* menggunakan *cluster genre* maupun *absolute MVA* terhadap tingkat *sparsity* masih perlu diuji terutama untuk tingkat *sparsity* kurang dari 93,6% untuk lebih meyakinkan pengujian dan menemukan kelemahan lain yang mungkin ada pada metode *User-Item Based Collaborative Filtering* menggunakan *cluster genre* maupun *absolute MVA*.
2. Pada penelitian selanjutnya, *rating* implisit yang digunakan dapat dikembangkan menjadi sudah menyewa, belum menyewa, dan tidak mau menyewa. Cara pengumpulan *rating* yang demikian ini bisa dilakukan antara lain dengan mengamati *behaviour member* ketika ia menyatakan bahwa ia tidak menyukai salah satu atau beberapa film yang direkomendasikan. Dengan lebih bervariasinya jenis *rating* yang digunakan, dapat diuji apakah kualitas rekomendasi yang dihasilkan bisa lebih baik lagi.
3. Untuk penelitian selanjutnya dapat mempertimbangkan fitur tahun rilis film dalam pembuatan rekomendasi untuk mengetahui lebih jauh bagaimana

*commit to user*



pengaruh jarak tahun rilis antara satu *item* dengan *item* lainnya terhadap kualitas rekomendasi yang dihasilkan.

